David McGookin
Stephen Brewster (Eds.)

LNCS 4129

Haptic and Audio Interaction Design

First International Workshop, HAID 2006
Glasgow, UK, August/September 2006
Proceedings

Springer

Lecture Notes in Computer Science 4129

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

David McGookin Stephen Brewster (Eds.)

Haptic and Audio Interaction Design

First International Workshop, HAID 2006
Glasgow, UK, August 31 – September 1, 2006
Proceedings

Springer

Volume Editors

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

Library of Congress Control Number: 2006930675

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 3-540-37595-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-37595-1 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11821731 06/3142 5 4 3 2 1 0

Preface

Overview

The International Workshop on Haptic and Audio Interaction Design was organized as part of the EPSRC-funded MultiVis project. The main focus of the workshop was to investigate how the modalities of sound and haptics (touch) could be used together in interaction, and what novel interfaces could be provided when they are used in conjunction. What are the relative advantages of each of the modalities when used alone and together? Are there reasons why haptic-based information is more useful in certain situations than equivalent audio information? How can different modalities be used together to create compelling and useful interaction with computer-based systems? We posed these questions to researchers around the world, asking them to submit novel work which sought to discover answers. Thirty papers were submitted of which 15 were accepted. Each paper was peer reviewed at least twice using an esteemed set of leading international figures from both academia and industry, to whom we are grateful for the quality of their reviews, time, patience and responding within our tight schedule.

The papers presented at the workshop come from a wide variety of disciplines ranging from psychology to art, showcasing how haptics and sound can improve user interaction with computers; challenging us to move beyond simple mouse and keyboard metaphors to produce interfaces for devices and applications that allow for the full range of human interactivity. Below the papers are categorized and summarized based on their application and focus.

Visual Impairment

Much of the benefit of haptic and audio interaction is in situations where users may be unable to use a visual display. In such situations the haptic and audio modalities must shoulder responsibility that would otherwise be undertaken by the visual sense; requiring that the bandwidth of this modality is used efficiently. Magnusson *et al.* investigate how sound can be used in haptic navigation tasks to improve target location in non-visual environments. McAllister *et al.* consider how sound and haptics can be combined to improve access to image-based graphs. Both Winberg and Sallnäs *et al.* show how sound and haptics can create rich environments that allow for collaboration between sighted and visually impaired people. Rassmus-Gröhn *et al.* investigate how haptics can be used to support non-visual drawing activities amongst visually impaired children. These papers show us how much, when used together, the haptic and audio modalities can overcome the loss of vision creating rich and engaging user experiences.

Research Approaches

An additional approach in considering how to use haptic and audio interaction effectively is from the more theoretical perspective, with existing design approaches being adapted to fit with new modalities. Murphy *et al.* consider how semiotics can be applied to the design of non-speech audio to make the resulting audio cues more easily learnable and understandable by the users. Coleman *et al.* discuss how the existing approaches of ethnography can be adapted to investigate new areas where sound can be used. O'Sullivan and Chang propose a new descriptive language for vibration that should allow for more effective use. Vickers argues that by considering the relationship between haptics and sound in musical terms, we can create more effective and richer interactions.

Psychophysics

Whilst applications showing how to improve haptic and audio interaction design are of vital importance, it is also important to consider how audio and haptic stimuli interact at a more basic level to provide clues as to where future effort should be focused, as well as where it should not! Avanzini and Crossato investigate the contribution of audio and haptics on the perception of contact stiffness. Tikka and Laitinen uncover design principles to successfully use vibration in conjunction with touch screen technology.

Mobile Applications

As with visual impairment, mobile devices restrict the user's ability to interact with a visual display. Crossan and Murray-Smith investigate how haptic input can overcome the problems of selection in mobile environments. Using gesture-based interaction their system allows selection of songs by rhythmic tapping. Vanacken *et al.* identify how sound and haptics can improve selection in virtual environments.

Art and Leisure

Whilst many of the interactions discussed by authors involve measured performance improvements and "work" style applications, there has been some consideration as to how effective haptic and audio interaction design can make more engaging or entertaining experiences for the user. Martens and Walker investigate how low-frequency vibrations delivered both through sound and directly via a haptic "platform" can improve the sensation of presence and realism within a movie. Barras discusses how the use of haptic and audio interaction can create an engaging interactive experience to communicate historical stories. Barras describes the development of a haptic and audio-based interactive art exhibit that describes the imminent extinction of the great apes.

Through the work contained in these papers we can begin to see the potential of using complementary haptic and audio modalities in interaction design, as well as the initial results that show us how to do so effectively.

August 2006 David McGookin and Stephen Brewster

Organization

The International Workshop on Haptic and Audio Interaction Design 2006 was organized by the Department of Computing Science, University of Glasgow, UK.

Workshop Chairs: David McGookin and Stephen Brewster (University of Glasgow, UK).
Organization: Andrew Crossan, Eve Hoggan and Johan Kildal (University of Glasgow, UK).

Referees

Stephen Barrass, University of Canberra, Australia
Graeme Coleman, University of Dundee, UK
Mikael Fernström, University of Limerick, Ireland
Christopher Frauenberger, Queen Mary, University of London, UK
Antonio Frisoli, Scuola Superiore Sant'Anna, Italy
Stephen Furner, BT, UK
Matti Gröhn, CSC Scientific Computing, Finland
Matthias Harders, ETH, Switzerland
William Harwin, University of Reading, UK
Andy Hunt, University of York, UK
Gunnar Jansson, Uppsala University, Sweden
Roberta Klatzky, Carnegie-Mellon University, USA
Jukka Linjama, Nokia, Finland
Charlotte Magnusson, University of Lund, Sweden
Graham McAllister, Queen's University Belfast, UK
Margaret McLaughlin, University of Southern California, USA
Sile O'Modhrain, Queen's University Belfast, UK
Connor O'Sullivan, Motorola, USA
Antti Pirhonen, University of Jyväskylä, Finland
David Prytherch, University of Central England, UK
Roope Raisamo, University of Tampere, Finland
Chris Raymaekers, Limburg University, Belgium
Jonathan Roberts, University of Kent, UK
Gabriel Robles-de-la-Torre, International Society for Haptics, Mexico
Tony Stockman, Queen Mary, University of London, UK
Paul Vickers, University of Northumbria, UK
Bruce Walker, Georgia Institute of Technology, USA
Tamar Weiss, University of Haifa, Israel
Fredrik Winberg, KTH, Sweden
Mark Wright, Edinburgh School of Art, UK
Wai Yu, Queen's University Belfast, UK
Hong Z. Tan, Purdue University, USA

Sponsors

The workshop was supported by the EPSRC-funded Multivis project (GR/S86150/01) and Advanced Research Fellowship (GR/S53244/01) and the EU FP6 IST MICOLE project. We are grateful for the support of the Glasgow Interactive Systems Group. For more details on the workshop see www.multivis.org.

multivis.org

EPSRC
Engineering and Physical Sciences
Research Council

UNIVERSITY OF
STRATHCLYDE

GIST
GLASGOW INTERACTIVE
SYSTEMS GROUP

UNIVERSITY
of
GLASGOW

Table of Contents

Session: Interaction

Session: Psychophysics

Session: Music and Gesture

Session: Visual Impairments I

Session: Visual Impairments II

Session: Design I

Session: Design II

Perception of Audio-Generated and Custom Motion Programs in Multimedia Display of Action-Oriented DVD Films

Kent Walker and William L. Martens

Sound Recording Area, Department of Theory, Schulich School of Music,
McGill University, 555 Rue Sherbrooke Ouest, Montréal, QC, H3A 1E9, Canada
kent.walker@mail.mcgill.ca
http://www.music.mcgill.ca/~kentw/

Abstract. This paper addresses a practical problem associated with multimedia display systems which utilize motion-platforms or chairs. Given audio-visual content for which motion data is not available, motion may be automatically generated from multichannel audio, usually from a Low-Frequency Effects channel (LFE) such as that distributed on Digital Versatile Discs (DVDs). Alternatively, custom motion programs may be created to accompany multimedia content. This paper presents the results of a study designed to test the sense of realism, sense of presence, and global preference for multimedia playback in these two distinct cases of platform accompaniment: motion which has been generated automatically from audio, and motion which has been designed expressly for the purpose of stimulating appropriate haptic and vestibular sensations.

1 Introduction

1.1 Research Context and Research Questions

Since the advent of the Digital Versatile Disc (DVD), affordable advances in multimedia display technology have evolved which offer consumers opportunities for enhanced experience in the form of devices capable of haptic stimulation and motion accompaniment (whole-body vibration). These devices are many and fall into several categories but generally may be described as platforms and chairs which are capable of moving and shaking observers. Given the variety of devices available and the absence of standardized practice, multimedia content providers such as Hollywood film houses have yet to supply this niche market with motion programs for home entertainment use.

As a result developers of haptic technology are faced with the question of how to generate motion programs for their customers. Two main options exist: the first is to generate motion programs automatically using existing content; the second is for hardware manufacturers to employ programmers to create custom-coded motion from scratch for content from media houses. Depending on the program material or the multimedia scene presented to the observer, either method of generating whole-body vibration may be sufficient in terms of enhancing an

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 1–11, 2006.

observer's experience. For example, previous investigations by the authors [15] and others [3] have indicated that for music-only material, generation of whole-body vibration using existing audio signals is sufficient for the enhancement of experience provided that the systems are properly coordinated. In other words, for music programs separate vibration recordings or custom motion programs are generally not required. However, in the case of more action-oriented program material this may not be the case. Though there is a significant body of research on how to create motion-cues for users of flight simulation technology via vestibular stimulation [18], there has been relatively little work done on the most effective integration of motion platforms into home entertainment oriented audiovisual content display. Different multimedia scenes may require different kinds of motion programs in order to create within observers the greatest suspension of disbelief.

The experiments described herein investigated multimedia display of commercially available DVD film titles. Motion data generated from audio Low Frequency Effects channels (the *.1* channel on a *5.1* channel disc) was compared to motion generated specifically for individual DVD titles by a human designer. Experiments were designed to determine if existing information present in LFE channels is sufficient for the enhancement of presence and realism, if observers prefer custom-coded motion when watching DVD titles, and if the extra costs involved in the creation of custom motion programs is perceptually justifiable. The tested hypothesis was: for action-oriented DVD titles is custom-coded motion more likely to be preferred over audio-generated motion, and is this preference linked to an observers sense of presence and sense of realism for a displayed multimedia scene?

Before further explaining the experiments conducted, background on the technology for motion-based display and human sensitivity to vibration is presented.

1.2 Enhancing Presence in Multimedia Display

In recent research and development of advanced multimedia display technology, great emphasis has been placed upon multichannel sound and the enhanced consumer experience associated with coordinated display of visuals and spatial audio content. The potential for user immersion in the presented virtual world is one benefit of such multimedia display which is most properly called *television-type telesensation* [13]. Compared to more conventional media, such immersive audio-visual content produces a higher proportion of user responses indicating higher *sense of presence* or "sense of being transported to the electronically-mediated space" [7]; consumers can forget that the virtual world presented to their eyes and ears is an electronic reproduction, and imagine instead that they are experiencing the virtual world first hand. However, this suspension of disbelief is weakened considerably by one factor that has often been ignored in the development of advanced multimedia displays: observers visiting these virtual worlds are *not* disembodied minds. Regardless of where observers' eyes and ears take them, their bodies most often stay put in the physical display space. On the other hand, if simulation includes touch and motion sensations which are consistent with what is seen and heard, a heightened sense of presence is to be expected.

There has been a small but growing interest in providing a means for creating such high-quality *multimodal* experiences for consumers in home theater and computer gaming applications, typically in the form of moving seats or motion platforms. Such multimodal content has the potential to not only enhance experience, but to spawn new markets for both entertainment and electronics corporations in a commercial age fraught with the terror of economic uncertainly, caused largely by the real threat of entertainment piracy. To date, the film industry's reaction to piracy has mostly been the so-called *high-definition* technological revolution of home-theater, resulting most recently in an emerging format war between High-Definition Digital Versatile Disc (HD-DVD) and Blu-Ray Disc (BD). Nevertheless, experts in these new formats question the ability of these systems to offer extra value over what's currently available to consumers via existing popular DVD standards, as many consumers have lesser quality sound reproduction systems and televisions that will benefit perhaps only marginally from these higher-definition media [12]. However, if DVD houses embrace and promote haptic and motion transducer technology together with their new formats, they may have a greater degree of success with the high-definition revolution: the addition of a mode of display is more immediately gratifying to consumers than just-noticeable differences in display quality for audio and video alone.

1.3 Haptics and Motion Sensation

Haptic comes from the Greek *hapt esthai* which may be literally translated as *feel sense* [1]. Haptic sensation can result from many forms of stimulation, ranging from local vibrations on the skin to whole-body movement, and relies on a variety of sense organs and receptors from multiple sensory systems. Cutaneous sensory organs such as *Meissner's Corpuscles* (used for the sensation of vibration below approximately 40 Hz), *Pacinian Corpuscles* (used above 40 Hz), and *Merkel Discs* (used for the sensation of pressure on the skin) [4]. The haptic senses also include the organs of human kinesthesia such as muscles, tendons, and joints (used for the sensation of movement below 30 Hz), although such vibration-related stimulation is distinguished from the stimuli for the vestibular organs of the inner-ear, namely the *semi-circular canals* (used for detection of angular acceleration of the head), and the *otoliths* (used for detection of linear acceleration). It is important to note that there is a great deal of overlap in our sensory systems with regards to vibration, as our hearing sensitivity extends down to approximately 20 Hz, while our vibration sensitivity grows with frequency to peak at approximately 250-300 Hz; therefore, low-frequency sound sources are often *both heard and felt.*

1.4 Motion Degrees of Freedom (DOF)

In control of motion platforms, it is necessary to distinguish how freely the platform may be made to move through space. Complete freedom of motion in space admits Six Degrees of Freedom (6DOF), which include the possibility of three directions of displacement and three angular gyrations. The terms that are used in the technical literature to describe an object or observer's motion through space are not in such popular use, and so these terms are related to more common language in this section. To begin with, there are three terms that are used to describe the rotation of an observer whose spatial position does not change: yaw, roll, and pitch. If a person who is standing upright rotates to the left or right about the vertical axis through their body, this rotation is termed yaw. If an airplane were to dive forward, the rotation about the left/right axis would be termed pitch, and if the airplane were to tilt to one side, the rotation about the front/rear axis would be termed roll (not to be confused with the acrobatic maneuver termed a *barrel roll* in which an airplane not only rotates along its longitudinal axis, but also follows a trajectory along the surface of a cylinder or barrel). The remaining set of three terms is used to describe the spatial translation of an observer whose orientation in space does not change. These three translational motion terms, describing movement leftward, forward, or upward, are respectively: sway, surge, and heave.

2 Methods

2.1 The Loudspeaker Array

The current study used a calibrated multichannel stereophonic sound system compliant with ITU standards [6] yet with full range capabilities at all standard angles (in degrees relative to the median plane the angles were -110, -30, 0, 30, and 110). Five low-frequency drivers (ranging from 35 Hz to 300 Hz) and five higher-frequency drivers (ranging from 300 Hz to well over 20,000 Hz). The low-frequency drivers were Mini-Mammoth subwoofers manufactured by Quebec-based D-BOX Technology, which were placed at standard locations for the 5 main speakers in surround sound reproduction. The higher-frequency drivers were dipole radiating, full range transducers featuring the Planar Focus Technology of Level 9 Sound Designs of British Columbia. These speakers were placed at the same azimuth angles as the 5 low-frequency drivers, but in a ring positioned 15 degrees above the observer's ear level. All listening tests were conducted in a treated room [11].

2.2 The Visual Display

A Panasonic model TH-50PHD6UY HDTV 50-inch plasma display with a resolution of 1366 x 768 and contrast ratio of 3000:1 was used to display the video portion of all DVD titles. A Pioneer model DV 563A DVD player was also used. Titles were of the widescreen variety and observers sat at a distance of 2 meters from the screen.

2.3 The Motion Platform

The research presented here used a commercially available motion platform, the Odyssée system from the Quebec based company D-BOX Technology [9]. This

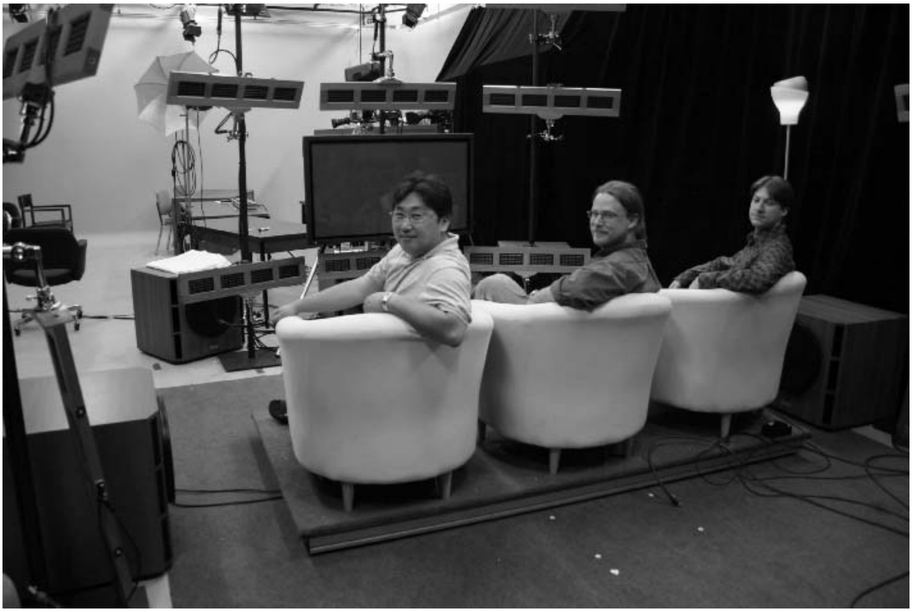


Fig. 1. A photograph of the installation at the Immersive Presence Lab at McGill University. The loudspeaker array, visual display, and motion platform are visible. Three participants (not the individuals shown) sat at the positions indicated.

system uses four coordinated actuators to provide multiple users with motion in three Degrees of Freedom (3DOF) for a home theater setting. No rotation about the vertical axis is allowed by this system, so no change in heading (or yaw) can be simulated. But two other rotations are enabled, and also translation along a single axis. The three terms describing these control possibilities are summarized in Table 1, along with an example of a command that could be sent to the motion platform to produce each type of motion.

The motions generated by the Odyssée system are based upon its use of four actuators positioned at the corners of the motion platform, each of which is independently controlled but only capable of motion along a vertical axis. When all four actuators move together, users can be displaced linearly upwards or downwards with a very quick response and with considerable force (the feedback-corrected linear system frequency response is flat to 100 Hz). The two angular motions are enabled as follows: upward motion of the two left actuators coupled with downward motion of the two right actuators enables pure roll, and contrary motion of the front and rear actuators allows for pitch control.

A detailed set of measurements was performed on the platform system using an Agilent 35670A Dynamic Signal Analyzer as the measurement device, as well as a PCB Piezoelectronics Model 356 416 tri-axial accelerometer. This was done in order to characterize general aspects of the performance of the platform, as well as its reaction to the stimuli used in perceptual experiments of motion quality.

Table 1. Motion Possibilities of Platform Used in Experiments

Motion Label	Description and Command Example
Roll (Y)	Rotation about the left/right axis ("tilt left")
Pitch (X)	Rotation about the front/back axis ("tilt back")
Heave (Z)	Upward/downward translation ("rise up")

It was noticed at this time that the measured results were affected by the location on the platform where the measurements were taken. Different areas had differing modal response when stimulated. It was also noticed during measurement that the weight on the platform had drastic effects on active-servo performance. As a result, measurements were taken with one of the participants sitting on the platform in the center seat used for perceptual testing. For these reasons, the majority of measurements were taken directly on the platform where the participants placed their feet, as well as directly from the hand of one participant. While measurements taken at the platform surface demonstrated that it was indeed moving only mostly with regards to heave (up and down) when heave was input to the system, measurements taken at the hand displayed movement in multiple dimensions; this makes sense as the observer was not asked to hold his body completely still during measurement (indeed given the force of acceleration applied to the platform this may have been impossible). The peak acceleration values along the heave axis were found to range from approximately 100 $\mathit{milli}\text{-}g$ to 197 $\mathit{milli}\text{-}g$ ($1g = 9.8\,m/s^2$) depending on the stimulus and the measurement location. Motion measured at the hand of an observer seated on the motion platform for an excerpt from the Hollywood blockbuster *Master and Commander* is

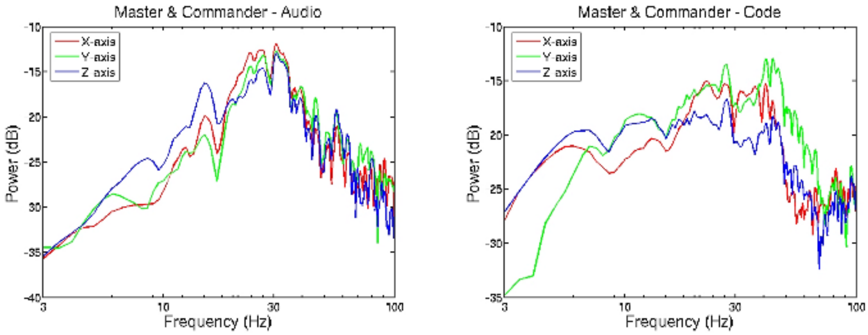


Fig. 2. Motion measured at the hand of an observer seated on the motion platform for an excerpt from the Hollywood blockbuster *Master and Commander*. Motion generated from the Audio LFE channel is shown on the left. Motion resulting from human-designed FX-Codes that were provided by D-Box Technologies are shown on the right. The Code motion measurement displays a greater variety of motion, as well as a lower frequency-extension, as frequencies below approximately 20 Hz are not normally part of multichannel audio programmes.

shown in Figure 2. Measurements were taken under two conditions: *Audio* (shown on the left) where motion was automatically generated by the LFE channel, and *Code* (shown on the right) where motion codes were custom crafted by a human motion-designer. D-Box Technologies offers more than 300 custom crafted motion codes to customers of its whole-body vibration products. While the *Audio* measurement displays axes which appear to be relatively more correlated than the *Code* measurement, the apparent decorrelation of *Audio* X Y Z are created by the human body's natural response to whole-body vibration (which has 6 degrees of freedom). The *Code* motion measurement displays a greater variety of motion, as well as a lower frequency extension, as frequencies below approximately 20 Hz are not normally part of multichannel audio programmes.

2.4 Measurements of Crosstalk Between Display Systems

Crosstalk between subwoofers and the motion platform was also measured, as subwoofers are capable of inducing structural vibration as well as haptic sensation in listeners. This was measured to be less than 20 *milli-g*. The reverse was also true as the motion platform did induce airborne vibration experienced as sound, however, this was measured to be approximately 40 dB C below the sound level of the loudspeakers.

2.5 Tests of Perceived Motion Quality

In a blind test, three observers were shown a total of 30 excerpts from seven different Hollywood DVD titles shown in Table 2. Each excerpt was presented with both audio-generated motion as well as custom motion in random order. Subjects were asked to rate motion on a 10-point scale (using pen and paper) with regards to three perceptual attributes: sense of realism, sense of presence, and global preference. As can be seen the measurements shown in Figure 2, the frequency content of the audio-generated and custom motion were significantly different. Observers could, therefore, easily discriminate a difference between each type of motion, however, they were not told which type of motion they were experiencing.

Table 2. Hollywood action-oriented DVD titles used for the generation of whole-body vibration and motion

DVD Title	Film Studio
Cast Away	20th Century Fox
Charlie's Angels (2000)	Sony Pictures Entertainment
I-Robot	20th Century Fox
Finding Nemo	Walt Disney Video
Master and Commander - The Far Side of the World	20th Century Fox
Perfect Storm	Warner Home Video
Terminator 3	Warner Home Video

3 Results

Figure 3 displays mean values for 30 LFE-derived and coded-motion excerpts of 1 minute in length from the seven action-oriented Hollywood DVD titles (see Table 2). There is a significant difference in the mean ratings observed for each subject (with regards to each attribute), and the mean *differences* between subjects for each attribute were relatively consistent. With regards to all attributes, subjects' mean ratings for coded-motion (Code) were higher than those for LFE-derived motion (Audio). Mean differences were greatest with regards to global preference (Pref) and smallest for sense of presence (Pres). It is clear that the coded motion was greatly preferred over audio-generated motion.

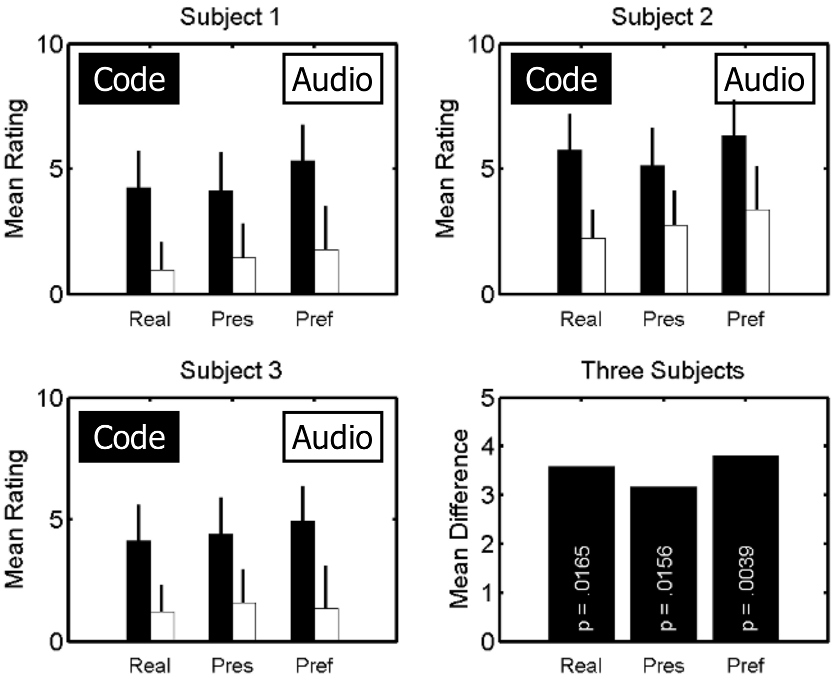


Fig. 3. Ratings of sense realism (Real), sense of presence (Pres), and global preference (Pref) from three observers exposed to both motion generated automatically from an LFE audio program and custom generated coded-motion. Observers were shown 30 excerpts of approximately 1 minute each, from seven different Hollywood DVD titles (see Table 2). Although there is a significant difference in the mean ratings each subject made with regards to each attribute, the mean *differences* between subjects is relatively consistent.

4 Discussion

Automatically derived LFE generated motion was taken directly from the *.1* LFE channel of the Dolby Digital signal; this channel of the AC-3 encoded signal is band-limited to frequencies below 120 Hz and is sampled at a rate of 240 Hz [14]. While automatically derived LFE motion moved the platform only up and down in the Z dimension, coded-motion was present in all 3 dimensions (Z, X, and Y).

Arguably, a significant step towards enhancing presence (and perhaps haptic sensation) in sound reproduction was taken with the development of the LFE channel for moving pictures [10]. This channel was originally designed for 70 mm film in the 1970s to allow for the addition of effects with lots of low-frequency energy [5]. This technology has since been disseminated to the consumer in the form of 5.1 multichannel (also known as 3-2-1) home theater systems of varying quality which are also used for music reproduction. It is important to note that a *dedicated channel* for LFE (the *.1* channel) should not be taken to be synonymous with the *number of subwoofers* in a reproduction system; both the ideal number and position of subwoofers in multichannel stereophonic speaker arrays (incorporating various bass handling processes) is a question still under investigation by many interested in acoustics and applied psychoacoustics.

That the LFE channel was originally designed for effects and that it is almost exclusively used for this purpose in home theater is of great importance to haptic-vestibular-motion systems which use this channel to provide motion to the observer. It is an assurance that only information relevant to the haptic experience will be transferred. In other words, low-frequency sounds which are not effect related but are present in the five full-range channels (perhaps music, plosives in speech, or such extraneous sounds) will not be routed to the vibration system in question provided that bass-management is not used. The LFE channel, therefore, represents a channel of information which provides haptic systems with a viable source for motion generation.

That said, the results of this study indicate that coded-motion is globally preferred and perceived as being more realistic for action-oriented film content. Coded-motion is more likely to create a stronger suspension of disbelief within the observer. However, given that mean differences for ratings of sense of presence between coded-motion and LFE-derived motion were the smallest of the three examined factors, LFE-derived motion appears to evoke within the observer an adequate degree of presence (more than would otherwise be achieved without any vibration).

While the frequency content of motion codes is determined by the type of motion simulated, generally speaking, frequencies below approximately 20 Hz may be associated with environmental motion commonly found in everyday situations, such as motion experienced in a boat (caused by waves) or vehicle (caused by changes in acceleration), whereas, higher frequency motion is experienced as vibration or shaking. Depending on the frequency and waveform of the motion in question, humans appear to have a natural ability to discern certain environments simply through the sensation of motion. For example, during the physical measurements of coded-motion shown in Figure 2, the observer remarked that he could tell which scenes were water scenes, and which were not, even if audio and video were *not* present (simply from the motion codes). This observation is mostly likely due to the fact that the coded-motion was a three-dimensional motion program (as opposed to audio-generated motion which manifested itself in the platform in the form of heave only). Future publications include a psychophysical study of simpler stimuli which will attempt to verify whether or not this is indeed the case.

5 Conclusions

In the absence of codes designed by hand (custom motion programs), low-frequency audio can generate some appropriate motion especially when taken directly from an LFE channel. Although this audio-generated motion is typically not very satisfying, it does somewhat enhance an observer's sense of presence. Nonetheless, as a first pass for later careful selection and editing, audio-derived motion data can provide savings in terms of human labor costs when programming custom motion. Of course, human intervention in the form of amplification and elimination of audio-derived motion data will almost always produce superior results. The conclusions for this paper are thus :

- Custom motion programs most successfully create in observers a sense of realism related to a virtual environment and are generally preferred to motion generated automatically from audio programs for action-oriented DVD films.
- The existing information in LFE channels are generally useful for subtle enhancement of observer presence, however, custom motion programs are globally preferred, evoke a greater sense of observer presence, and a sense of realism which is not possible with audio-generated motion.
- Manufacturers and media houses might profit from investing in the creation of custom motion programs for home theater applications.

Acknowledgements

The authors would like to thank: Dr. Atsushi Marui and Mr. Sungyoung Kim for their input, The Schulich School of Music at McGill University, as well as Dr. Durand R. Begault from the Human Factors Research and Technology Division of the NASA Ames Research Center for his direction and 3-D measurements of the platform motion and vibration. This research was supported in part by Valorisation-Recherche Québec (www.vrq.qc.ca) and the Centre for Interdisciplinary Research in Music Media and Technology (www.music.mcgill.ca/cirmmt).

References

1. Begault, D.(2006). Haptic Perception and the Development of New Tools for Digital Studio Technology. Unpublished lecture given to McGill University.
2. Bouguila, L., Ishii, M., Sato, M. (2000). What Impact Does the Haptic-Stereo Intergration Have on Perception in Stereographic Virtual Environment? In: S. Brewster and R. Murray-Smith (Eds.) Haptic Human-Computer Interaction: First International Workshop, Glasgow UK, 135–150.

3. M. Daub and M. E. Altinsoy (2004). Audiotactile simultaneity perception of musical-produced whole-body vibrations. In: Proceedings of the Joint Congress CFA/DAGA.
4. Griffin, M.J. (1990). Handbook of Human Vibration. Elsevier Academic Press.
5. Holman, T. (1999). 5.1 Surround Sound Up and Running. Focal Press.
6. International Telecommunications Union (1994). Recommendation BS.775-1, Multichannel stereophonic sound systems with and without accompanying picture.
7. Martens, W. L., Woszczyk, W. (2003). Guidelines for Enhancing the Sense of Presence in Virtual Acoustic Environments. In: H. Thwaites (Ed.) Proceedings of the 9^{th} International Conference on Virtual Systems and Multimedia, 306–313.
8. McGee, M.R., Gray, P., Brewster, S. (2000). The Effective Combination of Haptic and Auditory Textural Information. In: S. Brewster and R. Murray-Smith (Eds.) Haptic Human-Computer Interaction: First International Workshop, Glasgow UK, 118–126.
9. Paillard, B., Roy, P., Vittecoq, P., Panneton, R. (2000). Odyssée: A new kinetic actuator for use in the home entertainment environment. In: Proceedings of DSPFest 2000, Texas Instruments, Houston, Texas.
10. Rumsey, F. (2001). Spatial Audio. Focal Press.
11. Ryan, T., Martens, W.L., Woszczyk, W. (2005). The effects of acoustical treatment on lateralization of low-frequency sources. In: Proc. of 149th Meeting of the Acoustical Society of America **117**, 2392.
12. Taylor, J., Johnson, M.R., Crawford, C.G. (2006). DVD Demystified. McGraw-Hill.
13. Terashima, N. (2002). Intelligent Communication Systems. Academic Press.
14. Todd, C. *et al.* (1994). Flexible perceptual coding for audio transmission and storage. In: Proceedings of the 96^{th} Convention of the Audio Engineering Society.
15. Walker, K., Martens, W.L (2005).Haptic-Auditory Distance Cues in Virtual Environments: Exploring Adjustable Parameters for Enhancing Experience. In: Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their applications (*HAVE 2005*), Ottawa, Canada, 101–103.
16. Walker, K., Martens, W.L., Kim, S. (2006). Perception of Simultaneity and Detection of Asynchrony Between Audio and Structural Vibration in Multimodal Music Reproduction. In: Proceedings of the 120^{th} Convention of the Audio Engineering Society, Paris, France.
17. Woszczyk, W., Cooperstock, J., Roston, J., Martens, W. (2005). Shake, Rattle, and Roll: Getting Immersed in Multisensory, Interactive Music via Broadband Networks. Journal of the Audio Engineering Society, **53**, No. 4, 336–344.
18. Wu, W.(1997). Development of Cueing Algorithm for the Control of Simulator Motion systems. MS Thesis, State University of New York at Binghamton.

Evaluating the Influence of Multimodal Feedback on Egocentric Selection Metaphors in Virtual Environments

Lode Vanacken, Chris Raymaekers, and Karin Coninx

Hasselt University, Expertise Centre for Digital Media (EDM)
and transnationale Universiteit Limburg
Wetenschapspark 2, B-3590 Diepenbeek, Belgium
`{lode.vanacken, chris.raymaekers, karin.coninx}@uhasselt.be`

Abstract. Whether a user interface is intuitive depends amongst others on (multimodal) feedback. The addition of multimodal feedback can certainly improve interaction in Virtual Environments as it increases the bandwidth to the user. One of the most common tasks in Virtual Environments is object selection. This paper elaborates on the enhancement of some existing approaches with multimodal feedback. The proposed techniques have been evaluated through a user experiment and the results show that the addition of multimodal feedback is preferred by the user and depending on the selection metaphor, it can also speed up the interaction.

1 Introduction

3D Virtual Environments (VEs) are slowly moving on from the design state to commercialization. They are for instance used to visualize or interact with data and to model environments. In nearly all Virtual Environment applications, the user needs to select objects using some input device combined with a selection metaphor. Several metaphors have been thoroughly investigated. Until now most efforts concentrated on the efficiency and satisfaction of the user.

Recently, researchers are also investigating other influences on selecting an object: the devices being used [1], the hand being used [2] or the type of setup (CAVE, HMD, ...) [3]. Less research has been done on trying to improve the selection techniques by adding extra modalities like haptics, sound or less common modalities such as taste and smell. The addition of an extra modality could improve the user's experience by enhancing the immersion and/or the performance of the user.

In this paper, we look at several ways to enhance selection metaphors with haptics and sound. We first discuss some selection metaphors on which we have based our experiment. Afterwards, we elaborate on the addition of multimodal feedback to these selection metaphors. To conclude, we discuss our experiment in which the influence of extra modalities are tested. Finally, we draw some conclusions and give some pointers for future research.

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 12–23, 2006.

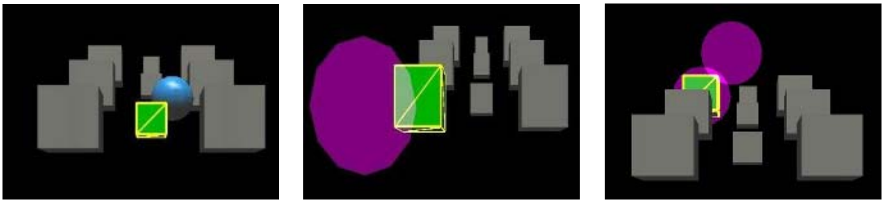

(a) Virtual Hand (b) Aperture (c) The Bubble Cursor

Fig. 1. The Selection Metaphors that were used in this research

2 Selection Metaphors

The selection of an object is one of the most common tasks in VEs, as an independent task as well as to start manipulation tasks. Due to the basic nature of this task, a lot of research has been done in order to find the "best" selection metaphor.

The selection metaphors investigated in this paper are limited to egocentric techniques [4] since interaction in VEs we investigate is done from a first person point-of-view. The reason for this assumption is the fact that our main research topic concerns modeling applications for Virtual Environments as these allow us to investigate interaction techniques which can be generalized for other VE applications. In previous work, several selection techniques have been compared using the dominant or the non-dominant hand [2]. Virtual hand and aperture showed to be the most promising techniques: virtual hand was preferred by most users, while aperture was the fastest technique.

Further, we elaborate on an interesting new technique which is adapted from 2D Graphical User Interfaces (GUIs), namely the bubble cursor [5]. This technique has not been tested in a 3D environment before and we would like to investigate whether this method might be an improvement of aperture.

2.1 Virtual Hand

The most simple and wide spread selection metaphor is called virtual hand. The user's hand position and orientation are virtually represented in the VE, but does not have to be represented as a real hand. In order to realize this, the user's movements are tracked and mapped to the virtual representation. In figure 1(a), one can see the virtual hand represented by a sphere. When the virtual representation intersects with an object, it can be selected immediately or it can be selected by pressing a button. The main disadvantage of this technique is the limited workspace, which is restricted to our own reach or that of the device. Solutions to this problem have been explored (eg. Go-Go [6]), but we assume that all objects are reachable or that the user can navigate to their vicinity.

2.2 Aperture

The aperture metaphor [7] can be seen as a flashlight with an invisible cone. Only the circle of light, the aperture, can be seen on the screen parallel to the projection plane (see Figure 1(b)), this is because the apex of the cone is in the user's eye point. When moving in the X or Y direction, the cone and aperture move along. When moving in the Z direction, the cone becomes smaller, making the aperture smaller as well. The object closest to the centerline of the cone becomes selected. For the user, this is the same as projecting the 3D view onto 2D and selecting the object that falls in the aperture.

When having multiple small objects close together, the aperture may have to decrease in size in order to disambiguate the selection. In a worst case scenario it has to be made as small as the smallest object. This is a disadvantage, as the movement along the Z axis costs extra time.

2.3 The Bubble Cursor

The bubble cursor originates from research into 2D area cursors. A single point cursor has a single point hotspot, whereas an area cursor has a much larger hotspot, the size of the area it encompasses. A 2D area cursor can have any shape: a square, a circle, an ellips, etc. A lot of research on area cursors has been guided by Fitt's Law [8, 9], which states that the movement time depends on the size of the targets and the distance to these targets. A detailed explanation of Fitt's Law is beyond the scope of this paper but more information can be found in [9].

The bubble cursor is represented by a circle which resizes dynamically and therefore selects only one target at a time. The cursor is also drawn semi-transparent and its size is adapted to the target it currently selects (see figure 1(c)). The details of the resizing algorithm are explained in [5]. In experiments, the bubble cursor appeared faster than a single point cursor in every situation, even when targets were very densely spaced.

When comparing aperture and the bubble cursor, we notice that both are based on the same principles, though coming from different backgrounds. It would be possible to design a bubble cursor in 3D resembling the Silk Cursor from Zhai et al. [10], which is a 3D semi-transparant area cursor with a rectangular volume. Due to the semi-transparency, depth cues are provided, but in Zhai's experiment there was only one target and the metaphor was only compared to a 3D wireframe equivalent area cursor. Informal testing in our lab shows that the 3D equivalent of the bubble cursor is not very useful, because it does not seem to improve depth cues sufficiently.

3 Adding Haptics and Sound to Selection Metaphors

Multimodal output during selection of objects has been thoroughly studied in 2D (GUIs) [8, 9]. In contrast, for VEs, the use of multimodal output has not yet been researched extensively. This section first elaborates on the possibilities of adding haptics[1] to selection and thereafter audio feedback will be discussed. Based on these discussions, section 4 introduces the different types of feedback that have been evaluated in our lab.

---

[1] We only concentrate on Force Feedback and disregard Tactile Feedback.

3.1 Adding Haptics to Selection

The addition of haptics to an application can improve the feeling of immersion and degree of intuition of the interaction to the user. The most common and widely spread haptic devices are those with one point haptic interaction, for example the PHANToM.

A lot of research concentrates on GUIs using haptics during selection. Oakley et al. [11] investigated the effects of different kinds of force feedback effects in 2D user interfaces: texture, friction, recess and gravity wells, which were overlaid on standard graphical buttons, and an haptically enhanced scroll bar. They concluded that the completion time was not reduced significantly, but the amount of errors made using gravity wells and recess effects did show a significant reduction. The texture and friction effect gave much worse results due to the fact that the user was not constrained to the target and the user's movements were perturbed. In their conclusion, gravity wells, a "snap-to" effect which pulls the user to the center of the target, seemed best in improving performance time and error reduction. Hwang et al. [12] also examined the use of gravity wells in 2D GUIs, for motion impaired users, but added distracter targets to enhance the realism of his experiment. These distracters were placed in all possible directions around the target. They found that when the distracters were in a straight line between the user and the target, problems arise and selection time diminishes. However, this was not proven to be due to oscillations between distracters.

In Virtual Environments, the work that has been done in introducing haptics for selection metaphors is rather limited. Wall et al. [13] investigated if the addition of haptic feedback, gravity wells and stereo graphics would improve selection of objects in 3D. The haptic feedback did improve the accuracy, but not the performance time, while stereo graphics improved both significantly. Magnusson et al. [14] asked users to test different sound and haptic feedback in a memory game. Gravity wells were among the conditions with the best results.

3.2 Adding Sound to Selection

A lot of research has been conducted concerning the addition of sound to GUIs and its advantages. Many different widgets (menus, buttons, scroll bars, progress bars,. . .) have been augmented with non speech sound, called Earcons, resulting in lower error rates and decreased task completion times [9].

For selection metaphors in GUIs, Akamatsu et al. [8] played a 2kHz tone while the cursor was over the target. They concluded that audio feedback did not improve overall selection time, but it did reduce the time spent over the target, thus sound made users react faster. Cockburn et al. [9] also played an earcon when the mouse was over the target, combined with a second earcon giving feedback when the target was successfully selected. They found that the addition reduced the mean selection time by 4.2%. Furthermore, when combining sound with other modalities, they discovered that a combination of two modalities, which improve selection time separately, does not assure the further improvement of selection time.

4 Adding Multimodal Feedback

4.1 Haptic Feedback

Based on the literature on haptic feedback during selection, gravity wells are chosen as a subject in our experiment. A problem with gravity wells is its many possible implementations. Therefore, a small experiment with five users is performed to determine the following parameters:

Forces: The attraction force can be consistent or degrading towards the center. The magnitude of the maximum force can also differ. Keuning et al. [15] concluded that users have different preferences regarding these parameters, but found that a higher force makes the users more precise. Therefore, it seems best to start with a high force and degrade towards the middle of the gravity well.

Size: The size of a gravity well is very important. When looking at Fitt's Law, the size influences the width thus the larger the gravity wells, the faster a user can select. On the other hand, larger fields overlap each other, causing oscillation problems and making it possible to get stuck on the wrong target.

Overlap: The easiest action to take when gravity wells overlap, is doing nothing at all. A second option is to activate the gravity well that is reached first. This can be enhanced with a simple prediction algorithm, where one of the overlapping gravity wells becomes active when we move 25% away from the closest ever distance to the center of the active gravity well. Another enhancement could be the recalculation of the force strength and size of the gravity well, using the overlapping percentage. Using half of the overlapping percentage seemed to give the best results.

Velocity: When incorporating the velocity at which the user is moving, we can disable the fields at a certain threshold, because the first movements during the selection process are at a high velocity [16]. Oakley et al. [17] already noticed that performance time improved when adjusting the force according to the velocity of the user.

The small experiment served as a pre-test for the final experiment, which incorporates another modality (sound) and other selection metaphors. In our small experiment, we use a strong force with force degradation to the center. We specify different sizes: 20%, 50% and 100% enlargement of the sphere bounding the object. For overlap techniques, we test both the prediction algorithm and the enhancement of the prediction algorithm, with the adjustment of sizes and forces. The gravity wells are disabled when the user reaches a speed of 2cm/s, which occurred frequently during informal testing. For the test setup, we use a PHANToM and 8 test-scenes in which 2-3 small and/or large boxes are present. The boxes sometimes serve as distracters, so that the largest gravity well sizes are not advantaged. When evaluating the results of our small experiment, no technique comes out on top, but looking at the fastest technique, we decided on using a 20% enlargement with no special recalculation of the size and strength of the gravity wells, leaving only the prediction algorithm.

In section 3.1 we discussed the addition of haptics for the virtual hand selection metaphor in GUIs and VEs, but in section 2 we also introduced two other metaphors. To our knowledge, no research has been done on combining haptics with any of these techniques, and neither for their equivalents in GUIs, the area cursors. Our research introduces two possibilities which try to emulate sticky icons on top of area cursors, as used by Worden et al. [18]. The first force feedback technique is a directional force in the opposite direction when moving over a target. Another is a viscous drag field with a negative constant, which is also activated when moving over a target. During informal testing the viscous drag field was preferred. Therefore the viscous drag field will be used in our final experiment which is discussed in section 5.1.

4.2 Audio Feedback

Most related work as described in section 3.2 makes use of earcons but different approaches are used. Based on these findings, we decided on the following: when reaching an object, an earcon sounds (F' on piano for 1s) to inform the user that a target is highlighted. If the user switches to another target, another earcon is played (C on celesta for 0.8s). That alerts the user for possible movement mistakes. When the user moves off a target, another earcon sounds (C on soprano sax for 1s). And finally, when selecting a target, an earcon is played (B on glocken for 1s) to confirm the selection of an object.

5 Evaluating Multimodal Feedback in VEs

We have presented several methods to add haptics and sound to selection metaphors. This section describes an experiment in which the selection metaphors from section 2 are enhanced with the above mentioned multimodal feedback. The goal of this experiment is to investigate the implications of their addition when selecting targets in a VE.

5.1 The Experiment

In the experiment three selection metaphors are used: virtual hand, aperture and the bubble cursor. Aperture is included in the experiment for comparison with the bubble cursor, since they are based on the same principle. Therefore, only the bubble cursor will be enhanced with modalities. For virtual hand we have two options regarding haptics: feeling an object (normal haptics) and the gravity wells proposed in section 4.1. As input device the PHANToM will be used. The audio feedback will use the approach proposed in 4.2. For the bubble cursor, the viscous drag field is used in combination with the same audio feedback as for virtual hand. Haptics and sound are added separately and in combination, which leaves us with eleven conditions: six conditions for virtual hand, one for aperture and four for the bubble cursor. In order to counterbalance these eleven conditions a Latin Square design is used.

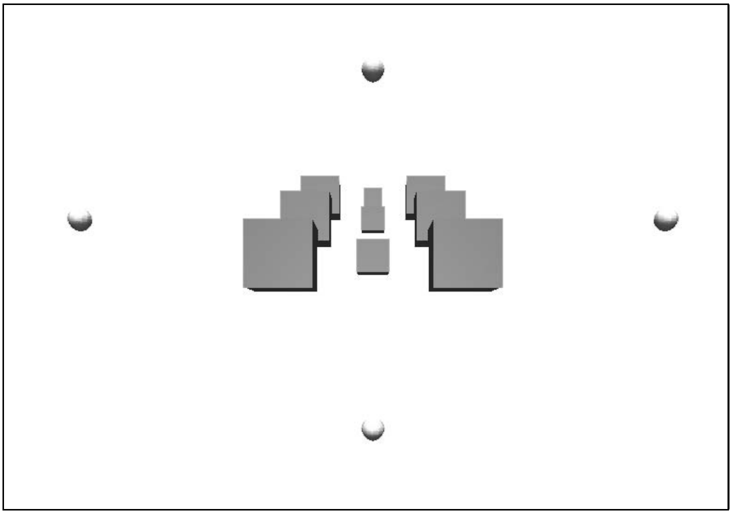


Fig. 2. The scene being used during the experiment. The figure was altered for a clearer view: the original background color was black.

Eleven users, nine male and two female, with a varying range of experience from none to expert participated. Their ages ranged from 21 to 30, with all of them being right handed.

The users were asked to select objects in a scene with four starting points defined, as can be seen in figure 2 (the spheres). The user is gently pulled towards one of these starting points, an earcon plays and the starting point disappears in order to let the user know this particular trial is started. The object that has to be selected colours green. After selecting an object, either correct or wrong, another starting point appears and another trail is started. Twenty setups, a start position and a block position, were defined and randomly given to the user for all eleven conditions. Data is logged after every selection including the total selection time, the time it took to highlight or reach the object and the time it took to finalize the selection by clicking. We also logged whether the correct object was selected and which trial had been performed. The first 5 trials for every condition were discarded, leaving 165 test results per user and 1815 test results for the complete experiment. After the test, the user was asked to fill in a questionnaire, querying his subjective impressions.

The main hypotheses of this experiment based on related work are the following: the bubble cursor performs the fastest (H1), adding audio or haptics speeds up the selection of an object (H2) and the combination of audio and haptics does not enhance the speed up furthermore during selection more significantly (H3).

5.2 Results

General Results. When looking at the selection metaphors and disregarding the combinations of modalities, it can be concluded that the bubble cursor is the fastest technique. The bubble cursor (BU) is significantly faster than aperture (AP) and virtual hand (VH) for selection and reach times. A single-factor ANOVA was conducted, resulting in $p < 0.001$ in every situation. Looking at the times it took to click, we would expect them to be equal but aperture was significantly slower than both the bubble cursor and virtual hand ($p < 0.001$). This can be explained due to the fact that users will move along the Z axis, even when already having the object highlighted. They want to be more certain that their selection is correct as the aperture often overlapped other objects.

Subjectively, almost all users, 8 out of 11, preferred the bubble cursor, 2 preferred aperture and one preferred virtual hand. We can conclude that, looking only at the selection metaphors, the bubble cursor is the fastest and most preferred technique of the three which is in confirmation with the first hypothesis (H1).

In figure 3, all the conditions are depicted, with their average selection, reach and click time. The order in which the techniques are fastest is clearly shown. We now discuss the combination of modalities separately for the bubble cursor and virtual hand, starting with virtual hand.

Virtual Hand: Modality Related Results. In the graph presenting all the conditions, virtual hand combined with gravity wells and sound is the fastest technique when looking at the average times. Taking statistical significance in account (p-values can be found in table 1), only the click time is significantly faster when comparing against the second fastest condition: virtual hand without haptics and with sound. The third fastest condition, virtual hand with normal haptics and sound, is not significantly slower than any of the two previous conditions for selection, reach and click time.

As soon as the addition of sound is removed, all conditions become significantly slower in their selection time, but not for all parts of the selection task.

Virtual hand with gravity wells and sound is significantly faster in selection and click time, compared to virtual hand with gravity wells and without sound. This means that the gravity wells only speed up the reach time, not the click time, or at least not significantly. The same conclusions hold when we compare virtual hand with gravity wells and without sound with virtual hand without haptics and sound, where the selection and reach time are significantly faster. Note that the combination of gravity wells and sound do not make the reach time significantly faster, although sound does improve the selection, reach and click time significantly when compared to virtual hand without sound.

These results are similar to those of Cockburn et al. [9] and according to our hypotheses, sound or haptics improves the selection time while the combination has no extra benefit. The users probably used the audio feedback to move faster, as they would hear a sound when reaching the object and could stop moving. This contradicts with the results of Akamatsu et al. [8], which concluded that only the click time would be reduced. Comparing both types of haptics, gravity wells are significantly faster. Considering the size of objects, we notice that all previous conclusions still hold except one: when selecting large objects, haptics have no influence on the speed. For small objects we notice that haptics do

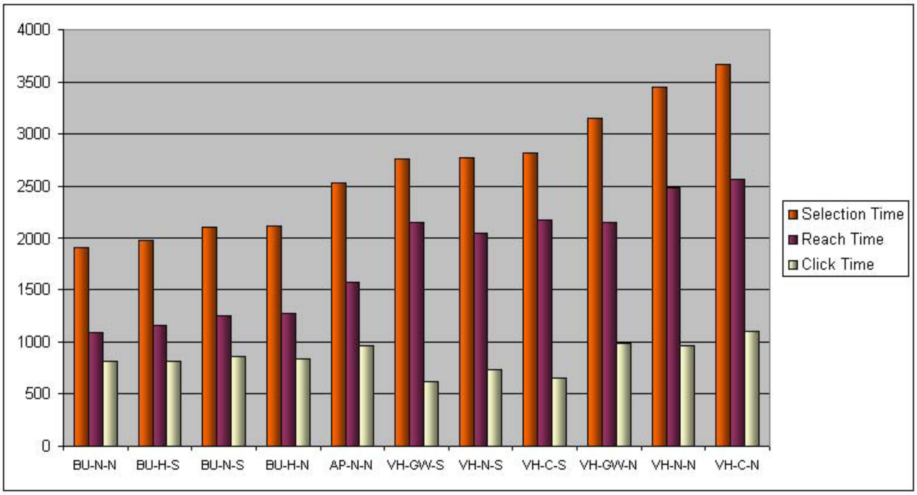


Fig. 3. Average selection, reach and click time for all conditions. BU, AP, VH stand for the respective selection metaphors. For the modalities with BU, haptics comes first and then sound. N stands for not present, H means haptics present and S means sound is present. For the modalities with VH, haptics comes first and then sound. N stands for not present, H means normal haptics, GW stands for gravity wells and S means sound is present.

matter. Gravity wells are significantly faster in selection and reach time than normal haptics, with $p < 0.015$ and $p < 0.025$ respectively.

The subjective results point out that users always prefer the addition of sound; we asked users which condition gave the best feedback for a certain selection metaphor. The addition of haptics is preferred 8 out of 11 times: four users preferred gravity wells, the other four preferred normal haptics.

The Bubble Cursor: Modality Related Results. The bubble cursor's results are somewhat counter-intuitive. On the one hand, the condition with no modalities at all was the fastest one, but it was not significantly faster than the bubble cursor with haptics and sound. On the other hand, the addition of sound or haptics on its own makes the bubble cursor significantly slower. This is a rather unexpected result; we expected both modalities on their own would improve the total selection time. The combination of modalities could be slower because of an overload of information, but since it is faster than the separate modalities, this is not the case. Further research on this result, which contradicts hypothesis H2 and H3, has to be done in order to draw meaningful conclusions.

We are not aware of related work that tested the bubble cursor or any other area cursor in combination with modalities, leaving us with nothing to compare against. Some clarification might be found in Fitt's Law. The bubble cursor already has a very large effective width, the width of the target which can be used to select it. This can be different than the visual width of the target. In

Table 1. Table comparing several conditions. Significance was tested using a single-factor ANOVA. A dash means that that particular modality was not taken into account.

Technique	Haptics	Sound	Reach (ms)	Reach	Click (ms)	Click	Total (ms)	Total
VH	GW	S	2151	.326	614	**<.001**	2766	.928
VH	N	S	2042		733		2776	
VH	GW	S	2151	.837	614	.131	2766	.629
VH	C	S	2176		649		2825	
VH	GW	S	2151	.943	614	**<.001**	2766	**<.003**
VH	GW	N	2042		733		2776	
VH	GW	N	2160	**<.02**	989	.5774	3149	**<.03**
VH	N	N	2483		967		3451	
VH	-	N	2123	**<.001**	666	**<.001**	2789	**<.001**
VH	-	S	2406		1019		3425	
VH	GW	-	2156	**<.03**	801	**<.02**	2958	**<.005**
VH	C	-	2373		874		3248	
BU	N	N	1086	.219	819	.809	1906	.311
BU	H	S	1158		811		1970	
BU	N	N	1086	**<.01**	819	.485	1906	**<.005**
BU	H	N	1278		839		2117	
BU	N	N	1086	**<.02**	819	.128	1906	**<.005**
BU	N	S	1247		866		2114	

case of the bubble cursor, it can select targets from a distance due to its resizing: this distance is the effective width. The addition of haptics, which slows down the movement over the target, increases the effective width even more. This addition could slow down the user because he already sees that he has reached the object and needs no slowing down. The same holds for sound: the adjusting of the bubble cursor might give enough feedback to the users. The combination of both haptics and sound could cancel the effect of too much feedback as the extra audio feedback could make the user less aware of him slowing down.

The subjective results for the bubble cursor are similar to the one's for virtual hand. All users except one prefer the audio feedback with 7 of them preferring the combination with haptics, leaving 3 users that prefer only audio feedback. The fact that the users prefer feedback, even though that the no feedback condition is the fastest, could be noted to the following: when moving the bubble cursor, the target being selected changes fast and the extra feedback could make the users more comfortable and aware of what is happening exactly, but also makes them slower.

Besides the logged timings, the amount of mistakes made during selection is an important factor. One particular trial, with the block on the middle left and the starting position on the right, had a significant amount of additional errors compared to the other trials. Therefore we removed this trial from the results. The virtual hand selection metaphor caused between 0 and 2.5% errors, aperture 5.5% and the bubble cursor between 3 and 5%. When conducting a chi square test, we noticed that no significance is present.

6 Conclusions and Future Work

This paper presents several selection metaphors for VEs and number of methods to combine them with haptics and/or sound, which are already successfully used in GUIs. Keeping these methods in mind, several combinations of audio and haptic feedback are implemented and tested in a user experiment.

This experiment shows that the addition of haptics and/or sound for the virtual hand metaphor has similar effects as in GUIs. The addition of one modality separately results in a decrease of the total selection time, but the addition of both haptics and sound does not decrease the selection time.

The results for the bubble cursor are in contradiction to those of the virtual hand or single point cursor GUI research. The addition of haptics or sound increases the total selection time. Contrary, the addition of both modalities speeds up the technique, but this is still slower than having only visual feedback.

Overall the bubble cursor had the best performance during our experiment. However, as little research into area cursors and multimodality has been conducted, it is possible that other combinations of multimodal feedback are beneficial for this selection metaphor. For instance an attraction force can be used in order to not only visually adapt the cursor but also haptically. In our current implementation it is possible that the user experiences a sensory conflict, however further research is needed to investigate this.

Acknowledgments

Part of the research at EDM is funded by ERDF (European Regional Development Fund), the Flemish Government and the Flemish Interdisciplinary institute for Broadband technology (IBBT).

References

1. Steed, A., Parker, C.: 3d selection strategies for head tracked and non-head tracked operation of spatially immersive displays. In: 8th International Immersive Projection Technology Workshop, Ames, Iowa, USA (2004)
2. De Boeck, J., De Weyer, T., Raymaekers, C., Coninx, K.: Using the non-dominant hand for selection in 3D. In: Proceedings of the first IEEE Symposium on 3D User Interfaces 2006, Alexandria, Virginia, US (2006)
3. Manek, D.: Effects of visual displays on 3d interaction in virtual environments. Master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA (2004)
4. Pouprey, I., Weghorst, S., Billunghurst, M., Ichikawa, T.: Egocentric object manipulation in virtual environments; empirical evalutaion of interaction techniques. Computer Graphics Forum **17** (1998) 30–41
5. Grossman, T., Balakrishnan, R.: The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In: CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press (2005) 281–290

6. Poupyrev, I., Billinghurst, M., Weghorst, S., Ichikawa, T.: The go-go interaction technique: non-linear mapping for direct manipulation in vr. In: Proceedings of the ACM Symposium on User Interface Software and Technology (UIST) 1996, Seattle, Washington, USA (1996)
7. Forsberg, A., Herndon, K., Zeleznik, R.: Aperture based selection for immersive virtual environment. In: Proceedings of UIST96. (1996) 95–96
8. Akamatsu, M., MacKenzie, I.S., Hasbrouc, T.: A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. Ergonomics **38** (1995) 816–827
9. Cockburn, A., Brewster, S.: Multimodal feedback for the acquisition of small targets. Ergonomics **48** (2005) 1129–1150
10. Zhai, S., Buxton, W., Milgram, P.: The ”silk cursor”: investigating. In: CHI ’94: Conference companion on Human factors in computing systems, New York, NY, USA, ACM Press (1994) 459–464
11. Oakley, I., McGee, M.R., Brewster, S., Gray, P.: Putting the feel in ’look and feel’. In: Proceedings of CHI 2000, The Hague, NL (2000) 415–422
12. Hwang, F., Langdon, P., Keates, S., Clarkson, J.: The effect of multiple haptic distractors on the performance of motion-impaired users. In: Proceedings of Eurohaptics 2003, Dublin, IE (2003) 14–25
13. Wall, S., Paynter, K., Shillito, M., Wright, M., Scali, S.: The effect of haptic feedback and stereo graphics in a 3d target acquisition task. In: Proceedings of Eurohaptics 2002, Edinburgh, UK (2002)
14. Magnusson, C., Rassmus-Gröhn, K.: Audio haptic tools for navigation in non visual environments. In: ENACTIVE 2005, the 2nd International Conference on Enactive Interfaces. (2005)
15. Keuning, H.: Augmented force feedback to facilitate target acquisition in Human-Computer Interaction. PhD thesis, Technical University of Eindhoven, Eindhoven, The Netherlands (2003)
16. McGuffin, M., Balakrishnan, R.: Acquisition of expanding targets. In: Proceedings of ACM Conference on Human Factors in Computing Systems (CHI) 2002, ACM Press (2002) 57–64
17. Oakley, I., Brewster, S., Gray, P.: Solving multi-target haptic problems in menu interaction. In: In Extended Abstracts of ACM CHI 2001, Seattle, WA, USA (2001) 357–358
18. Worden, A., Walker, N., Bharat, K., Hudson, S.: Making computers easier for older adults to use: Area cursors and sticky icons. In: Proceedings of CHI97: ACM Conference on Human Factors in Computer Systems, Atlanta, Georgia, USA (1997)

Haptic-Auditory Rendering and Perception of Contact Stiffness

Federico Avanzini and Paolo Crosato

Dep. of Information Engineering, University of Padova, Italy
{avanzini, crosatop}@dei.unipd.it

Abstract. This paper presents an experiment on the relative contributions of haptic and auditory information to bimodal judgments of contact stiffness using a rigid probe. Haptic feedback is rendered via a Phantom® Omni™ device, while auditory stimuli are obtained using a physically-based audio model of impact, in which the colliding objects are described as modal resonators that interact through a non-linear impact force. The impact force can be controlled through a stiffness parameter, that influences the contact time of the impact. Previous studies have already indicated that this parameter has a major influence on the auditory perception of hardness/stiffness. In the experiment subjects had to tap on virtual surfaces, and were presented with audio-haptic feedback. In each condition the haptic stiffness had the same value while the acoustic stiffness was varied. Perceived stiffness was determined using an absolute magnitude-estimation procedure: subjects were asked to rate the surfaces on an ordered scale of verbal labels, based on their perceived stiffness. The results indicate that subjects consistently ranked the surfaces according to the auditory stimuli.

1 Introduction

The importance of multimodal feedback in computer graphics and interaction has been recognized for a long time [1] and is motivated by our daily interaction with the world. Streams of information coming from different channels complement and integrate each other, with some modality possibly dominating over the remaining ones, depending on the task [2, 3]. Research in ecological acoustics [4, 5] demonstrates that auditory feedback in particular can effectively convey information about a number of attributes of vibrating objects, such as material, shape, size, and so on.

Recent literature has shown that sound synthesis techniques based on physical models of sound generation mechanisms allow for high quality synthesis and for a high degree of interactivity, since the physical parameters of the sound models can be naturally controlled by the gestures and the actions of a user. Sounds from solids are especially interesting since auditory cues frequently occur when we touch or interact with objects. Sound models for impulsive and continuous contact have been proposed e.g. in [6, 7]. Physically-based sound models of contact have been shown to be effective in conveying information about e.g. material

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 24–35, 2006.

properties [8], and have been applied in [9] to the development of an audio-haptic interface for contact interactions.

Bi-modal perception in continuous contact interaction (i.e., scraping or sliding) has been studied by many authors. In a classic work Lederman [10] compared the effectiveness of tactile and auditory information in judging surface roughness, and showed that when both were present the tactile one played the strongest role in determining experimental performance. More recent research by Lederman et al. [11] has focused on bi-modal roughness perception when the surface is explored using a rigid probe rather than with the bare skin, and *vibratory* roughness perception occurs. The results showed that, although tactual dominance is still found, sound plays a more relevant role when using a probe than in the case of direct contact with bare fingers. Guest et al. [12] have also focused on audio-tactile interactions in roughness perception. In their experimental setup, participants were required to make forced-choice discrimination responses regarding the roughness of abrasive surfaces which they touched briefly. Texture sounds were captured by a microphone located close to the manipulated surface and subsequently filtered in various ways before being presented to the participants. The authors investigated how the filtering biased the subjects' judgements. McGee et al. [13] studied bi-modal perception of *virtual* roughness, i.e. roughness of synthetic haptic and auditory textures. The latter were synthesized from the same sinusoidal waveforms used to describe the profiles of the haptic textures, and therefore did not provide a veridical feedback. Nonetheless, experimental results indicated that the presence of auditory feedback affected the likelihood that different textures were successfully judged as different.

Bi-modal perception in impulsive contact (i.e., impact) is apparently less studied. DiFranco et al. [14] studied the effect of auditory feedback on haptic stiffness perception, through headphone reproduction of prerecorded contact sounds between several pairs of objects. Experimental results showed that contact sounds influenced the perception of object stiffness. However the sounds used in [14] were chosen on a purely subjective basis rather than on an analysis of what timbral dimensions are mostly related to auditory perception of contact stiffness. Useful indications about the auditory cues that are most relevant to stiffness/hardness perception come from studies in ecological acoustics [15, 16].

This paper investigates the effectiveness of synthetic impact sounds in modulating the haptic perception of stiffness experienced by a user. In Sect. 2 we present the sound physical model used in the remainder of the paper, and describe how the sound model is integrated into an architecture for audio-haptic rendering. Section 3 reports upon an experiment on bi-modal stiffness perception that makes use of this architecture. Results are discussed in Sect. 4.

2 Impact Sounds

2.1 A Physically-Based Sound Model

When a generic solid object engages in some external interactions (e.g. it is struck, scraped, and so on), the forces at the contact point cause deformations to propagate through the body, and consequently its surfaces to vibrate and emit sound waves. A physically-motivated model for the simulation of vibrating objects is modal synthesis [17, 18], which describes the object as bank of second order damped mechanical oscillators (the *normal modes*) excited by the interaction force. The frequencies and dampings of the oscillators depend on the geometry and the material of the object and the amount of energy transferred to each mode depends on the location of the force applied to the object. Under general hypothesis, and with appropriate boundary conditions, linear partial differential equations describing a vibrating system admit solutions described as superposition of vibration modes. In this sense modal synthesis is physically well motivated and widely applicable. Techniques based on modal synthesis have been exploited by many authors for real-time synthesis of realistic sound effects for interactive simulations (see e.g. [7, 19]).

We have developed a physically-based sound synthesis model of interacting objects, simulated through a modal description. The objects can be coupled through non-linear interaction forces that describe impulsive and continuous contact. While the models described in[7] are linear and based on feed-forward computation, our force models are non-linear and dynamic. These features allow improved interactivity and better quality, at the expense of higher computation loads. In this work we make use of a real-time implementation of the model, realized as a plugin to the open source real-time synthesis environment `pd` (Pure Data[1]). The full model and the implementation details are presented in [6]. Here we only discuss the impact force model.

The audio impact force model [20] is based on an extension of the Hertz theory of normal collision between elastic bodies [21]:

$$f_{\mathrm{A}}(x(t), v(t)) = \begin{cases} k_{\mathrm{A}} x(t)^{\alpha} + \lambda_{\mathrm{A}} x(t)^{\alpha} \cdot v(t) & x > 0, \\ 0 & x \leq 0, \end{cases} \tag{1}$$

where the compression x at the contact point is the difference between the displacements of the two bodies, and $v(t) = \dot{x}(t)$ is the compression velocity. The condition $x > 0$ states that there is actual compression, while the complementary condition says that the two objects are not in contact. The force model (1) includes both an elastic component $k_{\mathrm{A}} x^{\alpha}$ and a dissipative term $\lambda_{\mathrm{A}} x^{\alpha} v$. The latter accounts for viscoelastic losses during collision. The parameter k_{A} in (1) is the force *stiffness* and is in general a function of the mechanical properties of the two bodies, while λ_{A} is the force *damping weight.* Additionally a variable exponent α is introduced, whose value depends on the surface geometry of the contact (e.g., $\alpha = 3/2$ for the particular case of contacting spheres).

2.2 Force Stiffness, Contact Time, Spectral Centroid

In previous studies we have investigated the influence of the impact force parameters on the spectral centroid of the sound attack transient, and on the duration τ

---

[1] http://crca.ucsd.edu/~msp/

of the contact between the two objects during the stroke. In particular a power-law dependence of the contact time τ on the force stiffness was found [22]: $\tau(k_\mathrm{A}) \sim k_\mathrm{A}^{-1/\alpha+1}$. A study in [23] on synthetic impact sounds obtained from model (1) provided quantitative results that show a strong correlation between the spectral centroid of the attack transients and the contact time.

The spectral centroid of the attack transient is known to influence the auditory perception of stiffness. Freed [15] has investigated the abilitiy of listening subjects to estimate the hardness of hammers made of various materials, from the sound that they generated when striking metallic pans of varying sizes. His experiments showed that the useful information for mallet hardness rating is contained in the attack transients of the sounds, namely in the first 300 ms of the signals. Loudness and descriptors related to the spectral centroid (average value and temporal variability in the first 300 ms) were used as predictors in a multiple regression analysis, and were found to account for 75% of the variance of the hardness ratings.

Giordano [16] has also investigated auditory perception of collision hardness. He argues that the contact time τ has an influence on hardness perception, and that τ variations are likely to explain at least in part data from [15]. Specifically, an increase in τ determines a decrease in the loudness of the radiated signal, and in the amount of energy at high frequencies (and thus in the spectral centroid), since vibrational modes with a period higher than τ are minimally excited.

In summary, the studies reported in [22, 23] have shown that manipulation of the impact force parameters k_A affects in a predictable way the contact time and the average spectral centroid during the attack transient. These parameters in turn have a major influence on the perception of impact hardness. Examples of these effects are provided in Fig. 1.

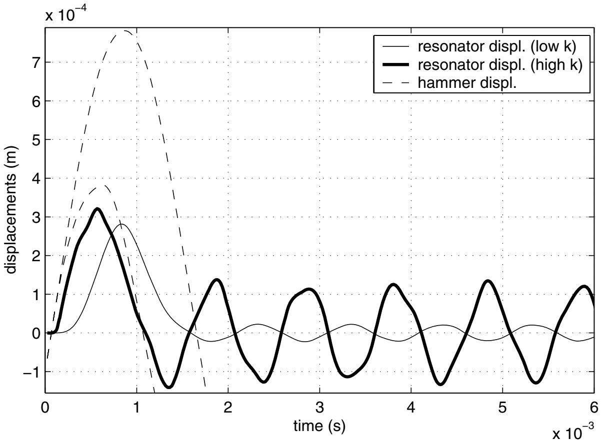


Fig. 1. Examples of transient attacks obtained from the impact model: short vs. long initial bumps, obtained by varying the force stiffness k_A

2.3 Audio-Haptic Rendering

The software experimental setup is composed of two processes which exchange information through a shared memory area (see Fig. 2). The first process renders graphics and the haptic freedback, and has been programmed with the Openhaptics$^{\mathrm{TM}}$ Toolkit developed by Sensable. An event catching engine driven by a function callback model is adopted to monitor contact events. When such an event occurs, data needed for sound synthesis is copied on the shared memory area. The second process renders contact sounds according to the current physical/geometrical parameters read from the shared memory area, and has been programmed with `pd`.

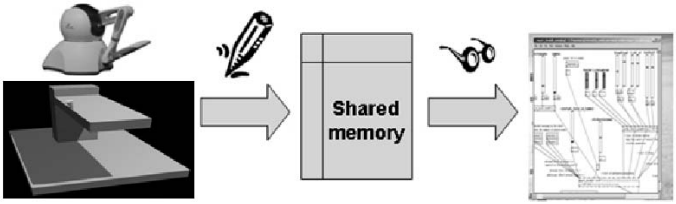


Fig. 2. The software architecture of the experimental setup

In order to achieve a realistic degree of interaction and unitary perception, the latency between the haptic, audio and visual feedback has to be very low. The comunication interface introduces some delay due to read/write access to shared memory. The code was heavily optimized so that the delay introduced by this process is negligible. We made many simulations of cyclic write/read access patterns, and found that in the worst case the delay introduced was in the order of μs, thus being negligible if compared to the latency due to haptic, sound and graphic rendering, which is in the order of some ms. During our experimental tests no subjects perceived any kind of noticeable intermodal latency.

Simulation of surface interaction in haptics is generally based on simple linear stiffness models [24]: a rigid immobile surface is modeled as a viscoelastic element, with a haptic stiffness k_{H} and a haptic viscosity coefficient λ_{H}, such that the haptic contact force is given as $f_{\mathrm{H}}(x(t), v(t)) = k_{\mathrm{H}} x(t) + \lambda_{\mathrm{H}} v(t)$, where x is the normal displacement relative to the surface. An ideally rigid wall should be simulated with k_{H} as high as possible. However limitations in the haptic sampling period T_{H} (typically $\sim$ 1 kHz), and the spatial resolution of the device, limit the range for k_{H}: using values that are too high can cause the system to become unstable, i.e., to oscillate uncontrollably. Sufficient conditions for the stability of the interaction can be found by requiring the system to be *passive* [25]. In this work we have used this linear physical model haptic rendering of stiffness, since it is the one implemented in the Openhaptics$^{\mathrm{TM}}$ Toolkit provided with the Phantom® device. This implies that k_{A} and k_{H} have different absolute values, because the physical models used for haptic and audio rendering are different.

3 Bi-modal Stiffness Perception

The architecture described in the previous section has been used to experimentally assess relative contributions of haptic and auditory information to bimodal judgments of contact stiffness using a rigid probe. More specifically, the experiment described in the remainder of this section is intended to assess the effectiveness of auditory feedback in modulating haptic perception of stiffness.

3.1 Participants

Sixteen subjects (between 19 and 30 years old) participated in the experiment. All participants reported themselves as being right-handed, and as having both normal hearing and normal tactual/motoric capabilities in their hands. All of them were naive as to the purposes and hypotheses of the test, and all of them volunteered. Some participants were musically trained.

3.2 Stimuli

The graphic display provided to subjects is shown in Fig. 3 (left). The small cone represents the position of the stylus. In every condition the haptic stiffness had the same value $k_{\mathrm{H}} = 400\,\mathrm{N/m}$. According to literature (see e.g. [26]) this can be considered an average value, with "soft" values being below $300\,\mathrm{N/m}$ and "hard" values starting above $600 - 700\,\mathrm{N/m}$. With this choice the haptic perception of stiffness is likely to be ambigous, and subjects are encouraged to rely on auditory judgement.

Auditory stiffness levels where obtained by varying the parameter k_{A}, while all the remaining parameters of the physical sound model were held constant. The fundamental frequency of the struck object and the modal frequency distribution were chosen based on the equations for the ideal bar: with length $L = 20\,\mathrm{cm}$, height $h = 1\,\mathrm{cm}$, density $\rho \sim 1 \cdot 10^3\,\mathrm{Kg/m^3}$, and Young's modulus $E \sim 3 \cdot 10^{10}\,\mathrm{N/m^2}$ (in between typical values for wood and glass), the fundamental frequency is $f_0 = \frac{\pi h}{8\sqrt{12}L^2}\sqrt{\frac{E}{\rho}} \cdot 1.194^2 \sim 220\,\mathrm{Hz}$, and the modal distribution is given as $f_0 \cdot [1,\ 6.27,\ 17.54,\ 34.37,\ 56.81,\ 84.87, \ldots]$. With this choice of values the sixth modal frequency is close to the upper limit of the range of human hearing, therefore the first five modes were simulated.

The modal decay times were also chosen to match intermediate values between wood and glass. Impact force parameters other than k_{A} (see (1)) were also held constant. Given this set of parameter values, the interval of variability $[k_{\min}, k_{\max}] = [1 \cdot 10^3, 6.4 \cdot 10^5]\,\mathrm{N/m}^{\alpha}$ for the stiffness k_{A} was determined empirically as the largest interval outside of which further stiffness variations do not produce noticeable effects in the physical model behavior. Finally a series of exponentially spaced values $k_i = 2^i \cdot k_{\min}$ was sampled within this interval, resulting in a set of seven auditory stiffness values.

3.3 Procedure

Subjects were seated in front of a 15 in. wide computer monitor. The Phantom® Omni$^{\mathrm{TM}}$ device was placed on their right-hand side, while a computer mouse was

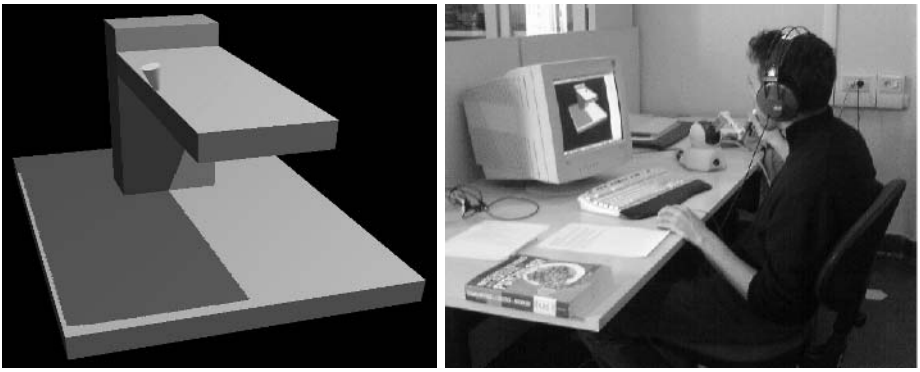


Fig. 3. Left: interactive graphic display presented to the subjects (the small cone represents the tip of the Phantom® stylus). Right: experimental setup.

placed on the left side. Auditory feedback was presented through headphones, connected to the output of a dedicated sound card. A picture of the experimental setup is provided in Fig. 3 (right).

Subjects were presented with the display depicted in Fig. 3 (left) and were instructed to judge the stiffness of the impact between a “hammer”, represented by the device stylus, and the bar of the graphical display. Every object in the scene could be felt through the haptic device, but only touching the upper bar produced a sound. The graphic display did not change between conditions, and was intentionally composed of stylized objects, in order to limit as much as possible the amount of visual information delivered to subjects.

Perceived stiffness was determined through an absolute magnitude-estimation procedure (similarly to the procedure reported in [11]): participants were instructed to assign the non-zero, positive number that best described the magnitude of the perceived stiffness of the stimulus, along a scale ranging from 1 to 8. Verbal labels were associated to each point of the scale, ranging from “extremely soft” (1) to “extremely stiff” (8).

Participants did not receive any training before the experiment. During the experiment, auditory feedback conditions were presented with the following internal organization (not known to the subjects): first the seven stiffness levels were presented once each, then they were presented again three times each, and the 21 (level × repetition) combinations were randomized. In this way the first seven conditions provided participants with a minimal hidden training phase. The random order was different for each subject. Participants were allowed to interact with each condition as long as desired. Finally, in a post-experimental interview, subjects were asked the multiple-choice questions reported in Table 1.

3.4 Discussion

The experiment presented here has some similarities with the study conducted in [14]. However, there are some noticeable difference as well, specifically in the

Table 1. Post-experimental interview

1. In your opionion what was varying between each condition?
   [*haptics* | *audio* | *haptics and audio*]
2. In order to express your judgements, you relied mainly on...
   [*haptics* | *audio* | *haptics and audio*]
3. In your opinion the conditions simulate changes in the stiffness of...
   [*the bar* | *the hammer* | *both*]
4. In your opinion changes in the stiffness are due to changes of...
   [*bar material* | *hammer mat.* | *mat. of both* | *hammer shape* | *hamm. mat.&shape*]
5. Did the visual display influence your judgements? If yes, how?
   [*yes* | *no*]

design of the auditory feedback. Sounds for the experiments reported in [14] were real impact sounds, recorded by tapping various tools (the authors mention a pen and a screwdriver) against surfaces of various materials (styrofoam, metal plate, and so on). In the setup, a contact detection event in the haptic rendering pipeline triggered the playing of one of the sound files.

We argue that using recorded sounds as those in [14] has two main drawbacks. First, auditory stimuli produced by such a wide variety of interacting object are likely to be very easily discriminated, and may allow subjects to perform an *identification* task rather than a *rating* task. Second, the recorded sounds were obtained by varying not only the material of the striker, but also that of the struck object. Several studies (see e.g. [16]) support the hypothesis of a strong link between impact stiffness perception and material perception.

The synthetic stimuli used in this study differ only in the values of k_{A}, while the modal parameters associated to the struck object are constant. Therefore the two perceptual dimensions of impact stiffness and material of the struck object are decoupled. Moreover, as described above, the modal parameters of the struck object were chosen to lie between values typical for wood and glass, in order not to provide a clear perception of material to the subjects. As a result, the auditory cues associated to variations in stiffness are very subtle.

4 Results

4.1 Stiffness Scaling

After a preliminary analysis of collected data, three subjects were classified as "outliers" and discarded from the results. One of the outlier misunderstood the meaning of the perceptual scale, thus giving inconsistent answers. All the outliers provided contradictory answers to questions 1 and 2 of the post-experimental interview, confirming that they were very confused by the contrasting auditory and haptic cues received when strucking the object, with sounds that were not "appropriate" to the haptic sensation. One of the outlier reported troubles in assigning stiffness values because the perceptual scale was too sparse.

The hidden training phase (the first seven conditions of each series) was not included in the data analysis. Magnitude estimates were extracted using the following procedure: first, for each subject, estimates were averaged across stimulus repetitions; then, in order to compensate for differences in individual scales, the averaged estimates for each subject were normalized by dividing each score by the individual participant mean and multiplying by the grand mean (across participants). This procedure is resemblant of that used in [11].

One-way ANOVA was conducted on the mean magnitude estimates, and the effect of the auditory stiffness level was found to be statistically significant ($F = 122.87$, $p < 0.001$). A boxplot of the data is presented in Fig. 4. On average subjects identified the increase in stiffness with good accuracy, especially in the range 2000 – 32000 N/m. Near the extremal values the judgements are clearly less accurate. In particular the 1000 N/m value is on average perceived as stiffer than the next one. The boxplot shows that the range of responses for these two levels are very wide, confirming that subjects had difficulties in identifying the stimulus. Two mild outliers are plotted on the fourth column (8000 N/m). Note however that the interquartile range for this condition is extremely narrow, justifying to some extent the presence of these two mild outliers.

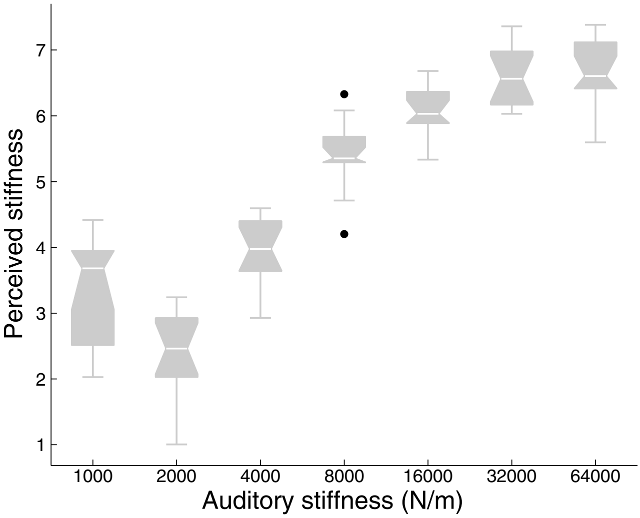


Fig. 4. Boxplots of perceived stiffness judgements (normalized magnitude estimates). For each box the central horizontal line represents the median, the top and the bottom represent the upper and lower quartiles, the vertical lines enclose the range of data.

4.2 Post-experimental Interview

The bar-plot in Fig. 5 shows the results of the post-experimental interview (see also Table 1). Question 1 and 2 were clearly related, although the first was mostly concerned with perception while the second asked about the strategy adopted by the subject in the rating task. Every subject's judgement was influenced, at least partially, by sound, but remarkably 5 out of 13 subjects perceived the haptic feedback changing together with audio and based their rating also on haptic feedback (although the haptic stiffness had the same value in all conditions, as explained in Sect. 3).

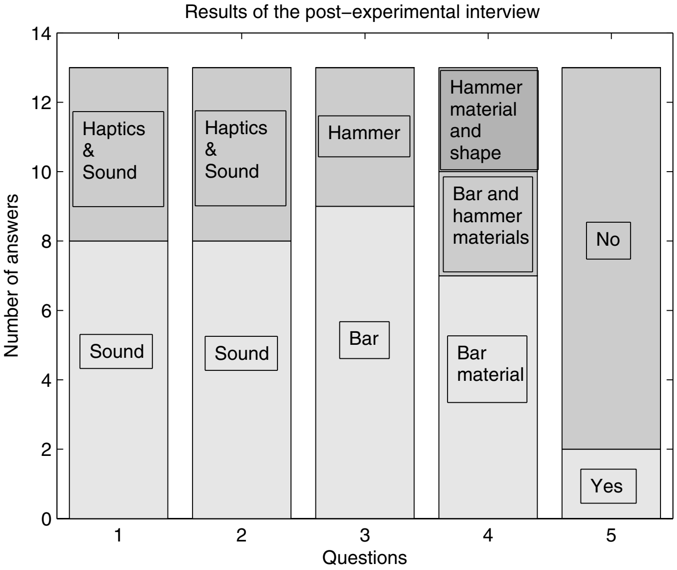


Fig. 5. Answers to the post–experimental interview. Each number corresponds to one question. The sixth question is discarded, since it was an open one.

Question 3 asked about which of the two objects underwent changes in stiffness. The "correct" answers would have been "both", since the stiffness value of the impact force relates to both the bar and the hammer properties. On the contrary, most of the subjects related different conditions to changes in bar stiffness, and a smaller percentage to changes in the hammer stiffness. This bias is in accordance with the findings by Giordano [16], although the graphic display may also have a role since the cone representing the hammer is a less veridical depiction that a parallelepiped for a bar. Since subjects are used to think about a cursor as a completely abstract representation of a position on a desktop, they may have associated the cone to a mouse cursor, and implicitly refused to give it a physical meaning.

This impression is confirmed by answers to question 4, since most of the subjects associated stiffness variations to changes in the material of the bar. The small fraction of subjects that related the change of stiffness to the hammer probably noticed that the acoustic properties of the bar did not change. In summary, listeners showed a somewhat limited ability to discriminate acoustically between hammer and struck object: this finding is compatible with the results reported by Giordano [16].

One of the strongest assumption in the experimental procedure was that the graphical display did not affect judgements: the answers to the fifth question clearly support this assumption. The two subjects that reported an influence of the graphical display commented that the graphics evoked some kind of hard material, like steel or thick wood.

5 Conclusion

The findings from the experiment reported in this paper support the effectiveness of auditory feedback in modulating haptic perception of stiffness. Magnitude estimates by the subjects provide clear indication that the perceived stiffness scales consistently with the physical parameter k_{A} which is varied in the auditory stimuli. Interestingly, a relevant portion (about 40%) of the subjects remarked in their answers to the post-experimental interview that they perceived variations in the haptic stiffness, although in every experimental condition the haptic stiffness had the same value.

The results suggest that auditory cues can be successfully used to augment and modulate the haptic display of stiffness, especially when the characteristics of the haptic system and the spatial resolution of the device, limit the range for surface stiffness rendering.

Acknowledgment. This research was supported by the EU Sixth Framework Programme – IST Information Society Technologies (Network of Excellence "Enactive Interfaces" IST-1-002114, http://www.enactivenetwork.org).

References

1. Hahn, J.K., Fouad, H., Gritz, L., Lee, J.W.: Integrating sounds in virtual environments. Presence: Teleoperators and Virtual Environment **7**(1) (1998) 67–77
2. Welch, R.B., Warren, D.H.: Intersensory interactions. In Boff, K.R., Kaufman, L., Thomas, J.P., eds.: Handbook of Perception and Human Performance – Volume 1: Sensory processes and perception. John Wiley & Sons, New York (1986) 1–36
3. Ernst, M.O., Bülthoff, H.H.: Merging the senses into a robust percept. TRENDS in Cognitive Sciences **8**(4) (2004) 162–169
4. Gaver, W.W.: What in the world do we hear? an ecological approach to auditory event perception. Ecological Psychology **5**(1) (1993) 1–29
5. Gaver, W.W.: How do we hear in the world? explorations of ecological acoustics. Ecological Psychology **5**(4) (1993) 285–313
6. Avanzini, F., Rath, M., Rocchesso, D., Ottaviani, L.: Low-level sound models: resonators, interactions, surface textures. In Rocchesso, D., Fontana, F., eds.: The Sounding Object. Mondo Estremo, Firenze (2003) 137–172
7. van den Doel, K., Kry, P.G., Pai, D.K.: Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In: Proc. ACM SIGGRAPH 2001, Los Angeles, CA (2001) 537–544
8. Klatzky, R.L., Pai, D.K., Krotkov, E.P.: Perception of material from contact sounds. Presence: Teleoperators and Virtual Environment **9**(4) (2000) 399–410

9. DiFilippo, D., Pai, D.K.: The AHI: An audio and haptic interface for contact interactions. In: Proc. ACM Symp. on User Interface Software and Technology (UIST'00), San Diego, CA (2000)
10. Lederman, S.J.: Auditory texture perception. Perception **8**(1) (1979) 93–103
11. Lederman, S.J., Klatzky, R.L., Morgan, T., Hamilton, C.: Integrating multimodal information about surface texture via a probe: Relative contribution of haptic and touch-produced sound sources. In: Proc. IEEE Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2002), Orlando, FL (2002) 97–104
12. Guest, S., Catmur, C., Lloyd, D., Spence, C.: Audiotactile interactions in roughness perception. Exp. Brain Research **146**(2) (2002) 161–171
13. McGee, M.R., Gray, P., Brewster, S.: Mixed feelings: Multimodal perception of virtual roughness. In: Proc. Int. Conf. EuroHaptics, Edinburgh (2002) 47–52
14. DiFranco, D.E., Beauregard, G.L., Srinivasan, M.A.: The effect of auditory cues on the haptic perception of stiffness in virtual environments. Proceedings of the ASME Dynamic Systems and Control Division, (DSC-Vol.61) (1997)
15. Freed, D.J.: Auditory correlates of perceived mallet hardness for a set of recorded percussive events. J. Acoust. Soc. Am. **87**(1) (1990) 311–322
16. Giordano, B.: Sound source perception in impact sounds. PhD thesis, Department of General Psychology, University of Padova, Italy (2006)
17. Adrien, J.M.: The missing link: Modal synthesis. In De Poli, G., Piccialli, A., Roads, C., eds.: Representations of Musical Signals. MIT Press, Cambridge, MA (1991) 269–297
18. van den Doel, K., Pai, D.K.: Modal Synthesis for Vibrating Objects. In Greenebaum, K., ed.: Audio Anecdotes. AK Peters, Natick, MA (2004)
19. O'Brien, J.F., Shen, C., Gatchalian, C.M.: Synthesizing sounds from rigid-body simulations. In: Proc. ACM SIGGRAPH 2002, San Antonio, TX (2002) 175–181
20. Hunt, K.H., Crossley, F.R.E.: Coefficient of restitution interpreted as damping in vibroimpact. ASME J. Applied Mech. **42** (1975) 440–445
21. Timoshenko, S.P., Goodier, J.N.: . Mc Graw Hill, New York (1970)
22. Avanzini, F., Rocchesso, D.: Modeling Collision Sounds: Non-Linear Contact Force. In: Proc. COST-G6 Conf. Digital Audio Effects (DAFx01), Limerick (2001) 61–66
23. Avanzini, F., Rocchesso, D.: Physical modeling of impacts: theory and experiments on contact time and spectral centroid. In: Proc. Int. Conf. Sound and Music Computing (SMC06), Paris (2004) 287–293
24. Basdogan, C., Srinivasan, M.A.: Haptic rendering in virtual environments. In Stanney, K.M., ed.: Virtual Environments Handbook. Lawrence Erlbaum Associates (2002) 117–134
25. Colgate, J.E., Brown, J.M.: Factors Affecting the Z-Width of a Haptic Display. In: Proc. IEEE Int. Conf. on Robotics & Automation, San Diego (1994) 3205–3210
26. Kuchenbecker, K.J., Fiene, J., Niemeyer, G.: Improving Contact Realism through Event-Based Haptic Feedback. IEEE Trans. on Visualization and Comp. Graphics **13**(2) (2006) 219–230

Designing Haptic Feedback for Touch Display: Experimental Study of Perceived Intensity and Integration of Haptic and Audio

Ville Tikka and Pauli Laitinen

Nokia Research Center, P.O. Box 407, FI-00045 NOKIA GROUP
`{ville.tikka, pauli.laitinen}@nokia.com`

Abstract. We studied the subjectively perceived intensity of the haptic feedback and the effects of the integration of the audio and haptic feedback. The purpose of the study was to specify design principles for haptic feedback on a piezo actuator enhanced mobile touch display device. The results of the study showed that the best corresponding physical parameter to perceived feedback intensity was the acceleration of the haptic stimulus pulse. It was also noticed that the audio stimuli was biasing the perception of the haptic stimuli intensity. These results clarify the principles behind haptic feedback design and imply that the multisensory integration should be stressed when designing haptic interaction.

1 Introduction

Mobile hand-held devices are getting smaller while the number of functions incorporated in a single device is growing. This trend sets high requirements for user interfaces that should simultaneously fit in a compact size and enable versatile use in an intuitive way. One possibility to tackle this challenge is to develop devices with touch displays that save space on the surface of a mobile device and allow variable configurations of buttons. The extreme of this development could be a device that is basically comprised only of the touch display and has no physical buttons at all.

At the moment there are many mobile devices on the market with touch displays, but generally they do not support two-way haptic interaction. Without the haptic feedback on touch display the user can only rely on audio and visual feedback, which breaks the metaphor of direct interaction [1]. Thus by adding the haptic feedback to the touch display, it would be possible to improve the usability of traditional use cases of touch displays [2] as well as create totally new effective modes of interaction in mobile devices without physical buttons.

Haptic feedback is a vital part of human perception and it is also fundamental for physical user interfaces. Through touching we convey effectively functional signals as well as emotion [3]. Even if touch may not be as rich as vision we have an amazing range of haptic sensations and touch displays should be able to take full advantage of them [1].

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 36 – 44, 2006.

The piezo actuators have been noted to be the best choice for providing haptic feedback for mobile touch display devices because they can be miniaturized, are durable, and most importantly offer efficient and versatile actuation for effective user interaction. The piezo actuators also produce natural sound and this can be utilized when designing feedbacks for mobile touch display interaction.

Although the importance of the haptic feedback is generally realized, there is limited amount of formal studies of how to design the haptic feedback on touch displays. There are no studies that we are aware of that would have researched how the intensity of the different haptic waveforms is perceived nor how the haptic feedback intensity is perceptually integrated with the visual and auditory feedback during the touch display use.

2 Background

The psychological perceptual studies have been exploring the issue of multisensory integration and how haptics are related to other sensory modalities. However, there is only sparse information how haptic and auditory modalities are integrated and how this integration affects perceived intensity.

The sensory integration depends on various factors both on perceptual and on higher cognitive level. The division between perceptual and higher cognitive level is usually called the processing level issue and it refers to the question whether the observed interactions originate in automatic perceptual processes or in later decisional ones. That is important because human responses are relevant to the intermodal coordination only if they reflect basic perceptual processes, rather than specific decisional strategies [4]. Generally the studies concerning multisensory integration have claimed to deal with perceptual processes and not with higher post-perceptual processes [5].

In perceptual studies vision has generally been noticed to bias the perceived location of touch [6] and perceived location of audio [5]. Also touch has been noticed to bias the perceived location of audio [7]. Consequently when pressing virtual buttons on a touch display the haptic and auditory feedbacks are perceived on the location of pressed virtual buttons even if the sound is coming from piezo actuators that are moving the whole display.

However, the question of modality dominance is not that clear when discussing how auditory and haptic modalities are affecting each other and how does this integration affect perceived feedback intensity. Typically the perception of surfaces is dominated by the haptic component of perception [8] but many studies have noticed that audio can bias and can concretely affect the haptic perception [9, 10, 11].

The integration between audio and touch has been studied also from attentional viewpoint and the studies suggest that there are no crossmodal links in attention between these modalities [7, 12]. Thus the users' attention to either of the audio or the haptic feedback is presumably not affecting the perceived intensity. This is contrary to the strong crossmodal links in attention reported both between audio and vision and between vision and touch [13].

However, it has been reported that auditory and haptic modalities are integrated when perceiving the intensity of the stimuli [14] and therefore it could be assumed that these two modalities are affecting each other in touch display use but the exact direction and effects of the integration are not known.

3 Aim of the Study

In this study we focused on researching the perceived intensity of the haptic feedback that was generated by piezo actuators. In particular, we studied the use of a virtual button on touch display because it is the most common use case of a mobile touch display device. The selected piezo technology intrinsically produces both haptic and audio feedback and thus we also studied the effect of the integration of these modalities on haptic feedback intensity perception.

We did two separate studies in order to find out how feedback intensity is perceived and how the perception of intensity is affected by the modality integration. In the first study, subjects evaluated feedbacks that consisted of both haptic and audio feedback. In the second study subjects evaluated only haptic feedbacks as audio feedback was excluded.

Our hypotheses were that the rise time and form of the haptic waveform would correspond with the perceived feedback intensity and that the audio feedback would integrate with the haptic feedback and bias the perception of the haptic feedback intensity.

4 Device and Stimulus Design

Device used in the study was a mockup handheld device with large touch display similar to the Nokia 770 tablet (Fig. 1). The haptic stimulus was generated by piezo actuator solution, which enables the production of various pulse shapes with displacement amplitudes on a scale of several hundred micrometers. The piezo actuator solution is similar to those that have been introduced by Tuovinen [15] and Poupyrev [16], but optimized to the requirements of mobile devices and use contexts.

The stimuli were generated with a robust and simple bending bimorph placed under the touch display module. The stimulus pulse magnitudes and dynamics can be controlled accurately with these actuators [17]. The pulse derives from the energy the piezo actuator supplies in a form of actuator deflection and the associated force subjected to loads [18]. The loads by device mechanics consist of masses and spring loads and are fixed in mechanical design whereas the loads produced by user interaction may vary.

The energy required for haptic stimuli production (force times amplitude) was at a level that inevitably causes a counter pulse to the opposite surface of the device by the

Fig. 1. Nokia 770 tablet with touch display

laws of conservation of momentum, although actuation is targeted at the immediate touch interface. This is characteristic of mobile devices that are small and lightweight by nature.

The modulation of the stimuli was done by controlling the driving voltage and the current of the piezo actuator and thereby altering two parameters, namely the rise time and the displacement amplitude (Fig. 2). This way it was possible to create virtually any kind of one-dimensional haptic stimuli with a relatively large dynamic range. In the context of the present device the range of the amplitudes varied from a few micrometers to a few hundred micrometers. The rise time of the stimulus pulse was varied at a range of 3-7 ms and the displacement amplitude range for stimuli was 3-180 µm. The fall time of the single pulse was fixed at 5 ms for all stimuli.

The piezo actuator also produces sound while actuating. The audio feedbacks were not separately designed, but the intrinsic sounds generated by the piezo actuator were

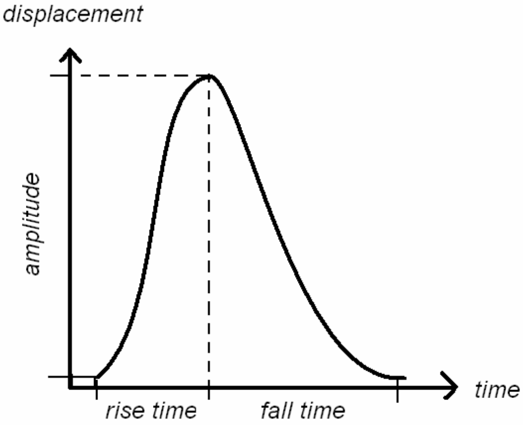


Fig. 2. The stimuli were generated by changing rise time and displacement amplitude. The fall time was fixed to 5 ms.

used as stimuli. As the sound originated from piezo actuator surface instead of touch display surface, the user interaction did not have significant effects on audio stimuli. The maximum sound levels associated with stimuli varied between 28-60 dB at 35 cm distance from the device.

The characteristics of the stimuli pulse were measured with Laser Vibrameter and the displacement amplitude was measured in respect of the displacement time without any load applied by the users. The damping force and its dynamics produced by user interaction were measured and recorded by the resistive touch display.

5 Procedure

We did two separate studies, one with haptic and auditory stimuli, and one with haptic stimuli only where the auditory stimuli were excluded by using earplugs and hearing protectors. 8 naïve participants took part in each of the studies. In both studies there were three females and five males participating and the average age of the subjects was in the first study 33 years and in the second study 27 years. All participants were staff members of Nokia Research Center.

Altogether 16 different physical stimuli were tested. The stimuli were generated altering amplitude and rise time of the stimulus pulse. The stimuli consisted of these 16 different stimuli that were repeated three times in a randomized order resulting to total 48 stimuli.

During the experiment subjects were holding the device perpendicularly in one hand and pressing the key number 5 on the virtual keypad on touch display with the other hand's thumb. In both studies users were told to rate the perceived haptic stimulus intensity without any notion of the purpose of the experiment. Each stimulus was rated verbally on a 1-to-5 rating scale right after each key press, where 1 was *clearly too weak,* 3 *moderate*, and 5 *clearly too strong*. Before starting the experiment users were able to try out the different stimuli.

6 Results

During the experiments, data was collected on the subjective evaluations of the stimulus intensities. In order to find out how the physical haptic pulses are correlated with subjectively evaluated stimulus intensities, the data sets from the two studies and physical parameters of the stimuli pulse were analyzed with linear regression analysis. To explore the integration between haptic and audio stimuli pairwise comparisons were made for the two data sets with Mann-Whitney U test.

The results suggest that the simple correspondence exists within the scale and accuracy used in the device and needed in general in virtual button applications. The average acceleration of the rising edge of the haptic pulse was found to offer the closest correspondence between the physical parameters of the stimulus pulse and the subjective evaluations of stimulus intensity. This was the case both with *haptics and audio* stimuli and with *haptics only* stimuli. The coefficient of determination that represents the percent of the data that is closest to the line of the best fit, was 0.92 ($p<0.01$) for *haptics and audio* (Fig. 3) and 0.90 ($p<0.01$) for *haptics only* (Fig. 4). There were no correlations between the perceived intensity and the input force or the time the users were pressing the touch display.

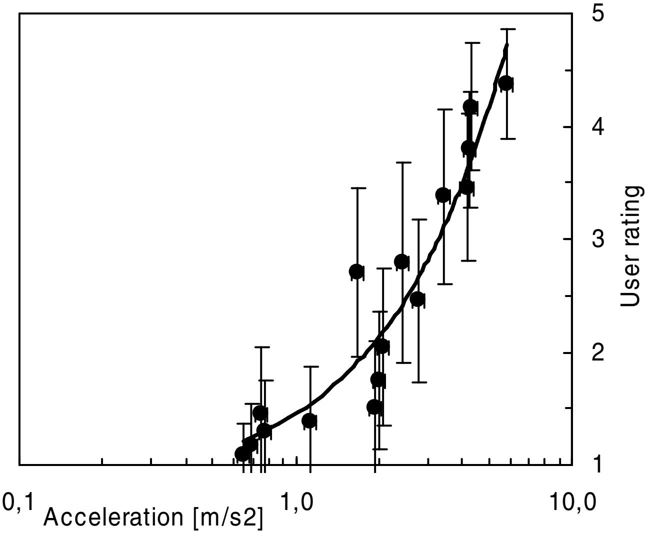


Fig. 3. The scatterplot shows the mean values and standard deviations for subjective evaluations of stimulus intensities for the *haptics and audio* stimulus and the average accelerations and error bars of the haptic stimulus pulses. The linear regression fits the data with a coefficient of determination $R^2=0.92$ ($p<0.01$).

The pairwise comparisons of the data sets from the *haptics and audio* and *haptics only* studies shows that there was difference in the stimulus intensity evaluations between the two studies. In the *haptics only* study three stimuli were evaluated weaker than in the *haptics and audio* study (Fig. 5). These differences imply that audio has an effect to perceived intensity of the haptic stimulus within the stimuli range used in the study (28-60 dB in audio and 0,6-8 m/s² acceleration) and it biases the perception and increases the perceived strength of the haptic stimulus.

This result was according the hypotheses although the biasing effect was not as clear as it was expected. The level of biasing was related to the level of audio as the stimuli that have higher sound levels were biased more than the stimuli that have lower sound levels.

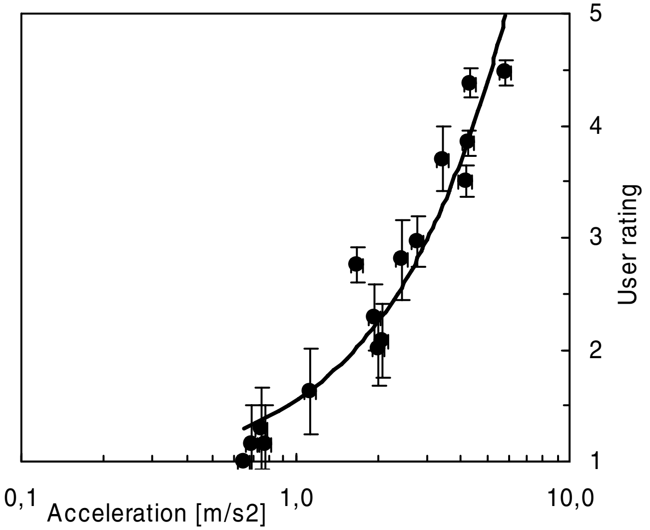


Fig. 4. The scatterplot shows the mean values and standard deviations for subjective evaluations of stimulus intensities for the *haptics only* stimulus and the average accelerations and error bars of the haptic stimulus pulses. The linear regression fits the data with a coefficient of determination $R^2=0.90$ ($p<0.01$).

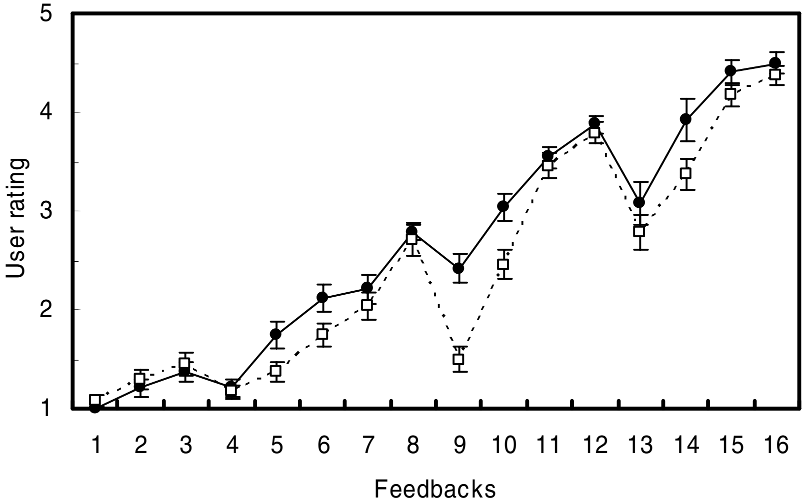


Fig. 5. The figure shows the mean values and standard error of means for subjective evaluations of stimulus intensities both for the *haptics and audio* (*continuous line*) and *haptics only* (*dotted line*) stimuli. The pairwise comparisons between the two studies show that there are statistically significant differences between the two studies in stimuli number 9 ($p<0.01$), 10 ($p<0.05$), and 14 ($p<0.05$).

7 Discussion

In the study it was found out that the displacement dynamics of the touch display correlated with the perceived intensity. The acceleration of the rising edge of the haptic stimulus pulse had the closest correspondence to the perceived intensity of the stimulus within the scale and the accuracy that is needed in mobile touch display devices in general and particular in virtual button applications. The displacement of the touch display surface was inadequate to explain the perceived intensity and there were no correlations between the perceived intensity and the input force or the time the users were pressing the touch display. These results were quite similar between the *haptics and audio* and *haptics only* conditions.

The pairwise comparisons between the two conditions suggest that audio stimulus has some effect on stimulus intensity perception. It was noticed to bias the stimulus intensity evaluations in a way that stimuli that have higher sound levels were biased more than stimuli that have lower sound levels. This was predictable but generally the biasing effect was weaker than expected. It could be partially explained by the small sample size and in a study with more subjects the difference would probably be more evident. However, it was noticed that the sound is affecting perceived haptic stimulus intensity and thus the effects of sound should carefully be taken into account in haptic feedback design in mobile devices.

There were also observable differences in stimulus evaluations between the subjects. It was noticed that some of the subjects preferred stronger stimuli as some of the subjects liked the weaker ones. This implies that different people and perhaps different cultures could have distinctive tastes for haptic feedbacks.

The challenges in designing haptic feedbacks are not only technological ones but also relate to broader research and design issues. First of all the haptic perception is affected by the simultaneous visual and auditory perception and therefore the haptic design should be linked to visual and audio design. Secondly the haptic perception is context dependent and application, device, and environment are probably affecting to the experience of the expedient feedback. Finally the personal and cultural differences are underlining the diversity of the haptic design issues. The haptic preferences presumably vary between individuals and between cultures and there is need of adjustability or even adaptivity in haptic design.

These above-mentioned issues are great challenges when developing and designing haptic user interfaces. Further studies are planned to tackle these challenges in a mobile context and study how to design both effective and pleasant haptic interaction for touch display devices.

References

1. Poupyrev, I, Maruyama, S.: Tactile Interfaces for Small Touch Screens. Proceedings of UIST 2003, ACM Press (2003) 217-220
2. Fukumoto, M., Sugimura, T: Active Click: Tactile Feedback for Touch Panels. CHI'2001, Extended Abstracts, ACM Press (2001) 121-122
3. MacLean, K.E.: Designing with Haptic Feedback. Proceedings of the 2000 IEEE International Conference of Robotics & Automation (2000) 783-788

4. de Gelder, B., Bertelson, P.: Multisensory Integration, Perception and Ecological Validity. Trends in Cognitive Neurosciences, Vol. 7(10) (2003) 460-467
5. Bertelson, P., de Gelder, B.: The Psychology of Multimodal Perception. In: Spence, C., Driver, J. (eds.): Crossmodal Space and Crossmodal Attention. Oxford University Press, New York (2004) 141-177
6. Pavani, F., Spence, C., Driver, J.: Visual Capture of Touch: Out-of-the-Body Experiences with Rubber Gloves. Psychological Science, Vol. 11(5) (2000) 353-359
7. Caclin, A., Soto-Faraco, S., Kingstone, A., Spence, C.: Tactile "Capture" of Audition. Perception & Psychophysics, Vol. 64(4) (2002) 616-630
8. Lederman, S., Klatzky, R.: Multisensory Texture Perception. In: Calvert, G., Spence, C. Stein, B. (eds.): The Handbook of Multisensory Processes. The MIT Press, Cambridge Massachusetts London (2004) 107-122
9. Jousmäki, V., Hari, R.: Parchment-Skin Illusion: Sound Biased Touch. Current Biology, Vol. 8(6) (1998) R190
10. Zampini, M., Guest, S., Spence, C.: The Role of Auditory Cues in Modulating the Perception of Electric Toothbrushes. Journal of Dental Research, Vol. 82(11) (2003) 929-932
11. Bresciani, J.-P., Ernst, M.O., Drewing, K., Bouyer, G., Maury, V., Kheddar, A.: Feeling What You Hear: Auditory Signals Can Modulate Tactile Tap Perception. Experimental Brain Research, Vol. 162 (2005) 172-180
12. Lloyd, D., Merat, N., McGlone, F., Spence, C.: Crossmodal Links Between Audition and Touch in Covert Endogenous Spatial Attention. Perception & Psychophysics, Vol. 65(6) (2003) 901-924
13. Driver, J., Spence, C.: Crossmodal Spatial Attention: Evidence from Human Performance. In: Spence, C., Driver, J. (eds.): Crossmodal Space and Crossmodal Attention. Oxford University Press, New York (2004) 180-220
14. Schürmann, M., Caetano, G., Jousmäki, V., Hari, R.: Hands Help Hearing: Facilitatory Audiotactile Interaction at Low Sound-Intensity Levels. Journal of Acoustical Society of America, Vol. 115(2) (2004) 830-832
15. Tuovinen, J.: Piezoelectric User Interface, EP1449267, (2002)
16. Poupyrev, I., Maruyama, S., Rekimoto, J.: Ambient Touch: Designing Tactile Interfaces For Handheld Devices. Proceedings of the 15th annual ACM Symposium on User Interface Software and Technology, (2002) 51-60
17. DeVoe, D., Pisano A., J.: Modeling and Optimal Design of Piezoelectric Cantilevermicroactuators. Microelectrical Systems, Vol. 6(3) (1997) 266-270
18. Giurgiutiu, V., Chaudhry, Z., Rogers, C.A.: Stiffness Issues in the Design of ISA Displacement Amplification Devices: A Case Study of a Hydraulic Displacement Amplifier. In: Chopra, I. (ed.) Proceedings of SPIE, Smart Structures and Materials 1995: Smart Structures and Integrated Systems, Vol. 2443 (1995) 105-119

Rhythmic Interaction for Song Filtering on a Mobile Device

Andrew Crossan[1] and Roderick Murray-Smith[1,2]

[1] Department of Computing Science,
University of Glasgow,
Scotland, UK, G12 8QQ
{ac, rod}@dcs.gla.ac.uk
[2] Hamilton Institute,
NUI Maynooth, Co. Kildare, Ireland

Abstract. This paper describes a mobile implementation of song filtering using rhythmic interaction. A user taps the screen or shakes the device (sensed through an accelerometer) at the tempo of a particular song in order to listen to it. We use the variability in beat frequency to display ambiguity to allow users to adjust their actions based on the given feedback. The results of a pilot study for a simple object selection task showed that although the tapping interface provided a larger range of comfortable tempos, participants could use both tapping and shaking methods to select a given song. Finally, the effects of variability in a rhythmic interaction style of interface are discussed.

1 Introduction

Mobile music players are an important part of everyday life for many people on the move. While the capacity of these devices and therefore the amount of information stored on them has increased dramatically in the last few years, they have got smaller and smaller. While this is important for portability, it presents interesting challenges in allowing people to interact with these devices in a fun, but efficient manner. Especially while on the move, where staring at a small screen while navigating a busy environment might prove difficult and potentially dangerous. One interesting example of a prototype system, where the music adapts to aid navigation is discussed in [1]. Directional information and ambiguity are presented to the user through volume and audio panning of the current track to direct the user to a destination or alert them when a choice of different paths is available.

This paper examines a novel concept for allowing users to browse their music library by tempo. It attempts to provide an intuitive, fun and, importantly, 'eyes free' method of interaction that allows the user to interact with the device while walking and potentially without removing it from a bag or pocket.

1.1 Mobile Gesturing

Mobile devices are now widely used for a variety of everyday tasks. However, due to the requirement for a small screen and keyboard, interacting with these devices often

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 45–55, 2006.

proves to be difficult. On-screen buttons are generally closely grouped together making interactions slow and error prone. The combination of reduced resources and the need for mobility may mean that a direct translation of the desktop interaction techniques may not provide the best method of interacting with these devices. The development of new interaction techniques and technologies provides the opportunity for a more continuous form of interaction. There are currently several examples of gesturing in mobile devices available in the literature. Pirhonen, Brewster and Holguin [2] demonstrate an example of stylus gesturing as an input technique for controlling a PDA based MP3 player. These interactions are designed to be intuitive for the task performed. Pirhonen, Brewster and Holguin were able to demonstrate significant usability benefits with the gesture interface over the standard interface, with users indicating that the gesture system required a lower workload to perform the task.

Recent commercially produced mobile devices have been developed with integrated low cost accelerometers. For example, both Samsung and Nokia have developed phones (the Samsung SCH-S310 and Nokia 3220) that use integrated accelerometers as an interaction mechanism. These sensors have previously been examined for use in context sensing applications, but also offer the potential for allowing a user to control a device through gestures. Strachan *et al.* [3] describe the bodyspace project where a mobile device can be controlled through movement. Acclerometers attached to the mobile device detect the movement of the device from one pre-defined position to different areas around the body. A control action can then be assigned to each area of the body allowing a user to open documents or applications through body-related gesturing.

Hinkley *et al.* [4] demonstrate how a combination of sensors can be used to interact with a mobile device. Recognition of orientation of the device was, for example, used to start a voice recording application (in combination with proximity sensors and touch sensors) or automatically change the screen orientation. Hinkley *et al.* demonstrate how these novel sensors can be used to provide a natural and intuitive interaction mechanism.

1.2 Rhythmic Interaction

The methods described above all use point-to-point gestures where the user starts in one position and moves through a specific trajectory to another set position to finish the gesture. Rhythmic gesture methods of interaction have largely been ignored as an input mechanism for computers. However, they have the potential to offer a natural method of interacting with a device for several types of task. This is particularly important in the case of mobile devices, where their small size and need for portability can lead to slow and frustrating interactions. The introduction of new, affordable sensor technology, allows greater possibilities for interacting with these devices.

One benefit of rhythmic gestures is that they can be designed to be naturally repeatable. Tapping out a rhythm on a desk with a finger or dancing to music would be two everyday examples of repeatable rhythmic gestures. If the user taps out a rhythm or performs a rhythmic gesture, it feels natural for the user to repeat this gesture until recognition occurs. This allows the user to perform the gesture to excite a state in the system (akin to finding the ‘resonant frequency’ of particular states or options).

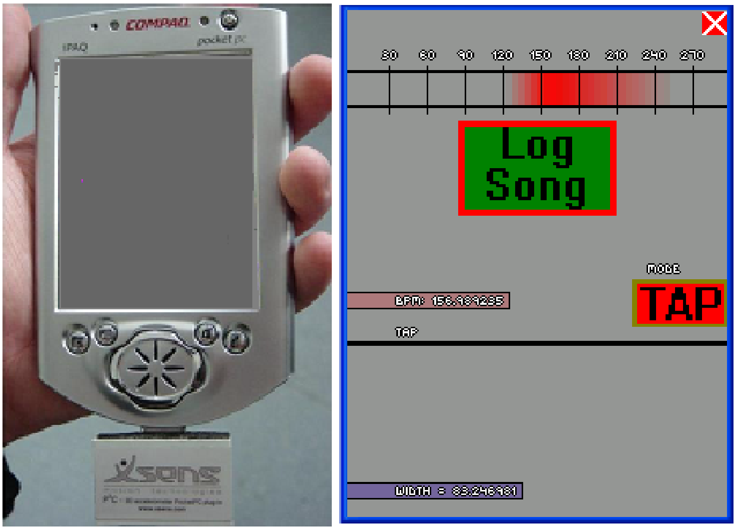


Fig. 1. PDA with the Xsens P3C accelerometer attached to the serial port (**left**). The song browsing environment (**right**).

The system can then respond by presenting a series of options based on a probabilistic model of how likely it is that the user is trying to select each option. More likely options can be displayed to the user more prominently, and the dynamics of the system can be altered such that it is easier to select more probable options. The user can then refine his or her movements, responding to the feedback, to select the target. Lantz and Murray-Smith [5] conducted an initial usability study using rhythmic gestures with a mobile device. They examined 10 different rhythmic gesture movements recorded through an accelerometer attached to a mobile device to look at consistency and ease of movement for each of the gestures. They demonstrate a dynamic movement primitive approach that is used to model the gestures to eventually provide recognition. Rhythmic interaction in the form of 'haptic dancing' with a PHANToM force-feedback device is described in [6]. One of the insights from this work is that future interaction with a computer might not be similar to the current command-and-control style of interaction, but more like the flowing transitions of control and the give-and-take of dancing.

Synchronisation of oscillators and phase entrainment has been studied extensively in the physics literature [7], and the theory has recently been applied to the design of rhythmic interaction methods. In [8], we examined extraction of gait phase from an accelerometer attached to a mobile device held by a user while walking applied to mobile usability. The acceleration signal, allowed extraction of each step and further allowed an estimate to be made at each point of the user's current phase within a step. Here we present a rhythmic interaction system for song filtering on a mobile device. For input, we use repeated screen tapping or shaking gestures sensed using an accelerometer. Songs are selected by synchronising the tapping or device oscillations with the tempo of the song being selected.

2 The System

This system was developed using an HP 5450 PDA with the Xsens P3C 3 degree of freedom linear accelerometer attached to the serial port (shown on the left of Figure 1). Its effect on the balance of the device is negligible (its weight is 10.35g). The accelerometer was used to detect movement of the device, sampling at a rate of approximately 90Hz.

The interface used for the study is displayed on the right of Figure 1. The top section of the screen shows a scale indicating the user's current beats per minute (BPM), and the uncertainty in tempo due to variability in the timing of the user's beats. To produce a beat, the user must either tap the designated area of the screen (below the dark line shown on the right of Figure 1), or shake the device. The button on the right allows the user to change between tap mode where the user taps out a tempo, and accelerometer mode where the user moves the device rhythmically at a tempo through the air. When processing the accelerometer data, a low-pass filter is applied in real time to remove muscle tremor and sensor noise from the trace, and a high pass filter is used to correct for drift over time (as the user's posture may change). Examples of processed acceleration traces for the device's vertical accelerometer axis are shown in Figure 2.

Songs of differing tempo are stored in the system. A song is presented to the user by a short clip of looped audio from the song and is associated in the search space with its tempo. The volume of the song played to the user is then a function of the distance between the tempo of the song and the current detected BPM. The volume is also dependent on the variability in the user's tapping. A high level of variability may be due to the user browsing the space. The system therefore presents the user with more context information, represented by a higher level of volume from nearby songs. Low BPM variability will occur when a user has decided on a target and adjusts his or her tempo to match the selected song. The system then presents a more focussed area of the space to the user thereby reducing the volume of songs whose tempo is nearby but does not match the current BPM. The variability is displayed to the user by the width of the distribution at the current position shown on the BPM scale.

This can be viewed as a form of the pointerless selection mechanism described by Williamson and Murray-Smith [9], which is based on detection of control behaviour in the user's observed actions.In this case, the metaphor is that each track has a 'resonant frequency' which can be 'excited' by the user's tapping actions, such that the initial beats narrow down the search space. The current uncertainty in the negotiation process is indicated by a volume which increases the closer the user is to the resonant frequency for that track. The display indicates the ambiguity in the tapping behaviour, by visually showing the spread of frequencies being excited, and also in sound by spreading the excitation to songs proportional to the spread in beat frequencies. Options close to the excitation frequency have a volume proportional to their likelihood, as a stimulus to the user, which allows the user to entrain more precisely with the desired song, homing in to the final selection.

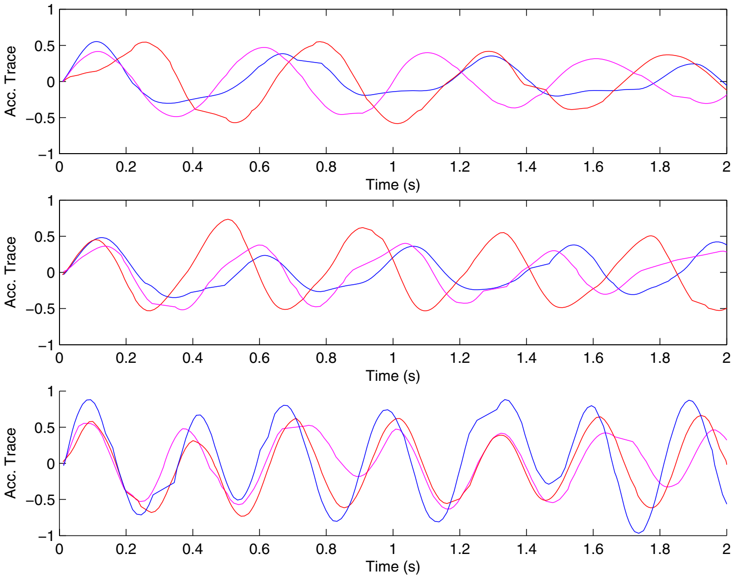


Fig. 2. Data from three participants when in acceleration mode and synchronising with a song of 112BPM (top), 135BPM(Middle) and 192BPM (Bottom)

3 Pilot Results

A pilot study was carried out with a small range of tracks to examine how the system was used to browse within the space. Seven participants took part in the pilot study using the application described with seven songs of differing tempo stored in the environment. The songs chosen had tempos of 72, 87, 112, 135, 150, 170 and 192 beats per minute. Users were asked to target each song using accelerometer and tapping modes in a counterbalanced order. Users successfully managed to target and maintain synchronisation for several beats with a mean of 6.7 of 7 songs using tapping mode and 4.9 of 7 songs using accelerometer mode. This difference is to be expected as the tapping interface allowed the users to easily look at the screen and therefore use the BPM scale to browse the space. Also, *post hoc* analysis of the data suggests that 5 unsuccessful targeting attempts in the accelerometer condition were due to poor performance of the recognition algorithm for slower tempo songs. The tapping mode allowed a larger range of motions to be incorporated due to the lower physical demand required to tap the stylus on the screen as opposed to moving the device through the air. The acceleration traces for one axis for three users targeting three songs of tempo 112, 135, and 192 beats per minute are shown in Figure 2. In this instance for the specific movement that the participants were instructed to perform, each cycle of the sinusoid corresponds to one oscillation with the device. This would not necessarily be true for different motions.

Observation during the experiment suggested that some of the participants were able to use the variability in tempo to browse the space. A common technique was to start with slow beats, and increase rapidly to move the current BPM position up through the scale. This presented the user with a range of tempos from the slowest to the fastest BPM while maintaining a high level of variability. This ensured that a wide distribution of the space was always presented to the user at different positions on the BPM scale, which is useful for browsing the space.

3.1 Considering Variability in Tapping

When considering this, or any gesture-interaction mechanism, it is important to take into account the effect of variability of the user's actions in the recognition process. When analysing beats per minute, it can be seen that:

$$BPM = \frac{60}{IBG}, \tag{1}$$

where BPM is beats per minute, and IBG is the inter beat gap in seconds. If we introduce a bias error in the time between taps into this equation, a new value for the predicted beats per minute is given by:

$$BPM_{varLim} = \frac{60}{(IBG + v)}, \tag{2}$$

where BPM$_{varLim}$ is the value of the BPM measured given where variability in inter beat gap is v. If the target BPM was 120, and therefore the target inter beat gap 0.5seconds, a variability of 0.1 would correspond to the user tapping with an inter beat gap of 0.6 seconds. The offset can then be obtained by looking at the the difference between these values,

$$BPM - BPM_{varLim} = x, \tag{3}$$

where x is the difference in target BPM and recorded BPM for a given variability. Substituting for BPM and BPM$_{varLim}$ using equations 1 and 2, we can derive an equation for v in terms of IBG such that

$$v = \frac{xIBG^2}{1 - xIBG}, \tag{4}$$

again, substituting BPM for IBG using equation (1), we can see that

$$v = \frac{60x}{BPM^2 - 60xBPM}, \tag{5}$$

This equation can be used to plot the effect of tap variability on BPM. Figure 3 shows the inter tap variability required to target at different BPMs for accuracy thresholds of +3 to +21BPM. Figure 4 shows the corresponding plot for an accuracy threshold below the target BPM for -3BPM to -21BPM. It can immediately be seen that positive inter tap variability have a far larger effect on the BPM level than negative inter tap variabilities. Also, it can be seen that as the beats per minute increase, a smaller variability is required to keep the BPM within a given threshold.

These show the effect of variability in one tap or a group of taps with a constant variability. However, what it does not take account of is a users ability to synchronise to a particular beat. For example, if a user is given a song to tap along with, he or she may display variability with individual beats, but can use his or her knowledge of the

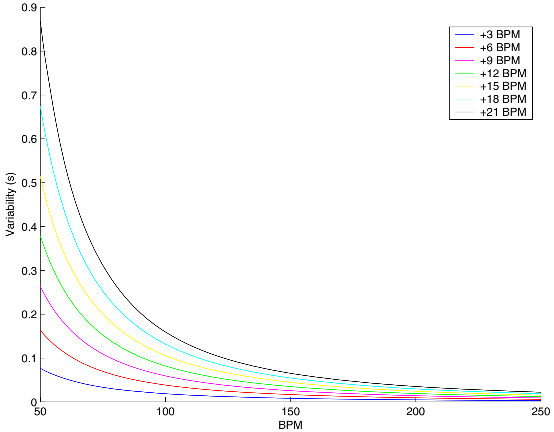


Fig. 3. Inter tap variability required to target within the given accuracy threshold above the target BPM for different BPM levels

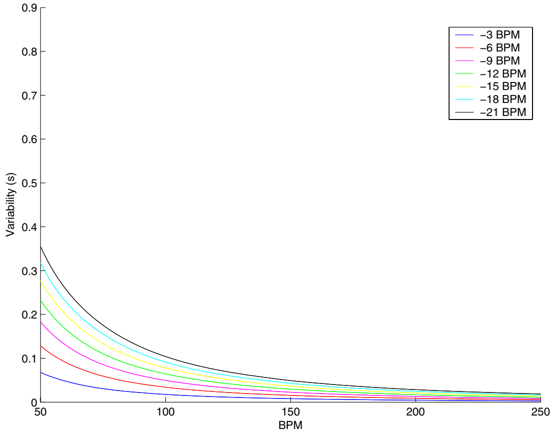


Fig. 4. Inter tap variability required to target within the given accuracy threshold below the target BPM for different BPM levels

song so far to predict the next beat, and therefore compensate for the initial variability. Measurement of actual user variability is important to determine how the user uses the music to synchronise.

In Figure 6 we show the effect of variability in tapping accuracy as a test user tries to locate songs at a range of BPM. The standard deviation of the Inter Beat Gap for songs at BPM of 72,112, 135, 170 and 192, was respectively 0.0383, 0.0435, 0.0200, 0.0210, 0.0171s, as shown in Figure 5, but although the slower songs have a lot of variability, in absolute terms, it does not translate into greater uncertainty, as shown in Figure 6. Note that the means are also extremely close to the target BPM, showing the precision users can achieve with the system.

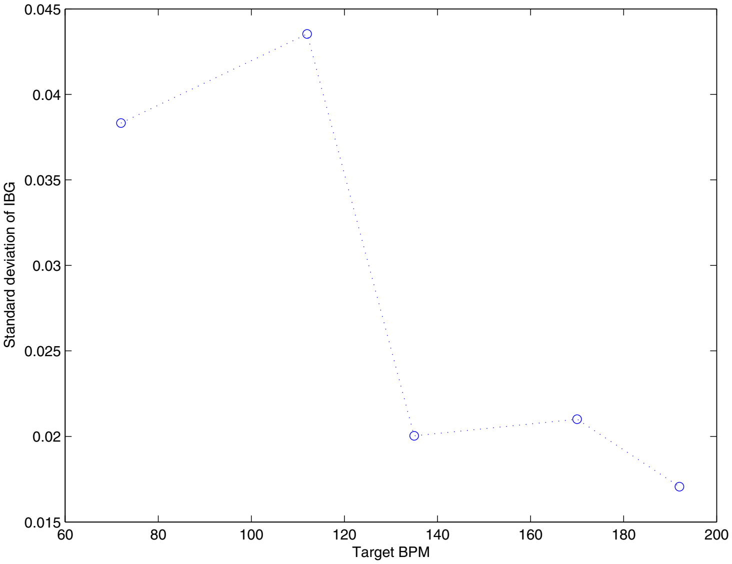


Fig. 5. Inter Beat Gap standard deviation at different BPM levels

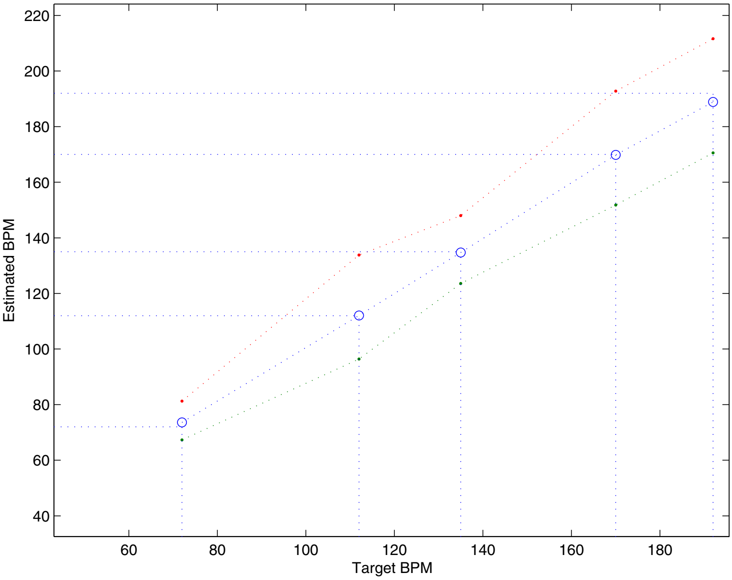


Fig. 6. Uncertainty of BPM estimate at different reference BPM levels (showing the spread of BPM associated with $\pm 2\sigma$ – two standard deviations in the inter-beat gap)

We can also get further insight into the working of the algorithm by looking at the time evolution of the mean and standard deviation of the BPM estimate, as shown in Figure 7, where 3 different runs are shown where a user acquires the desired BPM

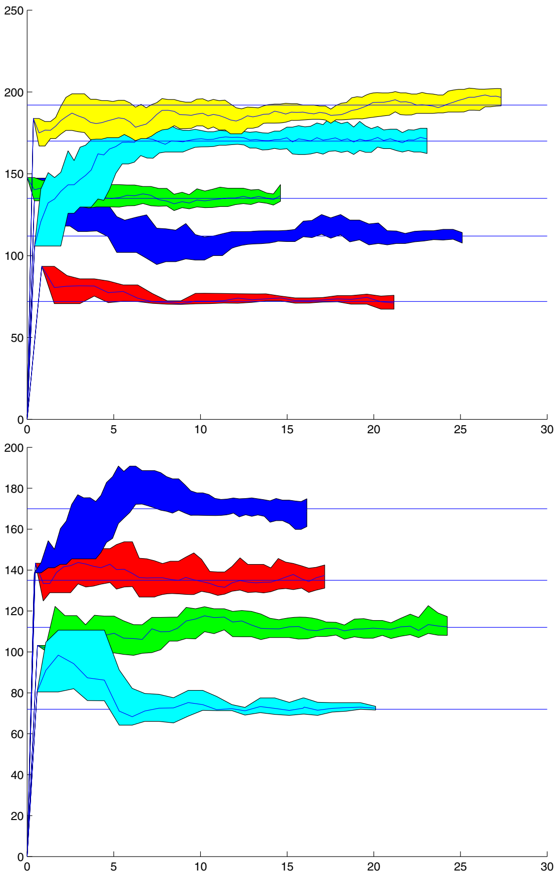


Fig. 7. Illustration of mean and standard deviation evolution as the user attempts to select each of 5 tracks with different BPM

level, while receiving feedback from the system. It can be seen that initially the user estimates the BPM for the appropriate song. There is a high level of variability as the user searches the space. The system supports this by increasing the volume of nearby songs due to this high variability of tapping. As the user hears and recognises the desired song, the mean tapping BPM moves towards the song's BPM and the level of variability of tapping tempo decreases presenting the user with a focussed area of the space centred around the song.

4 Discussion and Extensions

This system was successful in allowing users to browse an audio space of 7 songs. However, obvious scalability issues need to be examined before this interaction mechanism is shown to provide some benefit in this area. We envisage a system based on tempo to be used in conjunction with other complementary song selection mechanisms.

This should not be seen as a song selection technique, but as a system of filtering a selection of songs, as using tempo alone for selection would require the system itself to have more in-built intelligence. An extension that has been developed to the system described above combines different filtering mechanisms to provide more differentiability. The user rhythmically taps to screen to provide the tempo, while tilting the device to different orientations to browse different genres simultaneously.

In a rhythmic interaction system, it is important to consider how variability in the user's actions and timing will affect the performance of the system. In the initial implementations of the above interface, the user's BPM was determined by averaging the time between a fixed number of beats. This method was found to penalise convergence with slower songs, as with this system, synchronising with a slower song will always require more time to generate the same number of beats. An alternative method is to only consider beats within a certain time period. For example, if we only consider beats with 5 seconds of the current time, tapping with a tempo of 60BPM will use 5 beats when determining the mean. tapping at a rate of 120BPM will lead to 10 beats being considered when calculating the mean. This method provides a good compromise to robustness from the effects variability and rate of convergence for slower songs.

In this work we have used meta-data as a reference for the response frequency for each song. A more sophisticated approach is to derive the timing of beats for an individual track using signal processing techniques, and then require the user to tap in time with the automatically extracted beat-information. There are interesting challenges in this, including the subjective nature of music perception (different people might want to entrain with different aspects of the track, and this should be supported). For examples of modern automated analysis of music tracks see the Intelligent Sound project `www.intelligentsound.org`.

Finally, an application has been developed to demonstrate how rhythmic interaction methods can be used to adapt the user interface. The system analyzes the oscillatory acceleration history from the user walking or jogging, and will select the song of the closest tempo to the user's step rate in order to synchronise with the user. With an online Hilbert transform for gait phase estimates and techniques to adapt the playing of the music, such as granular synthesis, a self-tuning approach could adjust the music to synchronise with the jogging, creating a phase-locked loop between user and music.

5 Conclusions and Future Work

This paper describes an interaction technique for allowing users to filter songs by interacting with a mobile device through rhythmic gestures, and synchronising the tempo of the interactions with the tempo of a song stored in the environment. Different classes of object (not just songs) could be accessed in such a manner. Using tempo to browse songs seemed to be an intuitive and enjoyable concept to all users tested. With minimal or no training, all users were able to to pick up the device and synchronise with songs.

When selecting a song, the user is asked to produce a constant tempo. There is potential for a very rich and expressive method of interaction with the system by allowing differing rhythms to be entered. Particularly for the accelerometer case where the continuous interaction will allow users to adjust the form of their movements to provide more information to the system. This will become important for the scalability of the system where, for example, users could be more expressive, and modify their action between beats to select genre.

A key feature of this interface is the fun people had in relatively carefree exploration of the music space. Efficiency of selection might not always be the only metric for applications that focus on entertainment, such as audio or photo browsing. One user commented that it brought a bit of the more ‘grainy’ feel of old-fashioned radios with tuning dials to the relatively ‘antiseptic’ world of digital audio.

Finally the accelerometer offers the potential to produce a system to allow context sensitive selection of objects. For example, the system could detect a user walking or jogging, and select a random song from the user’s music library of the appropriate tempo to synchronise with the user’s step rate. Alternatively, the system could monitor the user’s heart rate and coax them to increase or decrease their step rate by playing songs of a slightly different tempo to the step rate and by exploiting phase entrainment. Future work will require a formal evaluation of these techniques, and fine-tuning of phase relationships during selection.

Acknowledgements

The authors are grateful for the support of SFI grant 00/PI.1/C067. RM-S is grateful for support of EPSRC grant GR/R98105/01, and BRG project *Continuous Gestural Interaction with Mobile Devices*. This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors’ views.

References

1. Warren, N., Jones, M., Jones, S., Bainbridge, D.: Navigation via continuous adapted music. In: ACM SIG CHI 2005, Portland (2005)
2. Pirhonen, A., Brewster, S.A., Holguin, C.: Gestural and audio metaphors as a means of control for mobile devices. In: Proceedings of ACM CHI 2002 (Minneapolis, MN), ACM Press Addison Wesley (2002) 291–298
3. Strachan, S., Murray-Smith, R., Oakley, I., Ängeslevä, J.: Dynamic primitives for gestural interaction. In Stephen Brewster, M.D., ed.: LNCS 3160: Proceedings of Mobile Human-Computer Interaction MobileHCI 2004: 6th International Symposium, Glasgow, UK, Springer-Verlag (2004) 325–330
4. Hinckley, K., Pierce, J., Sinclair, M., Horvitz, E.: Sensing techniques for mobile interaction. In: Proceedings User Interface Software and Technology (UIST 2000), ACM (2000) 91–100
5. Lantz, V., Murray-Smith, R.: Rhythmic interaction with a mobile device. In: NordiCHI ’04, Tampere, Finland, ACM (2004) 97–100
6. Gentry, S., Murray-Smith, R.: Haptic dancing: human performance at haptic decoding with a vocabulary. In: IEEE International conference on Systems Man and Cybernetics, IEEE (2003)
7. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A universal concept in nonlinear sciences. Cambridge University Press (2001)
8. Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., Musizza, B.: Gait phase effects in mobile interaction. In: ACM SIG CHI 2005, Portland. (2005)
9. Williamson, J., Murray-Smith, R.: Pointing without a pointer. In: ACM SIG CHI, ACM (2004) 1407–1410

Lemma 4: Haptic Input + Auditory Display = Musical Instrument?

Paul Vickers

Northumbria University,
School of Computing, Engineering, and Information Sciences,
Pandon Building, Camden St.,
Newcastle upon Tyne NE2 1XE, UK
`paul.vickers@unn.ac.uk`

Abstract. In this paper we look at some of the design issues that affect the success of multimodal displays that combine acoustic and haptic modalities. First, issues affecting successful sonification design are explored and suggestions are made about how the language of electroacoustic music can assist. Next, haptic interaction is introduced in the light of this discussion, particularly focusing on the roles of gesture and mimesis. Finally, some observations are made regarding some of the issues that arise when the haptic and acoustic modalities are combined in the interface. This paper looks at examples of where auditory and haptic interaction have been successfully combined beyond the strict confines of the human-computer application interface (musical instruments in particular) and discusses lessons that may be drawn from these domains and applied to the world of multimodal human-computer interaction. The argument is made that combined haptic-auditory interaction schemes can be thought of as musical instruments and some of the possible ramifications of this are raised.

1 Multimodal Challenges

Little research has been conducted into how haptic and auditory modalities can best be combined in human-computer interaction. However, in the non-computing world the two modalities have been partners for a long time. First Hollywood gave us *movies* (the visual display in computing terms). Then came the *talkies* (and auditory display in computing). In *Brave New World* (1932) Aldous Huxley offered us the notion of the *feelies* – the arm rests in theatre seats would provide haptic stimulation during erotic features. Whilst the feelies are not with us yet haptic displays are making inroads into the development of human-computer interfaces. As the auditory display community has discovered there are several hurdles that must be jumped before a new interaction modality is considered acceptable. Haptic interaction also raises its own usability issues which are being dealt with by researchers. However, the combined use of auditory display and haptic input and output in a single application raises a new set of

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 56–67, 2006.

design challenges. The field is, perhaps, in an analogous position to that of cinema in the late 1920s. Early film sound was largely causal and was thus a recording of the sonic events in the scene. The use of sound as a separate non-synchronous entity was advocated by Russian filmmaker Sergei Eisenstein whose 1928 *Statement on Sound* [1] suggested using sound as a counterpoint to the visuals. Eisenstein wanted to make montages of sound just as he used visual montage to great effect in his earlier silent films. The French filmmaker René Clair was an early promoter of non-synchronous sound which has today developed into the disciplines of voice-overs/narration, sound effects & Foley art, and film music. These early pioneers helped to make film a truly bi-modal communication channel. Today, we are now beginning to move away from direct causal haptics and sonification in which the touch and auditory modes are used to accentuate visual information and are seeing new true multi-modal interfaces in which the different senses are used for separate but complementary information streams.

1.1 Design Issues in Auditory Display

The larger questions of sonification design are concerned with issues of instrusiveness, distraction, listener fatigue, annoyance, display resolution and precision, comprehensibility of the sonification, and, perhaps binding all these together, sonification æsthetics.

There is a tension in auditory displays between the sonification being perceptible to its intended audience and being too intrusive or annoying. In their work on awareness support systems, Hudson and Smith [2] articulated the problem of intrusion in terms of awareness and privacy. They stated that this "*dual tradeoff is between privacy and awareness, and between awareness and disturbance*". The more information an auditory display provides the richer the sonification yet the greater the potential for disturbance, annoyance, and an upset in the balance of the acoustic ecology. Gutwin and Greenberg [3] claimed sonification is a tradeoff between being well informed and being distracted. Kilander and Lönnqvist noted the effect of such sonifications on people sharing the workspace: "*In a shared environment, one recipient may listen with interest while others find themselves exposed to an incomprehensible noise*" [4]. Indeed, commenting upon the design of their *nomadic radio* system[1], Sawhney and Schmandt [5, 6] cautioned that care must be taken to ensure that the auditory display intrudes minimally on the user's social and physical environment.

In dealing with intrusiveness, Pedersen and Sokoler [7] framed the problem as a balance between putting a low demand on attention versus conveying sufficient information. They studied this problem through an '*ecology of awareness*' thus acknowledging the importance of the acoustic ecology of a sonification.

---

[1] In *nomadic radio* a mixture of ambient sound, recorded voice cues, and summaries of email and text messages is used to help mobile workers keep track of information and communication services.

Pedersen and Sokola made the auditory, visual, and haptic representations of their AROMA system highly abstract – abstraction would allow useful information to be communicated without divulging too many details that would violate privacy[2]. It was hoped that abstract representations would be better at providing "*peripheral non-attention demanding awareness*" [7, p. 53]. It was also noted that such abstract representations lend themselves to being remapped to other media (what Somers [8] would call *semiotic transformation* – similar to Eisenstein's *transference*), or, in turn foster the accommodation of user preferences (an important aspect of *æsthetic computing* – see Fishwick [9]). Unfortunately, user studies showed that the abstraction led to users interpreting the representations in varied ways that were not always correct [10]. Furthermore, Kilander and Lönnqvist [4] warned that the "*monitoring of mechanical activities such as network or server performance easily runs the risk of being monotonous*" a finding observed by Pedersen and Sokola who reported that they soon grew tired of the highly abstract representations used in AROMA. It is interesting that some of the blame was attributed to an impoverished æsthetic, the feeling being that involving expertise from the appropriate artistic communities would improve this aspect of the work.

Cohen [11] identified a general objection to using audio: people in shared office environments do not want more noise to distract them. Buxton [12] argued that as audio is ubiquitous it would be less annoying if people had more control over it in their environments. Lessons from acoustic ecology would be helpful here.

1.2 Acoustic Ecology

The term *acoustic ecology* [13] comes from work begun by R. Murray Schafer in the 1960s as part of his *World Soundscape Project* at Simon Fraser University (see [14]). Schafer sees the world around us as containing ecologies of sounds. Each soundscape possesses its own ecology, and sounds from outside the soundscape are noticeable as not belonging to the ecology. In Schafer's worldview we are exhorted to treat the environments in which we find ourselves as musical compositions. By this we are transformed from being mere hearers of sound into active and analytic listeners – exactly the characteristic needed to benefit most from an auditory display. When the environment produces noises that result from data and events in the environment (or some system of interest) we are able to monitor by listening rather than just viewing. In regard to auditory display the term acoustic ecology means the internal ecology of the various sounds within the sonification. That is, we treat the sonification both as a real-world soundscape in its own right, the acoustic ecology of which is jumbled, and as part of the wider real-world soundscape in which it is situated. Again, its sonic components may sit uneasily within the acoustic ecology of the host soundscape.

---

[2] The system communicated information about elderly householders to relatives in remote locations.

Cohen [11] defined an acoustic ecology as "*a seamless and information-rich, yet unobstrusive, audio environment*". Kilander and Lönnqvist [4] tackled this problem in their FUSEONE and FUSETWO environments with the notion of a *weakly intrusive ambient soundscape*, or WISP. In this approach the sound cues for environmental and process data are subtle and minimally-intrusive[3]. Minimal- or weak-intrusion is achieved in Kilander and Lönnqvist's scheme by drawing upon the listener's expectation, anticipation, and perception; anticipated sounds, say Kilander and Lönnqvist, slip from our attention. For example, a ticking clock would be readily perceived and attended to when its sound is introduced into the environment (assuming it is not masked by another sound). As the steady-state of the ticking continues and the listener expects or anticipates its presence the perceived importance drops and the sound fades from our attention [4]). However, a change in the speed, timbre, or intensity of the clock tick would quickly bring it back to the attention of the listener. Intrusiveness can thus be kept to a minimum by using and modulating sounds that fit well with the acoustic ecology of the sonification's environment. The sonification is discriminable from other environmental sounds (either by deliberate attentiveness on the part of the listener, or by system changes to the sounds) yet is sufficiently subtle so as not to distract from other tasks that the listener (and others in the environment) may be carrying out. To increase the quality of the acoustic ecology further, Kilander and Lönnqvist used real-world sounds rather than synthesized noises and musical tones. They concluded that "*easily recognisable and natural sounds ...* [stand] *... the greatest chance of being accepted as a part of the environment. In particular, a continuous background murmur is probably more easily ignored than a singular sound, and it also continuously reassures the listener that it is operative*" [4].

The audibility of sonifications is an important factor and is tightly coupled to the issue of intrusiveness. The comprehensibility of sonifications depends on many factors including the production quality of the sounds, the quality of the playback system, and cultural and metaphoric associations. Many data require metaphoric or analogic mappings for audio representation as they do not naturally possess their own sound. The choice of metaphor may determine how learnable and comprehensible the mapping is. For example, Kilander and Lönnqvist found that the sound of a golf ball dropping into a cup was difficult for listeners to recognize "*except possibly for avid golfers*" [4] whilst the sound of a car engine was easy to identify. This highlights the fact that when using real-world sounds it is important to assess the cultural attributes of those sounds. Investigating musical tones for the monitoring of background processes Søråsen [15] found that sudden onset or disappearance of a timbre is easier to detect than changes in the rhythm and melody of that timbre. He concluded: "*changes within one single instrument should be very carefully designed to represent non-binary changes in state or modus*".

---

[3] Kilander and Lönnqvist actually used the adjective 'non-intrusive' to describe their sonifications. One could argue that this term is misleading as any sonification needs to be intrusive to some extent in order to be heard. Their term 'weakly intrusive' is more helpful and more accurate.

2 Applying Musical Æsthetics

Vickers and Hogg [16] argued that the æsthetics and composition approaches of electroacoustic and musique concrète[4] may potentially lead to great success in sonification design given their dependence upon the gestural encoding present in sounds. According to Smalley's [17] spectro-morphological[5] classification we hear the physical, gestural qualities in sounds, and these in and of themselves (though usually in combination with timbre and volume) carry sufficient information regarding movement, atmosphere, size, material quality, and so forth to offer information to the perceiver that serves to generate meaning akin to that generated by musical harmonic/tonal systems. It has the added advantage of being arguably less culture specific, that is it is not classical, or pop, or anything we already recognize – it is rather a system that is more open to reading than it is a musical style that is recognised as such. Criticisms of cultural imperialism are often raised when sonifications based upon tonal (or even atonal or Schoenbergian serialist) structures are presented at conferences. The potential offered by electroacoustic musical forms to avoid cultural stereotyping means much more serious consideration should be given to their use in auditory display research.

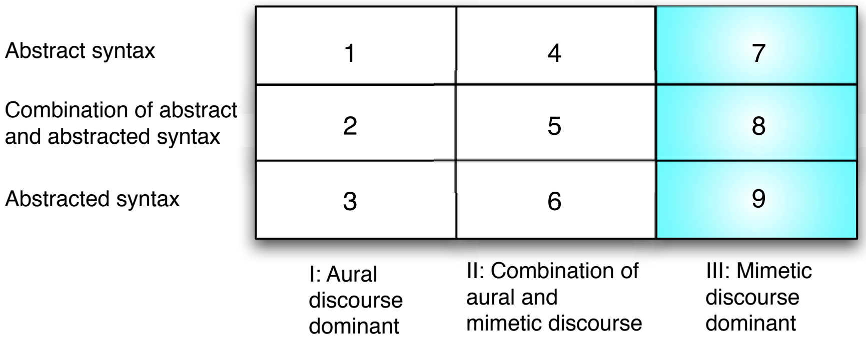
To help make the link between auditory display and electroacoustic music clearer, let us consider Emmerson's [18] *language grid* in which he classifies electroacoustic music into a nine-sector space on two axes: the level of syntactical abstraction of the music and the use of mimetic reference vs. aural discourse (see Fig. 1). Mimetic sound imitates or represents nature and aspects of human culture. In the language grid pitch-oriented music (both tonal and atonal) occupies grid sector 1 (abstract musical syntax with aural discourse dominant). We have shaded positions 7, 8, and 9 because they bound the region occupied by sonification. Sonifications are by definition mimetic in that their goal is to represent objects, events, or data of the domain in question. Some sonifications (sector 7) are grounded in the tonal music/melodic paradigm and so use *abstract* syntax[6] (e.g. Vickers and Alty's CAITLIN program sonification system [19] whilst others (sector 9) rely more on sounds *abstracted* from real objects (e.g. auditory icons [20]. In region 8 are those sonifications that use both abstract and abstracted syntax (such as Barra et al.'s WebMelody [21]). Sonification, then is concerned with mimetic discourse along the abstract-syntax/abstracted-syntax dimension.

---

[4] Musique concrète is a branch of electroacoustic music pioneered by Pierre Schaeffer (1910-1995) in which music is produced by editing together processed fragments of natural and industrial sounds.

[5] In musique concrète and electroacoustic music conventional pitched tones are only a subset of spectro-morphologies within a much broader world of spectra [17]. The reliance on architectures based around harmonic progressions of pitches is removed.

[6] In Emmerson's classification an *abstract* syntax is one in which the musical ideas have been organised and constructed independently from the sound materials. An *abstracted* organisation is one in which the music is abstracted from the sound-generating materials themselves. Thus, traditional music composition uses abstract syntax in which the notes on the score are related only to each other.



	I: Aural discourse dominant	II: Combination of aural and mimetic discourse	III: Mimetic discourse dominant
Abstract syntax	1	4	7
Combination of abstract and abstracted syntax	2	5	8
Abstracted syntax	3	6	9

Fig. 1. Emmerson's Language Grid [18]

2.1 Indexicality

Vickers and Hogg [16] introduced the idea of *indexicality* to sonification discourse. Indexicality is associated with mimesis as it is a measure of how strongly a sound sounds like the thing that made it. In sonification practice indexicality is related to whether sonifications make more use of direct data-to-sound mappings (high indexicality – the sound is derived directly from the data) or more use of metaphoric or interpretive mappings (low indexicality). Hayward's [22] auditory seismograms are an example of the former in which seismographic data were scaled and frequency-shifted until they lay in the human audible range. Vickers and Alty's [23] program sonifications are metaphoric: tonal musical motifs were used to stand for data and objects. From this, we can adapt Emmerson's language grid to the sonification domain to give Fig. 2 in which Emmerson's abstraction dimension has been retained but only the mimetic sectors have been carried over. In addition, the polarity of indexicality is indicated. Note that sonifications relying predominantly on abstract syntax (e.g. Vickers and Alty's program sonifications) possess lower indexicality than those making use of predominantly abstracted syntax (e.g. Hayward's seismograms).

There is a mapping, then, between sonifications and music compositions in that direct sonifications and concrete music possess high indexicality and metaphorical sonifications and abstract music possess low indexicality. That is, 'direct' (abstracted syntax) and 'metaphorical' (abstract syntax) in the sonification domain map to 'concrete' and 'abstract' respectively in the music domain. Musicologists would argue that music is as much a construct of the listener's mind as of the composer's; if the listener perceives something as music then it *is* music (though Smalley [17] claims that the listener must "*discover a perceptual affinity with its materials and structure*" in order for this to happen). Thus, Vickers and Hogg [16] offer us:

Lemma 1. $Sonification \implies Music$

Kramer [24] notes the similarity in structure between sonification and music creation: sonification renders data in sound to allow a human listener to detect

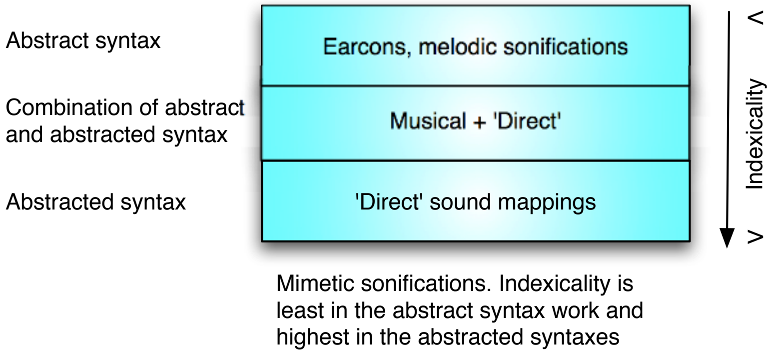


Fig. 2. Sonification Indexicality (adapted from Emmerson's *Language Grid* [18])

and comprehend patterns and structures in that data, whilst a musician renders a musical score so as to make it audible and thus make perceptible the music's structure and even give clues as to the composer's and the musician's emotional states. Thus, a piano is a sophisticated auditory display machine [16]: through the agency of the musician, the piano renders in sound (albeit in a highly complex and abstract way) the score, the technique of the musician, the physics of the piano, the emotional state of the musician and the composer, and even the musician's response to the feedback loop offered by his own ears. This gives us [16]:

Lemma 2. $Music \implies Sonification$

Thus, Vickers and Hogg [16] argue that sonification and music are mutually implicated and thus we get the logical biconditional:

Lemma 3. $Sonification \iff Music$

That is, if something is music then it is also a sonification and vice versa [16]. As an illustration of the link between music and sonification consider Barra et al's WEBMELODY system [21]. Drawing on the ideas of futurist composer Luigi Russolo (1885-1947), principles of Pierre Schaeffer's musique concrète, and inspired by Edgard Varèse's *Poème Électronique* (1958) and John Cage's aleatoric compositions (e.g. *Music of Changes* (1951)), Barra et al [21] tried to construct sonifications for monitoring a web server that were "*neutral with respect to the usual and conventional musical themes.*" They attempted to move away from the idioms of tonal and atonal (serialist) music and towards the more concrete compositions found in the musique concrète and electroacoustic traditions.

2.2 Acousmatics to Haptmatics

One device employed very successfully by WEBMELODY (and many other sonification systems) is *acousmatic sound*. The term was introduced by Pythagoras who reputedly taught his students whilst standing behind a screen[7]. Acousmatic sound, then, is that which one hears without the originating cause being visible to the listener. Auditory display research is replete with acousmatic sound as sonifications are often designed to highlight unseen or hidden data or events. We may also use haptic feedback to communictate information about unseen (or unheard) objects, data, or events (e.g. notification of incoming email). As this is an analogue of acousmatic sound we shall call it *haptmatic* sensation. An example of an existing haptmatic display technique is Brewster and Brown's [25] *tacton* – a haptic icon (c.f. auditory icons and earcons).

In fact, there is a very direct relationship between sonification and the kind of haptic feedback found in tactons and other similar approaches. Sound is simply a transverse wave with properties of frequency, phase, timbre, and amplitude. The vibro-tactile display offered by the tacton possesses similar properties – indeed, if a tacton emitter is placed on a resonant surface it becomes a loud speaker. Sound and touch are very closely related in our everyday experience, especially in the region of low frequency sounds which are easily conducted through floors and other surfaces. At night clubs we can feel the bass frequencies in our chest; on the street we can feel heavy lorries approaching. The relationship between touch and sound is not only in this direction. In the next section we look at the relationship between haptic (gestural) input and sonification.

3 Spectro-morphology and Haptic Input as Performance

Smalley's [17] classification of sounds via spectro-morphology provides us with an approach to sound and musical structures that focuses on the spectrum of available pitches and frequencies and how they are shaped (morphed) over time. In Smalley's model every sound contains gestural information that codes for the identity of movement, atmosphere, size, material quality, and so forth of the sound object in question. We are used to listening to sounds and decoding such gestural information. We can tell how fast a car is approaching us; we can estimate how hard a drum was hit and what size the drum is. Indeed, there is a very strong association between gesture and interpretation. This is especially true of music performance in which very complex and subtle gestural control is used to shape the music we hear and experience.

One of the challenges faced by audiences of electroacoustic music concerts is the lack of visible gestural interaction between the musician and the instrument. When the instrument is tape decks and laptop computers the familiar frame of reference of the direct relationship between a performer bowing strings, striking keys, moving sliders, and banging drums and the resultant sound is lost. It is replaced by a visible gesture set that often, at best, has a seemingly indirect relationship to the sonic experience and, at worst, no perceptual link at all (a reiteration of Eisenstein's contrapuntal sound, perhaps).

---

[7] Jérôme Peignot (1955) and François Bayle (1974) reintroduced the term in respect of musique concrète.

The ‘traditional’ musical instrument has been with us so long that it has become the natural mechanism for making music and we are well attuned to watching musicians physically play their instruments. We note that for every input gesture there is a corresponding output sound. The expert listener and observer notes the subtlest of gestures but even the neophyte can observe the link between gross gestures (such as strumming a guitar) and the overall sonic output.

How, then, does this inform us about the haptic control of interfaces that also use sonification? Spectro-morphology tells us that we listen for and can identify gestures in a sound. Experience shows us that we are used to associating physical gestures with consequent sound. If we are so sensitive to gesture, it might be that the haptic gestures become associated with the system’s auditory output whether they are directly related or not. Might the subtle gestures needed for fine control of systems such as Sensable’s Phantom[8] be translated in the mind of the listener as being somehow related to any auditory display? We have shown above Vickers and Hogg’s assertion that sonification and music are in a mutual implication relationship. If a sonified interface is thus a music playing system and the interface also provides gestural input in the form of haptic control, we suggest:

Lemma 4. *Auditory Display + Haptic Input = Musical Instrument*

That is, because of the ubiquitous frame of reference from musical instruments in which input gestures result in musical output we argue that a device/application interface that combines auditory display with haptic input might be viewed or interpreted as a musical instrument in the *mind* of the user. It does not necessarily matter that the sonification produced by the system is not related either directly or indirectly to the haptic input (though a lack of even an indirect relationship is questionable) as one would still naturally seek links between the gestural inputs and the sonic outputs. Where there is no direct relationship it becomes quite possible that artificial cause-effect relationships will be constructed in the user’s mind which may cause usability problems. A user moves a lever, grabs a virtual object, or makes a circular gesture to a mobile phone’s camera, and the system happens to emit a sonified data stream immediately afterwards, it is very possible that a causal link will be established in the user’s mind even if no link is present.

3.1 Haptic Input and Auditory Display

In other words, a musical instrument is a system in which manual gestures result (however visibly indirectly) in musical output. A multimodal system that combines haptic input with auditory display may thus be considered to be a musical instrument. If causal relationships between the haptic control and the auditory display are intended, then system designers might find it useful to draw lessons from the musical instrument design community for ways of improving the

---

[8] See `http://www.sensable.com/products/phantom_ghost/phantom.asp`

interaction[9]. If causal relationships between gestures and audio are not intended or desired (Eisenstein again), then interface designers also need to be aware of the dynamics of instrument design in order to try to 'design out' any user perceptions of causality that arise through familiarity with the musical instrument paradigm.

3.2 Haptic Output and Auditory Display

Where systems combine auditory display with haptic output, other possible interactions between the two modalities arise. We commented above about the relationship between sound and touch, how certain sounds also generate vibrations that are perceived physically. Again, the question arises whether causal relationships between the haptic and auditpory outputs are inferred by the user. Where no causal relationship exists such inference would likely be detrimental to the user's interpretation of the system's outputs. Again, the system might be considered a musical instrument.

4 Further Study

Researchers who have reported the most success with their sonifications also tended to deal directly with the issue of the æsthetics and acoustic ecology of their sonifications. As the role of æsthetics is increasingly entering the consciousness of designers of computing systems (e.g., see Fishwick [9]) so it needs to inform the work of the auditory display community. It has been proposed that sonifications be viewed as works of musical art as they could then benefit from the application of the æsthetic practices employed by artists [16][10]. When haptics are added to the mix there is a great potential for causal associations between the sound and the touch to be created in the user's mind. What is needed then, is research that explores the cognitive and artistic issues surrounding our haptic-sonification musical instruments. In what ways does haptic control affect the way we perceive an auditory display? If force feedback and sonification are used in tandem, does the brain automatically assert a link between the two data streams even when none is present? In the world of television and the movies, it is usually the case the the soundtrack affects the perceived meaning of the visual track and not the other way around. Will sound also have such an influence on our perception of force feedback?

It would be instructive to look at the semiotics of haptic-auditory interfaces. Whilst much has been written about the semiotics of visual and auditory messages less attention has been paid by semioticians to touch [26]. However, as musicologists and designers of new interfaces for musical expression[11] are very interested in the relationships between physical gesture and sound with the machine as intermediary [27], it may benefit the HCI community to temporarily cast

---

[9] Where direct links do exist between the input and the output, there is also scope for users playing the interface like a musical instrument just for fun.

[10] St. Augustine's Confessions act as a valuable cautionary tale against going too far down the art for art's sake route.

[11] See `www.nime.org`.

their multimodal interfaces as musical instruments and see what design lessons can be learnt by studying these systems as a music researcher might. *Speech Act Theory* [28] studies how people use language (rather than establishing the truth-value of statements) [26] and offers three dimensions of act: *locutory*, *illocutory*, and *perlocutory*. The locutory dimension deals with material aspects of an act's generation (e.g. strength of a physical movement). The illocutory aspects are to do with the intention behind an act. The percolutory dimension deals with the effect an act has upon the receiver. By drawing together the skills of the HCI practitioner, the music researcher, and the semiotician, we may be better placed to understand the locutory, illocutory, and perlocutory nature of haptic and auditory signals and thus able to explore the rich interactions and the acousmatic and haptmatic effects that will result in the new generation of multimodal systems.

References

1. Eisenstein, S., Pudovkin, V., Alexandrov, G.: Statement on sound. In Taylor, R., Christie, I., eds.: The Film Factory: Russian and Soviet Cinema in Documents, 1896-1939. Harvard University Press, Cambridge, Massachusetts (1988) 234–235
2. Hudson, S.E., Smith, I.: Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In: CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM Press (1996) 248–257
3. Gutwin, C., Greenberg, S.: Support for group awareness in real time desktop conferences. In: Proceedings of The Second New Zealand Computer Science Research Students' Conference, Hamilton, New Zealand, University of Waikato (1995)
4. Kilander, F., Lönnqvist, P.: A whisper in the woods – an ambient soundscape for peripheral awareness of remote processes. In: ICAD 2002 – International Conference on Auditory Display. (2002)
5. Sawhney, N., Schmandt, C.: Nomadic radio: Scaleable and contextual notification for wearable audio messaging. In: Proceedings of the CHI 99 Conference on Human factors in computing systems: the CHI is the limit, New York, NY, ACM Press (1999) 96–103
6. Sawhney, N., Schmandt, C.: Nomadic radio: speech and audio interaction for contextual messaging in no- madic environments. ACM Transactions on Computer-Human Interaction (TOCHI) **7**(3) (2000) 353–383
7. Pedersen, E.R., Sokoler, T.: Aroma: abstract representation of presence supporting mutual awareness. In: CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press (1997) 51–58
8. Somers, E.: A pedagogy of creative thinking based on sonification of visual structures and visualization of aural structures. In Brewster, S.A., Edwards, A.D.N., eds.: ICAD '98 Fifth International Conference on Auditory Display. Electronic Workshops in Computing, Glasgow, British Computer Society (1998)
9. Fishwick, P., ed.: Aesthetic Computing. MIT Press (2006)
10. Pedersen, E.R.: People presence or room activity supporting peripheral awareness over distance. In: CHI '98: CHI 98 conference summary on Human factors in computing systems, New York, NY, USA, ACM Press (1998) 283–284

11. Cohen, J.: Monitoring background activities. In Kramer, G., ed.: Auditory Display. Volume XVIII of Santa Fe Institute, Studies in the Sciences of Complexity Proceedings. Addison-Wesley, Reading, MA (1994) 499–532
12. Buxton, W.: Introduction to this special issue on nonspeech audio. Human-Computer Interaction **4** (1989) 1–9
13. Wrightson, K.: An introduction to acoustic ecology. Soundscape: The Journal of Acoustic Ecology **1**(1) (2000) 10–13
14. Schafer, R.M.: The Tuning of the World. Random House (1977)
15. Søråsen, S.: Monitoring continuous activities with rhythmic music. In: Proceedings of the Student Interaction Design Research Conference, Sønderborg, Mads Clausen Institute, University of Southern Denmark (2005)
16. Vickers, P., Hogg, B.: Sonification abstraite/sonification concrète: An ‘æsthetic perspective space’ for classifying auditory displays in the ars musica domain. In Edwards, A.D.N., Stockman, T., eds.: ICAD 2006 - The 12th Meeting of the International Conference on Auditory Display, London, UK (2006)
17. Smalley, D.: Spectro-morphology and structuring processes. In Emmerson, S., ed.: The Language of Electroacoustic Music. Macmillan, London (1986) 61–96
18. Emmerson, S.: The relation of language to materials. In Emmerson, S., ed.: The Language of Electroacoustic Music. Macmillan, London (1986) 17–39
19. Vickers, P., Alty, J.L.: Musical program auralization: Empirical studies. ACM Trans. Appl. Percept. **2**(4) (2005) 477–489
20. Gaver, W.W.: Auditory icons: Using sound in computer interfaces. Human Computer Interaction **2** (1986) 167–177
21. Barra, M., Cillo, T., De Santis, A., Petrillo, U., Negro, A., Scarano, V.: Multimodal monitoring of web servers. IEEE Multimedia **9**(3) (2002) 32–41 TY - JOUR.
22. Hayward, C.: Listening to the earth sing. In Kramer, G., ed.: Auditory Display. Volume XVIII of Santa Fe Institute, Studies in the Sciences of Complexity Proceedings. Addison-Wesley, Reading, MA (1994) 369–404
23. Vickers, P., Alty, J.L.: Siren songs and swan songs: Debugging with music. Communications of the ACM **46**(7) (2003) 86–92
24. Kramer, G.: An introduction to auditory display. In Kramer, G., ed.: Auditory Display. Volume XVIII of Santa Fe Institute, Studies in the Sciences of Complexity Proceedings. Addison-Wesley, Reading, MA (1994) 1–78
25. Brewster, S.A., Brown, L.M.: Tactons: Structured tactile messages for non-visual information display. In: Proceedings of Australasian User Interface Conference, Dunedin, New Zealand, Austalian Computer Society (2004) 15–23
26. McGee, K., Harup, A.: Contact expressions for touching technologies. In Mitchell, G., ed.: Proc. 3rd Conf. Computational Semiotics in Games and New Media, University of Teesside, UK, University of Teesside (2003) 68–76
27. Magnusson, T.: Screen-based musical interfaces as semiotic machines. In Schnell, N., Bevilacqua, F., Lyons, M., Tanaka, A., eds.: Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, IRCAM – Centre Pompidou (2006) 162–167
28. Winograd, T., Flores, F.: Understanding Computers and Cognition. Ablex Publishing Corp, Norwood, NJ (1986)

Navigation and Control in Haptic Applications Shared by Blind and Sighted Users

Eva-Lotta Sallnäs, Kajsa Bjerstedt-Blom, Fredrik Winberg,
and Kerstin Severinson Eklundh

Department of Human-Computer Interaction,
Royal Institute of Technology, S-100 44 Stockholm, Sweden
{evalotta, kblom, fredrikw, kse}@csc.kth.se

Abstract. Haptic feedback in shared virtual environments can potentially make it easier for a visually impaired person to take part in and contribute to the process of group work. In this paper a task driven explorative evaluation is presented of collaboration between visually impaired and sighted persons in three applications that provide haptic and visual feedback. The results show that all pairs could perform all the tasks in these applications even though a number of difficulties were identified. The conclusions made can inform design of applications for cooperation between visually impaired and sighted users.

1 Introduction

In the study presented here, cooperative interaction between visually impaired and sighted people is investigated in a shared virtual environment with haptic and visual feedback. The focus in this evaluation has been on how users manage to explore these types of environments together, finding reference points and objects, identifying objects and their properties, discriminating between objects and handing off objects to each other. Three different virtual environments that provide visual and haptic feedback were used in this evaluation. In these environments different kinds of collaborative functions have been implemented such as one proxy for each user, grasping of objects, hand over of objects between users, haptic guiding of one another and lifting of objects together. The aim of this study is to investigate how a sighted and a blind person can interact with the three virtual environments that provide visual and haptic feedback. Furthermore, the aim was to investigate how a sighted and a blind person could use these environments in order to solve four tasks together.

2 Background

Not even half of the population of adult visually impaired persons that could work in Sweden have an employment [8]. To have an employment and to work is not just important economically, it is equally important for building social relations and for the feeling of self-esteem. Even if integration in society has improved, reports show

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 68 – 80, 2006.

that visually impaired persons often feel excluded from both information regarding the society and cultural events [9]. The ability to function in groups is generally becoming more and more important in society. It is therefore crucial to investigate how computer interfaces should be designed to make it possible for visually impaired and sighted persons to contribute equally well in collaborative situations.

A number of studies have investigated different interaction techniques including tactile and audio feedback in order to support visually impaired people using computer interfaces individually [11, 12]. One investigation has been performed that investigated collaboration between blind and sighted users in auditory interfaces [10]. In that study a blind user interacted with an audio interface and the sighted with a graphical interface. Special attention was in this study put on observing turn taking, awareness, communication, and error recovery. The study showed that both participants could form a mental understanding of the contents and the layout of the environment. They could both refer to specific objects and form a common ground regarding what happened in the environment during collaboration and they could both take active part in the problem solving.

A number of authors have investigated issues regarding sighted users' joint manipulation of virtual objects in a haptic collaborative virtual environment [1, 2, 3, 5, 6]. These studies have shown that haptic feedback in general makes joint manipulation of objects more efficient and that perceived presence in the virtual environment and perceived performance increases with haptic feedback. The results regarding the feeling of social richness in interaction measured by the concept of social presence is a bit more complex. Haptic feedback increases perceived social presence when participants do not communicate verbally but this effect is overshadowed when they communicate by telephone [7].

This earlier research serves as a point of departure for starting to also investigate collaboration between sighted and a visually impaired persons in haptic environments. The motivation for these kinds of studies is to accumulate knowledge on how to design interfaces that can make it possible for visually impaired persons to contribute in work and learning activities where group work is required that involves graphical information.

3 Aim of the Evaluation

A study was performed in order to investigate how a sighted and a blind person can interact with three shared virtual environments that provide visual and haptic feedback. Furthermore, the aim was to investigate how a sighted and a blind person could use these environments in order to solve four tasks together.

4 Method

The study presented in this paper was a task driven explorative evaluation of collaborative task performance in three different virtual environments that provided haptic and visual feedback. The methodologies used in this study were observation of video recordings, focusing on the dialogue between the collaborating participants and their interaction in the shared virtual environment, and semi structured post-test interviews with pairs of participants [4].

The observation technique was explorative as the situation investigated in this study is novel. The observation of the video recordings was thus not guided by predefined categories, but categories were instead defined as a result of the analysis. The reason for not making any quantitative analysis was that the situation investigated is novel and therefore valid hypotheses are hard to formulate. Furthermore, the sample of participants is very heterogeneous and the study was not comparative which makes quantitative data less interesting.

In the semi-structured interviews 26 questions addressed the dimensions collaboration, communication, situation and social awareness, interface design, spatial orientation and how haptic feedback affected the interaction in the environments. Furthermore, 8 questions enquiring about background information was used.

4.1 Participants

Six visually impaired and six sighted adults participated in the laboratory test. The age was between 41-50 for three, 51-60 for two and 21-30 for one of the visually impaired participants. All the sighted participants were between 21-30 years old. Two participants were visually impaired since birth, one could differentiate between light and darkness and the other could also differentiate between strong colors. Three participants were blind since childhood, and one participant had become blind when adult. Only one of the sighted and one of the visually impaired participants had previous experience of using interfaces with haptic feedback. Those two persons were put in one pair.

4.2 Apparatus and Software Environments

The haptic feedback system that each person used in this study was a desktop Phantom device with a Reachin display system (Figure 1). The two haptic systems were connected serially to one computer.

The sighted person had stereovision through Stereographics CrystalEyes 3 shutter glasses. Touch and vision were co-located for the sighted person. The navigation and interaction techniques used were the one-point haptic feedback in all of the environments. This made it possible to feel objects, the walls, floor and ceiling when such were present in the environments.

The collaboration during the evaluation was video recorded with two video cameras. One camera recorded what happened in the interface whereas one video camera recorded the movements of the participants. The dialogue between the participants was recorded with both video cameras, whereas the interview was recorded with one of the cameras. One interview failed to be recorded but the notes from this interview were used in the analysis.

Fig. 1. The two test participants sitting in front of one Reachin Display system each

Design Environment. In the design environment users could collaboratively identify and explore objects and object properties. In task one, this environment contained one sphere and one cube. In task two, it contained three differently sized cubes and a screen that was built up out of four flattened cubes (Figure 2).

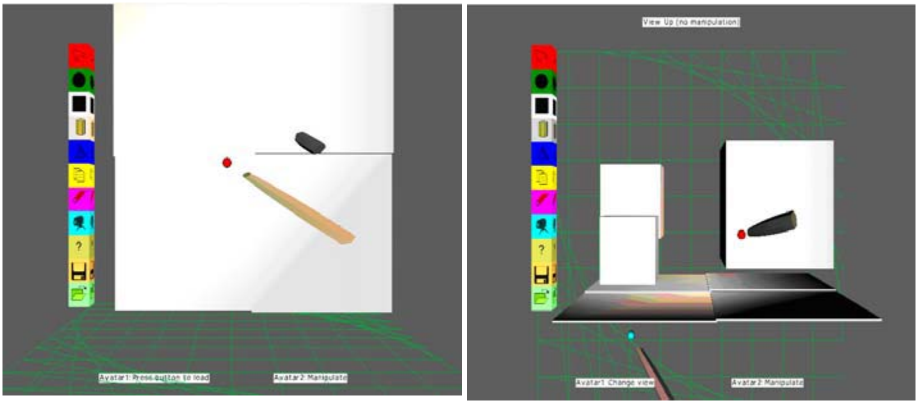


Fig. 2. The left picture shows the design environment viewed from the front and the right picture shows it viewed from above

The test participants were able to feel the geometrical shape of objects and their tactile properties such as surface friction, size and softness. Each participant had a proxy in the environment that they investigated the environment with. The sighted participant could also see a visual representation of the shared environment.

The experimenter could in this application build a scene by using a haptic menu. The participants did not use this menu in the study. When building a scene the experimenter could push at the symbol of a shape on the menu and then place the object anywhere in the virtual space. It was also possible to save a scene, and to upload an old scene. An object could be deleted by first pushing the button with the eraser on the menu and then pushing at an object. The experimenter could also change an object's size, surface friction and softness by moving a slide bar that could be activated through the menu bar.

Box-Moving Environment. In the box-moving environment the participants could jointly lift objects. In this application there were eight boxes that moved quite easily on the floor that was perceived as rather slippery (Figure 3).

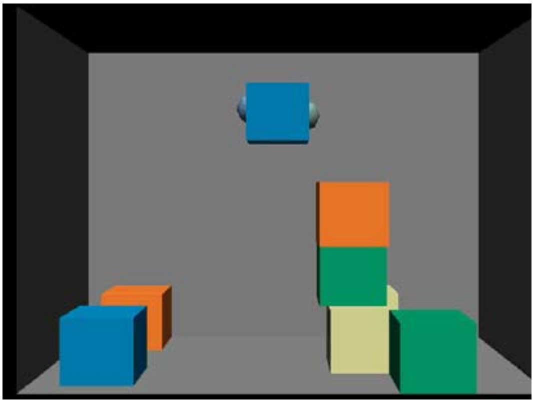


Fig. 3. The box-moving prototype in which two participants push from each side of a cube in order to lift it

The test participants were represented in the environment by one blue and one green sphere respectively. The participants could in this environment feel collisions between objects and could lift a cube by pushing from each side of it and lift it. The participants then felt the collaborator's forces on the object. It was also possible to feel the shape of the other person's proxy and to hold the other person's proxy by pushing the button on the Phantom pen. In this way participants could "shake hands" or guide each other. This environment also included gravitation that made it possible to feel the weight of the cubes.

Hand-Off Environment. In the hand-off environment users could grasp an object and hand it to another user in the shared environment. The test participants were as in the previous environment represented by one blue and one green sphere respectively. The application consisted of four shelves and four boxes that sat on the top shelves (Figure 4).

In this environment users could experience the force from collisions between objects and also gravitation forces. It was also like in the other environments possible to feel all the shapes in the virtual environment.

What makes this application special is that participants could grasp a cube themselves by placing their proxy at the cube and pushing the button on the Phantom pen. Then the participant could lift the object and hand it off to the other participant. When both participants held the object they could both feel the other person's pulling forces through the object until someone let the cube go. In this way it was possible to know when the other participant had the cube securely before letting it go.

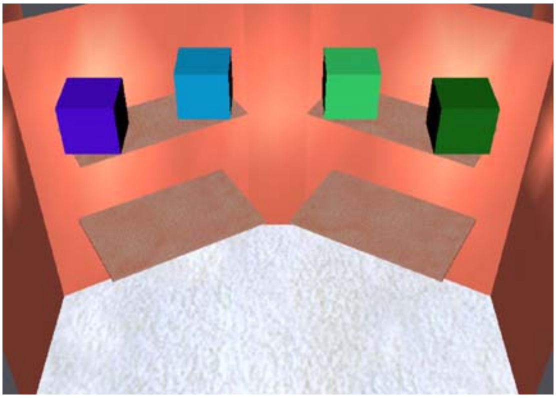


Fig. 4. The hand-off environment with four cubes and four shelves

4.3 Tasks

Task number one was to find one cube and one sphere in the first design application and to decide together which object that was the hardest, which was the softest, which had most surface friction and which had the least surface friction.

Task number two was also performed in the design application but the scene was different. The task was to identify the shape of three objects that were hidden behind a screen and to decide together which object was the largest.

Task number three was to take turns to hold a cube against the floor in the box moving application while the other person explored the shape. Then the participants should push from each side of a cube and lift it. Finally, the participants should hold on to each other's proxies and "shake hands" virtually in the environment.

Task number four was performed in the hand-off environment. The participants were instructed to first locate the shelves and cubes. After that, one participant was instructed to take one cube from the upper left shelf and give it to the other participant that should put it on the lower right shelf. Then the person that placed the cube on the lower right shelf should take one cube from the upper right shelf and give it to the other person that should place this cube on the lower left shelf.

4.4 Procedure

The participants were placed in the same room during the tests. A paper screen was placed so that the sighted person could not see the visually impaired person (Figure 1). The participants were able to talk to each other during the test.

Before the test the participants explored one of the applications in order to get acquainted with the visual/haptic environment. Before each task the experimenter gave a verbal description of the environment and a task description to both subjects at the same time. The test sessions were video recorded for later analysis. Performing the four tasks took between 30-40 minutes for all pairs of participants.

The test session ended with interviews with the two users in a pair together and they lasted between 30 and 60 minutes. The interview questions included 26 questions about the joint task performance, 6 questions about background information and finally 2 questions that were only asked to the visually impaired person. The interview was recorded with a video camera.

4.5 Analysis

The video recordings of the interaction in the shared visual/haptic virtual environments including the communication between the participants were analysed. Annotations were made of each evaluation session making notes describing the users' interaction with the interface and with each other for each pair respectively. Interpretations of interesting parts in the interaction were also annotated in parallel to the descriptions in a separate column. The annotations from each pair were analyzed in three iterations where the data was interpreted and reduced. Finally, the data was categorised and the data from each category from each evaluation session was compared in order to derive general findings.

The interviews were transcribed in their entirety but neither expressions of emotion nor pauses were transcribed. The data was subsequently reduced in two iterations so that only information relevant for the dimensions used in the interview was further analysed. Annotations of interpretations were made for each meaningful unit of the data material and a number of categories were defined. The data was analysed in this way for each pair and finally all pairs were compared in order to obtain general findings and interesting patterns in the results as well as unique but yet informative findings.

5 Results from Observations

The overall result showed that all participating pairs managed to perform the four tasks even if they encountered a number of difficulties when interacting with the interfaces.

5.1 Verbal and Haptic Guiding

All participants guided or were guided by the other person in the virtual environment. In most cases the sighted person took command and started to guide the blind person verbally. In some cases however, the blind persons first investigated the environment themselves before the pair began to solve the task. Guiding was mostly done verbally by the sighted person that made deictic references like “go right”, “left”, “up” or “backwards”. In a number of instances the haptic force feedback was used as guiding behaviour. One example of this was when one pair performed task one. The sighted person then followed the visually impaired person's graphical representation and was exploring what the visually impaired person explored in order to feel what the she felt, so that the sighted person could more easily explain what she meant regarding the issue they where discussing. Another example was when a pair performed task number three. In that situation the pair had built a tower and the sighted person grasped the visually impaired person's graphical representation and haptically guided the visually impaired person around the shape of the tower. In the same task, the sighted person in another pair used haptic feedback in order to get the visually impaired person's attention and to show direction. In that case the sighted person then bumped into the visually impaired person's graphical representation on purpose and at the same time said "come" and moved away. The visually impaired person's graphical representation then followed in the sighted person's direction. Haptic feedback was used for coordination in subtle ways, which means that not so many details were needed in verbal descriptions. Participants could adjust what they expressed verbally based on what they felt or saw and what they knew that the other person was experiencing like for example that they could feel that they found an object.

In one situation the visually impaired person guided the sighted person verbally. This happened when a pair performed task number two. The sighted person was in that situation placed in front of the screen behind which the objects that the pair was supposed to explore were hidden. Because of this the sighted person did not find the objects. The visually impaired person then explained what the objects felt like and how they were placed spatially in relation to the screen. The sighted person finally understood that the objects were behind the screen and then they could solve the task together. The collaborating pairs in this task had a more equal work situation because of the fact that the sighted persons had to rely more on their tactile sense than in the other tasks.

5.2 Navigation and Spatial Orientation

When the pairs had explored the environments they learnt how the objects were arranged in the environments and if they were movable or fixed. Both the visually impaired and the sighted persons seemed to form a mental representation of the environments. This was also expressed later in the interviews.

Participants used deictic referencing during navigation in the environments, with expressions like "left", "right", "forward", "up" and "towards me". This was however not unproblematic as a lot of misunderstandings occurred before they had established a common ground and agreed on what concepts to use. One common misunderstanding was that the sighted person said "go forwards" whereby the visually impaired person moved into the environment, that was forward relative to the person's body centre. The sighted persons understood the problem after a while and instead said "towards you with the pen" or "no, forwards in the image".

Another problem with developing a common ground during the test was that the participants knew that they were sitting opposite to each other physically in the test room. Furthermore, the visually impaired persons did not always know how they were placed in relation to the other person in the virtual environment. This made it hard when for example the sighted person said that the visually impaired person should go towards him. The sighted person meant towards him in the virtual environment but the visually impaired person sometimes then moved towards the sighted person's physical location in the test room. It was easier for the participants to refer to objects in the environment. Then they for example said "this one to the left", "number two" or "the big one".

5.3 Joint Manipulation of Objects

Joint manipulation of objects was possible in the environments by the sighted and the visually impaired person. When the participants handed over objects to each other in task four they could feel the force feedback from the other person through the object that that person was holding with the Phantom pen. However, the coordination of the handover was mainly verbal in combination with haptic feedback. When one pair was performing task four, the visually impaired person first explored the environment by her self without any guiding of the sighted person. The visually impaired person (V) was the whole time talking aloud about what she was doing, like a manifestation of her perception of the environment. When V found a shelf she said that "this is the lower right shelf". Then she said "then I go to the upper shelf" and moved straight up and hit the intended shelf. V then took a cube with her Phantom pen and lifted it from the shelf and dropped it. V asked where she put it and the sighted person said that it was on the lower shelf whereby V went directly to that shelf and checked this. When they started to perform task four V took the initiative and she knew exactly where the shelves were and knew where to put the cube. When the sighted person was going to hand over a cube to V he verbalized what he was doing: "I take the cube now and am on my way towards you" "Now I push on you a little with the cube, do you feel that? Tell me when you have it". This illustrate how verbalization of behaviour was combined with haptic feedback in order to coordinate the joint handling of objects. One visually impaired person said to the sighted participant "…now I should be on the lower left shelf and there should not be any blocks there. But then you also said that but I just wanted to check it". This example shows that verbal communication between the participants was important to support awareness of the status of the shared environment.

6 Results from Interviews

6.1 Haptic Mental Representation

Fixed reference points like walls, floor, ceiling and fixed objects like shelves are important in an environment to support orientation. The blind participants reported that they were able to form a haptic mental representation of the environment after some exploration. In the environments with gravity, the weight of the objects and the fact that a dropped object ended up on the floor underneath the position of the user made it easier to control the objects and find them again when they were lost.

Even though the blind participants thought that getting a pre-understanding of the context from a verbal description is very important, they thought that the verbal descriptions did not correspond very well to their haptic experience of the environment. The communication with the co-worker was a great help when forming a mental representation of the environments. The blind participants argued that they could form a good mental picture of an environment but that they needed a bit more time to explore the interface. They sometimes felt a bit slow compared to the sighted person. One visually impaired participant said that “ one can do what everyone else can do but much slower”.

There was a problem with perspective in the environments. The visually impaired participants said that angles and inclinations were problematic because they sometimes did not feel natural or realistic and there were very few 90-degree angles. It was sometimes problematic to know the difference between walls and the floor and in one environment there was no ceiling. The participants said that it is better to have restricting surfaces in all directions. Another problem was to understand the perspectives because it was hard to know if the perspective was from above like looking down into a box or from the front like looking into an aquarium.

6.2 Referring to Objects

Referring to objects verbally was relatively easy. This had to do with the fact that both sighted and blind participants felt that they got a good mental understanding of the content and layout of the environments. Participants reported that they discussed how the different objects felt and where they were placed in the environment. The blind persons said that they had a good understanding of the changes in the environment of the objects that were relevant for the task.

One difficulty was however, that the haptic hardware volume limits sometimes were mistaken for being an object in the environment. If the environment does not have restricting surfaces all around, the volume borders of the environment become confusing for the blind person and create problems in the discussion between the sighted and the blind person.

The participants reported that they mostly coordinated their actions by talking about what they were doing but they also used haptic feedback for coordination and communication a number of times. One pair said for example that they used haptic feedback when handing over a cube in order to know if the other person really had a good grip of the cube.

6.3 Referring to Direction

A problem that both the sighted and the blind persons experienced was that it was hard to talk about directions in the environments. One reason for this problem was that the participants sat almost facing each other physically in the laboratory while they at the same time were talking about directions in the environment. The blind person wondered if they were facing each other in the virtual environment also which they in fact were not. In the virtual environment they had exactly the same view just as if they were sitting beside each other. It was also hard for the blind person to understand what the sighted person meant when she said “go backwards”. For the blind person that meant moving the hand with the Phantom towards himself whereas the sighted person meant the opposite, which is moving the hand forwards so that the proxy moves more into the virtual environment.

6.4 Perceived Presence

The participants reported about their shared experience of solving the tasks together in the environment as if they had both experienced approximately the same place. They referred to situations in the interaction that both of them remembered and they understood what objects and behaviour the other person was describing. One pair reported that they experienced a high degree of social presence in the environments and one person said that “you really are in the same environment and that you are actually close to each other”. They also said that they almost wanted to say I am sorry when pushing another person by accident. Another pair said that the environment felt realistic. The visually impaired person reported the following experience ”there was one situation when we should lift blocks and I then suddenly realized that it was so realistic, at the same time, that we should lift the block upwards we had to push it to the right, kind of sideways, and then it was the person standing on the left side that had to push to the right. I could not drag the block, I had to push against it so that it did not slip from our hands”.

7 Conclusions and Discussion

It is possible to derive a number of important conclusions concerning collaboration between a sighted and a visually impaired person in shared haptic/visual environments from the results of this study. These should be considered when designing applications in the future that aim at supporting joint interaction by a visually impaired and a sighted person in environments similar to the ones tested. Generally it is a positive result that all participating pairs managed to perform the tasks in the study together without any major breakdowns. It made it possible to get a multifaceted picture of the problematic aspects as well as the successful design features when analysing the interaction.

The most important conclusions that can be drawn for design of environments for collaboration between visually impaired and sighted are the following:

- Verbal communication between the participants is important to support awareness of the status of the shared environment and for coordinating joint activities.
- Haptic feedback is important in order for the visually impaired persons to form a mental understanding of the layout of the environment and of the objects in it.
- Referring to objects was relatively easy because users experienced that they had a shared mental understanding of the content and layout of the environments.

- It was hard to talk about directions in the environments when both participants were in the same room facing each other. Referencing thus depends on the physical location of the participants during the interaction in the virtual environment.
- Joint manipulation of objects was possible and haptic and verbal feedback was used in order to coordinate joint handling of objects.
- It is better to have restricting surfaces in all directions otherwise haptic hardware volume limits are sometimes mistaken for being an object in the environment.
- Predictable behaviour that mimics nature like gravity makes it easier to control objects.
- Fixed reference points like walls, floor, ceiling and fixed objects like shelves are important in an environment in order to support orientation.

We argue that the results from this study show that it is possible to develop interfaces that can support a shared work process between sighted and visually impaired users. It is clear that the users perceived the environments used for collaboration in this study to be truly shared environments that supported a common ground regarding the task that made joint problem solving possible. The results from this study also show, however, that there are a number of problems remaining to be solved.

In this study it becomes apparent that haptic feedback can serve as a substitute for the visual mode to the degree that visually impaired people can orient themselves in the 3D space and with the objects that are in it. If the environment is very dynamic however and the objects are moving so easily that the user does not have the time to explore them the spatial orientation is suffering.

It is possible for visually impaired persons to coordinate fairly complex tasks such as lifting a virtual object and giving it to a collaborator, as well as taking an object handed to them and placing it on a specific shelf. We have to remember that the only input the visually impaired person gets is one point haptic feedback. The verbal communication is of great importance and can be seen as a valuable resource in this collaborative situation.

This study show that the effects of the physical context cannot be ignored as it is very much integrated in the collaborative experience, especially for the blind person. The physical orientation of the two co-located participants had an effect on the words that could be used about direction and navigation in the environment. Move towards me meant moving towards the sound of the sighted person for the visually impaired person instead of moving towards the sighted person’s proxy in the environment. Move backwards meant for the visually impaired participant to move towards his stomach but it meant moving into the 3D environment for the sighted person. The physical constraint of the input device volume was also perceived as objects by the visually impaired participant.

In this study qualitative results have added knowledge about the usability aspects of adding haptic feedback in interfaces that visually impaired and sighted share with the aim of performing tasks together. This complements the quantitative results about increased efficiency and precision achieved by haptic feedback for sighted persons in other studies [1], [2], [3], [5], [6].

References

1. Basdogan, C., Ho, C., Srinivasan, M. A., & Slater, M. (2000). An Experimental Study on the Role of Touch in Shared Virtual Environments. *ACM Transactions on Computer-Human Interaction*, **7**(4): 443-460.
2. Jordan, J., Mortensen, J., Oliveira, M., Slater, M., & Srinivasan, M. (2002). Collaboration in a Mediated Haptic Environment. *The 5th Annual International Workshop on Presence. Porto, Portugal.*
3. Oakley, I., Brewster, S., & Gray, P. (2001). Can You Feel the Force? An Investigation of Haptic Collaboration in Shared Editors. In C. Baber, M. Faint, S. Wall & A. M. Wing (Eds.), *Proceedings of Eurohaptics 2001,* Birmingham, pp 54-59.
4. Patel, R. & Davidsson, B. (2000). *Forskningsmetodikens grunder.* [The foundations of research methodology]. Studentlitteratur. Lund
5. Sallnäs, E-L., Rassmus-Gröhn, K., & Sjöström, C. (2000). Supporting Presence in Collaborative Environments by Haptic Force Feedback. *ACM Transactions on Computer-Human Interaction,* **7**(4): 461-476.
6. Sallnäs, E-L., & Zhai, S. (2003). Collaboration Meets Fitts' law: Passing Virtual Objects With and Without Haptic Force Feedback. In M. Rauterberg, M. Menozzi & J. Wesson (Eds.), *Proceedings of INTERACT´03,* Amsterdam, IOS Press, pp 97-104.
7. Sallnäs, E-L. (2004). The Effects of Modality on Social Presence, Presence and Performance in Collaborative Virtual Environments. Ph.D. thesis, Royal Institute of Technology, Stockholm.
8. SRF – Synskadades riksförbund. [The swedish association of the visually impaired]. (2004-12-15).
   http://www.srfriks.org/VVILLSRF/SRFarmbmarknprogram/flersynskadade.htm
9. Szabo, M. (1989). Synskadades Villkor i Samhället. [The Situation in Society for the Visually Impaired]. In A. Hansson & M Szabó (eds.), Vit Käpp Förstoringsglas Punktskrift, Synskadade Berättar. [White Stick, Magnifying Glass, Braille, Visually Impaired People Report]. Nordiska museets förlag, Helsingborg
10. Winberg, F. & Bowers, J. (2004). Distilling Knowledge: Assembling the Senses: Towards the Design of Cooperative Interfaces for Visually Impaired Users. *Proceedings of ACM CSCW*, Chicago, Illinois, USA, pp 332-341.
11. Yu, W., Cheung, K. & Brewster, S.A. (2002). Automatic Online Haptic Graph Construction. *Eurohaptics 2002*, Edinburgh, UK, pp128-133.
12. Yu, W. & Brewster, S.A. (2002). Comparing Two Haptic Interfaces for Multimodal Graph Rendering. *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (Haptics 2002)*, Florida, USA, IEEE, pp 3-9.

User Evaluations of a Virtual Haptic-Audio Line Drawing Prototype

Kirsten Rassmus-Gröhn, Charlotte Magnusson, and Håkan Eftring

Certec, Design Science, Lund University
PO Box 118, S-221 00 Lund, Sweden
{Kirre, Charlotte.Magnusson, Hakan.Eftring}@certec.lth.se
http://www.certec.lth.se/haptics

Abstract. A virtual haptic-audio drawing program prototype designed for visually impaired children, has been gradually developed in a design-evaluation loop involving users in four stages. Three qualitative evaluations focused on recognizing drawn shapes and creating drawings have been conducted together with a reference group of 5 visually impaired children. Additionally, one formal pilot test involving 11 adult sighted users investigated the use of a combination of haptic and sound field feedback. In the latter test the relief type (positive and negative) was also varied. Results indicate a subjective preference as well as a shorter examination time for negative relief over positive relief for the interpretation of simple shapes such as 2D geometrical figures. The presence of the position sound field with a pitch and stereo panning analogy was not shown to affect task completion times.

1 Introduction

The present paper describes preliminary tests with a prototype of a virtual haptic-audio drawing application for low vision and non-vision users. The purpose of this application is to allow visually impaired users to create and access graphical images. The application is and will be developed in close collaboration with a user reference group of five blind/low vision school children. The objective of the prototype is twofold. During the early development stages, it is used as a research vehicle to investigate user interaction techniques and do basic research on navigation strategies and helping tools. Later, the prototype will be tailor-made for use in schoolwork and the final application should be possible to use in different school subjects.

Getting access to 2D graphics is still a large problem for users that are severely visually impaired. Using a haptic display in combination with audio feedback is one way to enable access. There are many issues to address, e.g. how to provide an overview, to what extent users are able to interpret a combination of lines or line segments into a complex image, how to design the lines to get appropriate haptic feedback, what hardware to use etc. General guidelines to create and develop haptic applications and models are collected in [1]. Applications making practical use of non-spoken audio and force-feedback haptics for visually impaired people are e.g.

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 81 – 91, 2006.

applications supporting mathematical display [2], [3] & [4], games [5-7] and audio-haptic maps [5;8].

There are few tools that enable blind people to create computer graphics. As described in [9] and [10], there are indeed people who are blind who have an interest in hand drawing. In [11], a CAD application is presented that enables users to create drawings with the help of audio and keyboard. This is accomplished by a structured approach of dividing a drawing into small parts and to enable the user to draw small segments of a drawing. In [12], a study on a haptic drawing and painting program is presented, and that work can be seen as a pre-study to the work in this article.

2 Interface and Equipment

The presented prototype makes it possible to make black & white relief drawings and tries to incorporate improvements suggested by [12]. The Reachin 4 beta software is used to control the haptic device, which can be either a PHANToM OMNI or a PHANToM Premium. The sound control is based on the FMod API.

The application consists of a room with a virtual paper sheet, which a user can draw a relief on. The virtual paper is inscribed in a limiting box. When the PHANToM pen is in touch with the virtual paper the user can draw on it while pressing the PHANToM switch. The haptic image is produced as positive or negative relief depending on which alternative is selected. The relief height (depth) is 4 mm. The drawing can be seen on the screen as a grayscale image – a positive relief is seen as black, and a negative relief is seen as white. The paper color is grey.

At the time of the test the haptically touchable relief was updated every time the user released the switch on the pen. It was at the time a problem to let the user feel the exact line that was drawn, since it caused instability in the PHANToM. However, this problem has been fixed (May 2005) partly by including GDI+ software and it is now possible to feel the line drawn (except the last segment).

To enhance the user's perception of localization a sound field was added. When the cursor moves in the virtual room, the pitch of a position tone is changed, brighter upwards, and mellower downwards. The mode information is conveyed by the volume and timbre of the tone. In free space, a pure sine wave is used. When the user is in contact with the virtual drawing paper (not pressing the PHANToM switch) the volume is louder. And when the user is drawing (and thus pressing the PHANToM switch) the tone is changed to a saw-tooth wave. Also, to differentiate the walls of the limiting box from the virtual paper a contact sound is played when the user hits a wall with the PHANToM pen

Drawings can be saved in png format, and a png import function has also been implemented. The files imported must be grayscale and a multiple of 256*256 pixels. A complete grayscale is actually translated into different relief heights, which makes it possible to import any grayscale image and get some haptic information from it. Images not adapted to haptic/tactile reading for blind users are very hard to understand, however, the grayscale can also be used e.g. to smooth out relief lines.

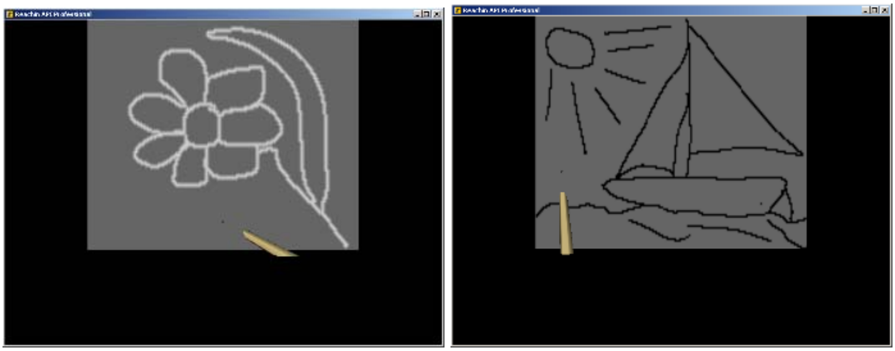


Fig. 1. Screenshots of the drawing program in negative relief to the left and in positive relief to the right

3 User Evaluations

The application has been gradually developed since 2005. It has been continuously evaluated by a reference group of 5 school children, aged 10 to 16. Two (2) of the participants are blind from birth, and three (3) participants have various forms of low vision. All of them read Braille and are integrated in normal schools in southern Sweden.

In March 2006, the application was tested in a formal pilot test with 11 sighted adults (aged 25 – 72). The users made the test without visual feedback from the screen.

3.1 Qualitative Evaluations with Reference Group with Children

The reference group has (to date) used the drawing program at 3 different group meetings. Not all of the reference group participants have been present at every meeting. Design work has been iterative and the users have been presented with new features and changes in the prototypes at every meeting. All three evaluations have been qualitative. The first two evaluations were of an informal nature, with few and loosely formulated test tasks. Instead an open discussion took place in which children and their parents or other close relations and the researchers discussed topics triggered by the prototypes tested. The third evaluation also incorporated some formal test tasks. During these tests, drawing has been tested with and without audio feedback, with positive and negative relief and with program interaction governed by virtual haptic buttons and keyboard commands. We have used the PHANToM Omni and Premium models for our tests. The tasks tested are summarized below:

- Draw and feel lines in negative and positive relief (haptics).
- Use vertical or horizontal work area.
- Change the relief using virtual button or keyboard button (haptics).
- Draw and feel an image (haptics + audio).
- Draw a specified shape – a rectangle and an Arabic number (haptics + audio).

- Explore and recognize Arabic numbers (haptics + audio).
- Explore and recognize two simple geometric shapes (haptics + audio)

3.2 Formal Pilot Study with Sighted Adults

A formal pilot user test was conducted on the latest prototype version with complete sound field mapping to pitch and pan, vertical drawing paper, limiting box and png file save, whose features are described in more detail above (section 2). The PHANToM Premium was used for this test, and a pair of headphones was used when sound field feedback was available. The users were presented with different recognition tasks and were asked to describe or reproduce the experienced relief drawings in different ways either using the drawing program itself, pencil and paper or verbal description (table 1). Four different test cases (see table 2 and 3) were designed to overcome the learning effect bias in the test. The users were also asked to rate the difficulty of the recognition tasks on a scale from 1 to 5, where 1 means least difficult, and 5 means most difficult.

In total, there were 16 test tasks regarding recognition and reproduction of 2D geometric figures in outline and abstract or mathematical curves with one or two parts (fig 3). The complex tasks were symbols from road signs either in positive or negative relief (fig 4). For each task the time to examine figures was measured. When the user considered himself/herself ready with the examination the time was stopped and the user was free to use as much time as he/she needed to reproduce the task figures. For the road sign recognition, the user was also asked to point out the test task road signs among a collection of road signs with similar features.

Table 1. Reproduction methods for test tasks

Task 1 geometry	Task 2 curves	Task 3 geometry	Task 4 curves	Task 5 road signs	Task 6 road signs
Verbal descr.	Drawing-program	Verbal descr.	Drawing-program	Pencil & paper	Pencil & paper

Table 2. Test cases for formal pilot test, sound field variations

Test case	Task 1 2D geometry	Task 2 curves	Task 3 2D geometry	Task 4 curves	Task 5 road signs	Task 6 road signs
1, 2	No sound	No sound	Sound	Sound	No sound	Sound
3,4	Sound	Sound	No sound	No sound	Sound	No sound

Table 3. Test cases for formal pilot test, variations in relief type. For each task there were 4 subtasks (for complex tasks 2) with varying relief type.

Test case	Subtask X.1	Subtask X.2	Subtask X.3	Subtask X.4
1, 3	Positive	Negative	Positive	Negative
2,4	Negative	Positive	Negative	Positive

Fig. 2. Sample recognition task material for outline tasks. Two geometrical figures and two curves, one open and one closed.

Fig. 3. Sample recognition task material for area tasks, Swedish road sign symbols

4 Results

4.1 Results from Qualitative Evaluations

All users seem to enjoy the program, and have found it easy both to draw and feel the lines. For line following negative relief seems to be preferred, although one user expressed a liking for positive relief. In general both types of relief seem to be wanted. Our first test was done using the PHANToM Omni model, while later tests due to practical reasons were done using a PHANToM Premium. This caused problems since the Premium pen is less easy to hold (particularly for blind children who are not as used to holding a pen as sighted children are). Furthermore the tiny switch on the Premium was harder to manipulate than the buttons on the Omni – particularly for children with more problems with their fine motor skills. Both vertical and horizontal work areas have been tested. The horizontal work area puts less strain on the arm, and allows for external physical constraints (e.g. the table) to stop the user from pushing through the haptic surface too much. The vertical work area on the other hand seems to generate more distinct haptic feedback – uses express that shape differences may be easier to feel with this orientation.

There was no clear preference for keyboard buttons over virtual buttons. The advantage of the keyboard is that it can be accessed without moving the PHANToM, but our users have to remember a lot of keyboard commands already, and thus keyboard use may lead to an unwanted increase of the memory workload.

For the tests with both haptics and audio, all three test users were able to use the application as intended, and the different task results for the users seem to match personal differences in fine motor skills and the ability to remember images. Two of the three users could draw a square and a number, two could recognize the Arabic numbers and all three could recognize the simple geometric shapes (circle/triangle/square). During the test session some general observations were also made. It seemed as if some of the users were helped by the difference in sound character to know which program mode they were in. This helped especially one user who previously had had big problems releasing the switch to feel the painting. The sounds, however artificial, did not disturb the users very much, although one musically trained user found them disturbing. That same user also indicated the similarity of the sound with the aiming sound used for target shooting for blind users. Another user expressed great enjoyment with the sound and spent quite much time playing with it.

4.2 Results from Formal Pilot Study

The formal test took approximately one hour to complete. 10 out of 11 test persons were able to complete the test, although a few test tasks had to be cancelled due to time restraints. One out of the 11 users had an elbow injury leading to less controlled fine motor movements, which first caused overheating in the PHANToM motors, and then caused the user to break the test due to pain.

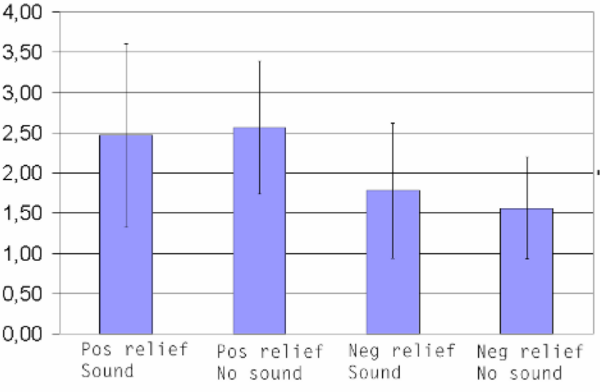


Fig. 4. Mean estimated difficulty of geometry recognition tasks on a scale from 1 (least difficult) to 5 (most difficult)

In general, the geometry recognition tasks were found to be the easiest. The time to complete the examination of the figures was shorter than the time to examine curves and road signs, and there can also be found indications that negative relief is to be preferred over positive relief, both subjectively (figure 5) and by time measure (figure 6). However, there appears to be no significant difference between the presence and the absence of the sound field.

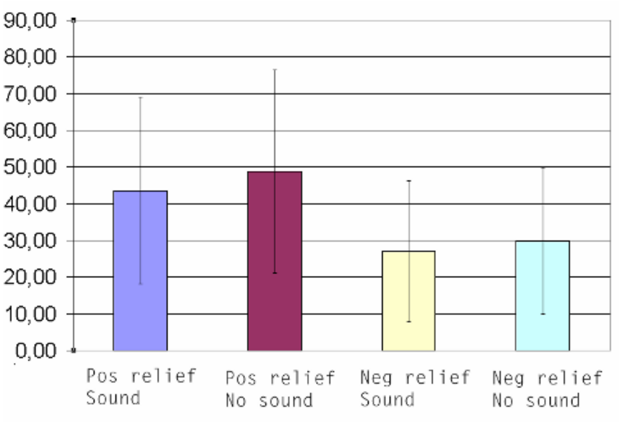


Fig. 5. Mean measured time in seconds spent on geometry recognition tasks

Since the task did not incorporate any reproduction of the geometrical figures by drawing, there is no information on how users perceived the figures in exact detail. However, 3 out of 10 users described the pentagon (se figure 3) as being "a house side" or "a square with a triangle on top". One user mistook the octagon (8 sides) for a heptagon (7 sides), another user for a nonagon (9 sides). Two users also mistook the hexagon for a pentagon.

Further, in the curve recognition tasks, there appears to be no major difference between the different conditions neither concerning subjective measures or time measures (figure 7).

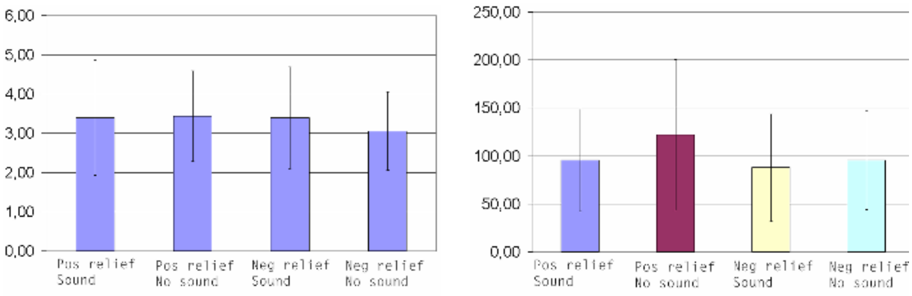


Fig. 6. Left: Mean estimated difficulty of curve recognition tasks on a scale from 1 (least difficult) to 5 (most difficult). Right: Mean measured time in seconds spent on curve recognition tasks.

As can be seen in table 4 a majority (8) of the ten test users drew more than half of the figures reasonably correct.

Table 4. Number of correctly drawn curves by user number

User	Nr of correct curves	Total nr of curves
1	7	7
2	7	8
3	6	8
4	7	7
5	2	8
6	3	8
7	5	8
8	6	7
9	5	7
10	6	8

The most common type of major error is that part of the figure is missing. This occurs particularly when the figure contains separate parts or when there is an intersection inside the figure. But also in the case of a single line, this may occur when the figure contains sharp direction changes such as the curve illustrated in figure 8.

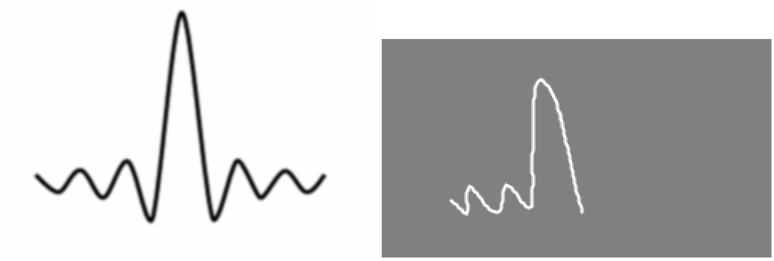


Fig. 7. Curve recognition task: the original curve and one example of the curve drawn by a user

For this type of figure several users missed the part of the figure before or after the big “bump” in the middle. The drawings also illustrate the need for better feedback during drawing, since minor mistakes such with regards to exact/relative positions and shapes were quite common – most users drew the figures from memory (as if drawing in the air) and would easily lose their orientation within the virtual environment.

The road sign recognition test was considered even more difficult; the mean rating was 4 on the 1 to 5 difficulty scale. However, despite the obvious problems the users had to examine and to reproduce most of the signs with pencil and paper, the users still pointed out the correct road sign on average 3 out of 4 times when presented with a chart of 24 different road signs.

It is hard to extract any information about benefits of choosing either one of negative or positive relief. Since it was a whole area that was embossed (positive or negative) the scanning of the area was difficult in either mode. Nevertheless, some observations were made that indicate that negative embossment is easier to scan because it more clearly constrains the user to an area.

5 Discussion

The formal pilot study shows that there is a tendency towards the preference for negative relief when it comes to following lines for the simpler line based drawing (e.g. geometrical figures). Also the time measurements indicate that negative relief is easier to use. This approach is also used by e.g. Yu et al. in [13]. It seems that this effect is less obvious in recognition of more complex line drawings. The study in [14] also shows that individual preferences for relief vary.

The examination time results for the positive relief rely much on the strategy adopted by the user. If a user follows the inside of a closed figure, recognition tends to be found easier and object examination times shorter. This effect is furthermore indicated by the complete failure of one user to follow the open curve in figure 8 in positive relief.

The error made by 3 of the users, who mistook the pentagon with sloped sides for a figure with straight vertical lines is also exemplified by Riedel and Burton in [15]. The sound field present did not give users any help in determining the slope either, since the stereo panning of the sound has too low resolution.

The sound information was shown not to affect the examination times and recognition accuracy. Pakkanen and Raisamo [16] have previously shown that exact recognition of geometrical objects using a combination of vibro-tactile feedback and audio is hard. Some users also expressed annoyance with the sound, whereas some users enjoyed it despite its artificial sound. One user suggested that the sound feedback should convey information about the placement of the center of the paper rather than the height of the PHANToM pen. Another user suggested that the sound information adjusted with a larger pitch range and better stereo effect might give information about the size of objects or relative shape of similar objects (like a sphere and an ellipse for example).

On two occasions, the absence of sound field feedback did have an impact on a single user's result. Due to technical problems, the contact sound with the walls stopped working after a while, which affected the examination times in the test cases without sound field feedback, since the user mistook the edges along the limiting walls for lines. With the sound field feedback present, the limiting wall contact sound was not as crucial.

All user tests showed that the application could be used also to create own line drawings in either positive or negative relief. However, since it (at the time of the test) was not possible for the user to feel the exact line that he/she was drawing, there was no easy way to connect the end point with the start point or start segment of the same line to produce closed curves.

6 Conclusions

- Both positive and negative relief is possible to feel and to work with.
- Negative relief is preferred when working with simple line shapes.
- There are indications that negative relief shortens examination times.
- Drawing lines with a haptic drawing tool is not too easy, but not too difficult either.
- Both vertical and horizontal virtual paper will work in the short run – but what about ergonomics?
- Simple shapes can be recognized when they are kept in a specific context.
- The sound feedback can be used to get information about the program mode.
- Sound feedback did not seem to have a positive effect on task completion times.
- The PHANToM Premium is hard to use especially for blind users who also don't handle an ordinary pencil very easily.

Acknowledgements

This research work is funded by the EU/IST eInclusion project MICOLE (IST-2003-511592 STP).

References

[1] Sjöström, Calle, "Non-Visual Haptic Interaction Design - Guidelines and Applications." Ph D Dept. for Design Sciences, Lund University, Faculty of Engineering, 2002.

[2] Yu, W. and Brewster, S. A. Comparing Two Haptic Interfaces for Multimodal Graph Rendering. 2002. Florida, USA.

[3] Yu, W., Kangas, K., and Brewster, S., "Web-based haptic applications for blind people to create virtual graphs," *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003.HAPTICS 2003.Proceedings.11th Symposium on*, pp. 318-325, 2003.

[4] Bussell, L. *Touch Tiles:* Elementary Geometry Software with a Haptic and Auditory Interface for Visually Impaired Children. 512-515. 2006. 2003.

[5] Iglesias, R., Casado, S., Gutierrez, T., Barbero, J. I., Avizzano, C. A., Marcheschi, S., and Bergamasco, M., "Computer graphics access for blind people through a haptic and audio virtual environment," *Haptic, Audio and Visual Environments and Their Applications, 2004.HAVE 2004.Proceedings.The 3rd IEEE International Workshop on*, pp. 13-18, 2004.

[6] Magnusson, C., Rassmus-Gröhn, K., Sjöström, C., and Danielsson, H. Navigation and Recognition in Complex Haptic Virtual Environments - Reports from an Extensive Study with Blind Users. Wall, S. A., Riedel, B., Crossan, A., and McGee, M. R. 2002. Edinburgh, UK.

[7] Magnusson, C. and Rassmus-Gröhn, K. Audio haptic tools for navigation in non visual environments. 2005. 11-17-0050.

[8] Magnusson, C. and Rassmus-Gröhn, K., "A Virtual Traffic Environment for People with Visual Impairments," *Visual Impairment Research*, vol. 7, no. 1, pp. 1-12, 2005.

[9] Kennedy, J. M., *Drawing and the Blind - Pictures to Touch* New Haven and London: Yale University Press, 1993.

[10] *Art Beyond Sight* Art Education for the Blind, Inc. and AFB Press, 2003.

[11] Kamel, Hesham M., "The Integrated Communication 2 Draw (IC2D)." Ph D Electrical Engineering and Computer Sciences Department, University of California, 2003.

[12] Hansson, Christine, "Haptic Drawing Program." Master Design Sciences, Lund University, Faculty of Engineering, 2003.

[13] Yu, W., Brewster, S., Ramloll, R., and Ridel, B., "Exploring computer-generated line graphs through virtual touch," *Signal Processing and its Applications, Sixth International, Symposium on.2001*, vol. 1 pp. 72-75, 2001.

[14] Sjöström, C., Danielsson, H., Magnusson, C., and Rassmus-Gröhn, K., "Phantom-based haptic line graphics for blind persons," *Visual Impairment Research*, vol. 5, no. 1, pp. 13-32, 2003.

[15] Riedel, B. and Burton, A. M. Perception of Gradient in Haptic Graphs by Sighted and Visually Impaired Subjects . 2002. 2002.

[16] Pakkanen, T. and Raisamo, R., "Perception strategies in modal-redistributed interaction," *Computer Architecture, 2004.Proceedings.31st Annual International Symposium on*, pp. 641-644, 2005.

Creating Accessible Bitmapped Graphs for the Internet

Graham McAllister, Jacopo Staiano, and Wai Yu

Queen's University Belfast, Sonic Arts Research Centre
University Road, Northern Ireland , BT7 1NN, UK
`{g.mcallister, w.yu}@qub.ac.uk`
`http://www.qub.ac.uk/sarc`

Abstract. Bitmapped graphs are the most frequently found form of graph contained on web pages. However, users who are blind or visually impaired currently find it difficult or impossible to access the data contained within such graphs, typically relying only on the ALT text description. This paper details an approach to creating bitmapped graphs for visually impaired users to access on the Internet. The process employs a combination of manual intervention from a web developer, and novel automatic algorithms that are specific for graph-based images. The approach identifies the important regions of the graph and tags them with meta-data. The meta-data and bitmap graph are then exported to a web page for sonification and exploration by the visually impaired user.

1 Introduction

Graphs play an important role on web pages, they can be used to summarize pages of textual information, compare and contrast between different data series or show how data varies over time. However, for people who are blind or visually impaired their degree of usefulness is often limited to the textual description in the image's associated ALT text. The ALT text is used on web pages to provide a textual description of an image and is often used by people who are blind. The level of text contained in the ALT text description depends on the annotation skills of the web developer, and may contain no text at all.

The main problem with the ALT text description is that it is a static textual description, intended to represent in text what the image does pictorially. For graphs, this poses a particular problem as a graph typically represents multiple dimensions of information. Consider the graph in Figure 1. The user may wish to know how many countries are in the survey, which country spends the least or most on health care, or perhaps the user requires the value of a specific country. This information cannot be easily embedded within the ALT text, and would exceed the recommendation by the RNIB, that ALT text should ideally be no longer than one sentence. A theoretical approach to automatically generating textual summaries of graphs has been proposed by McCoy [1].

In order to solve this problem, researchers have tried to make graphs and images accessible [2] by focussing on graphs created using Scalable Vector Graphics

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 92–101, 2006.

(SVG). The advantage of this approach is that it is relatively easy to gain access to the structure of the graph as the information is represented using XML. Thus, any unnecessary information can be removed from an SVG graph, leaving only the most important information [3]. By isolating only the most important aspects of a graph it makes it easier for a blind user to explore using audio and / or haptic feedback [4, 5]. A summary of the use of audio for graphs has recently been published by Stockman [6], and research on a generic sonification approach, which includes graphs, has been conducted by Franklin [7].

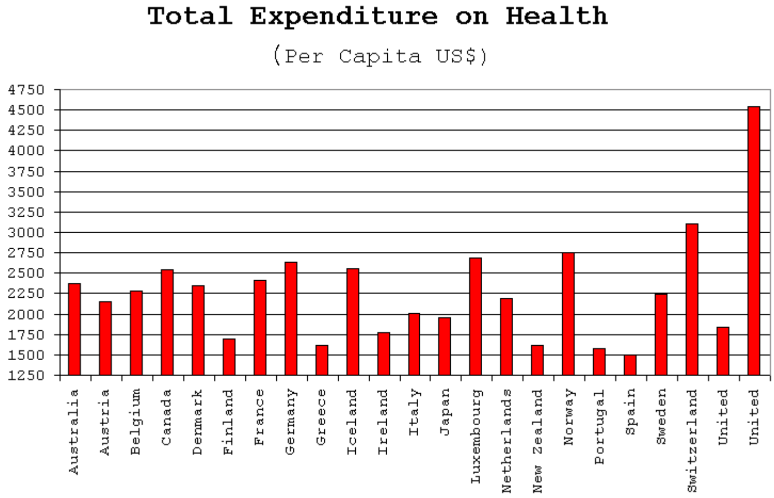


Fig. 1. Typical graph on a web page

However, such approaches are not altogether practical as SVG-based images are not frequently used for representing graphs on the internet due to bitmap-based images being easier to generate.

As a result, the research presented here which is part of ENABLED, an on-going EU project, focuses on an approach to make bitmapped graphs accessible to blind users on the Internet.

2 Bitmapped Graphs

From preliminary investigations, the most common type of graph found on the Internet is bar graphs, therefore the research focuses on this type initially. Before an algorithm can be designed to analyze bar graphs, it is critical to understand the variations that are present on web pages.

Figure 2 shows a typical example of a moderately undecorated bar graph. The bars are in 2D only (no perspective or shadowing), and the bars are represented by a colour that is distinct from the background. This example demonstrates the problem of irregular gaps between each bar.

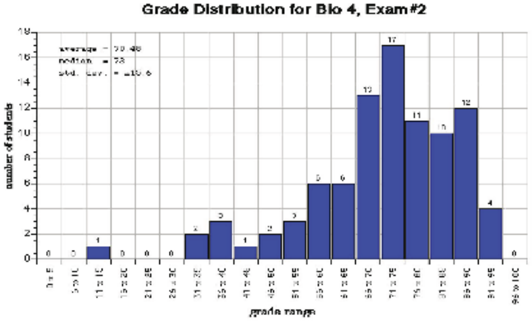


Fig. 2. A simple bar graph

Figure 3 shows an example of a double bar graph, each bar having its own unique colour. This example demonstrates bars that have a 3D shadow effect applied.

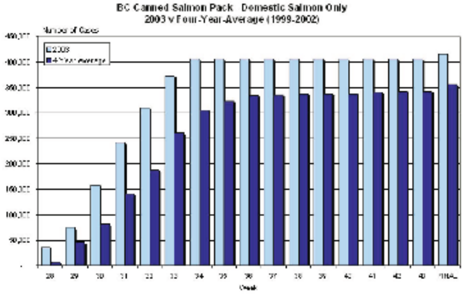


Fig. 3. A double bar graph

Figure 4 shows an example of a bar graph with stacked bars with 3D extrusions.

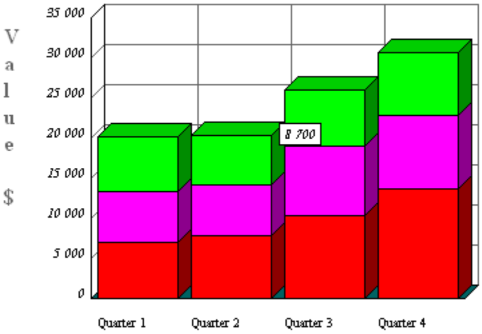


Fig. 4. A stacked bar graph

Figure 5 shows an example of a bar graph which is both stacked and has horizontal bars.

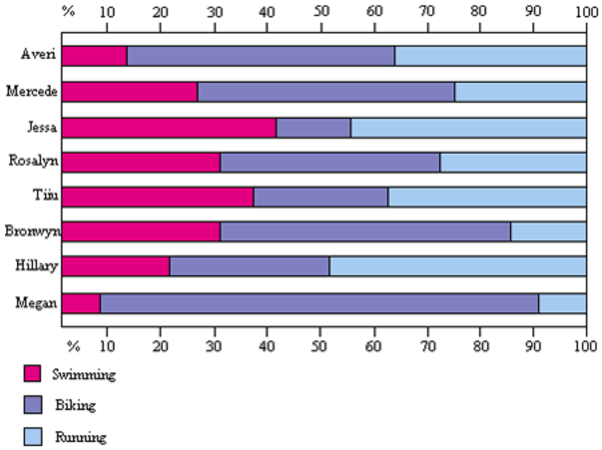


Fig. 5. A horizontal stacked bar graph

These examples highlight that even when focusing specifically on bar graphs, there are many variables involved in their construction, namely:

- directions of bars
- stacked bars
- 3D effect / shadowing
- gaps between bars / irregular formatting
- colour of bars / background image

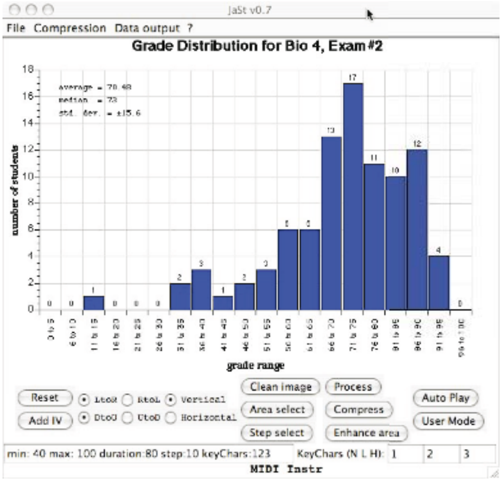
Now that the variables in creating bitmapped graphs have been discussed, the next section will detail the design rationale of the accessible bitmapped graph tool.

3 Creating Accessible Bitmapped Graphs

As web developers can create their own custom graphics to suit both the purpose of the graph and their own personal aesthetic requirements, it is difficult to develop a fully automatic image processing algorithm which can adapt to all image differences. The approach applied in this paper is to use the human visual system (HVS) of the web developer to aid in the segmentation process, followed by an automatic process of calculating the data values and applying the sound mapping.

3.1 Image Markup

Assuming that the graph has already been created, the first stage of the markup process is for the web developer to open the original source image into the



min: 40 max: 100 duration:80 step:10 keyChars:123

KeyChars (N L H): 1 2 3

Fig. 6. Accessible bitmapped graph application GUI

application. Figure 6 shows the application user interface with a graph that is to be marked-up.

Once loaded, the web designer can confirm that any meta-data information is cleared by clicking on the Reset button. The next step requires the identification of the data bars, this is achieved by clicking on the Add IV (Information Vector) button, followed by clicking on any region inside the data bar. If there is more than one data series, the user repeats this step as many times as necessary. This stage records the pixel intensity value of the data bars of interest. If more than one IV is added, a new MIDI instrument option is added, one per IV. The web developer can assign a specific MIDI instrument to a particular data series.

Next, the web designer selects the orientation of the bars, i.e., are they vertically or horizontally aligned, are the axis values ascending left to right, right to left, down to up, or up to down. This allows for graphs to be tagged regardless of which orientation the graph creator used. For example, the graph creator could have made a graph where the data bars go from top of the image downwards, the application allows for the tagging of such graphs. These variables inform the direction of the search path for determining the values of data bars.

The next step requires the web developer to identify the rectangular graph area region. This is performed by clicking on the Area Select button then selecting the top left and bottom right hand coordinates of the spatial region. The system records the pixel coordinates of this region.

The final step in the graph markup process is to identify the step size. The step size of a bar graph is the distance between any two consecutive axis tick marks (typically the width that a data bar occupies). The web designer inputs the step size by firstly selecting the Step Select button, then clicking twice in the graph region. The first click denotes the start of the step, the second click denotes the end of the step. In the example graph shown, this is very easy to determine as there is a grid in the background which can be used as a guide. For graphs without such a grid, the web designer can select the left hand side of the data bar, followed by the right hand side. The y-coordinate is ignored in this case, ensuring that the web developer does not have to make perfectly straight lines.

The web designer has now completed setting the parameters for the graph, and can hand over the generation of the meta-data to the application's algorithms by clicking the Process button.

3.2 Image Analysis

Sonification mappings of the image are calculated by analyzing the image according to the input parameters provided by the web developer. The algorithm starts at the region specified in the graph orientation variable, for example if the web developer clicked the L to R and D to U radio buttons, then the graph values are ascending left to right (x-axis), down to up (y-axis). Thus, the start location for image analysis is the lower left hand corner of the spatial region identified. However, the search will not start at the first pixel, but will start indented to the right, at a value of half the step size. This ensures that the search algorithm will 'look' along the middle of the data bars and not the edge information, where mistakes can be more easily made. The algorithm will determine the number of data bars in the graph by dividing the x-value of the spatial region by the step size, this value is used to process each data bar in turn. Starting with the first data bar, the algorithm examines along the orientation of the data bar (either horizontal or vertical, determined from the input parameters). By counting the number of pixels of same pixel intensity as stored in the Information Vector, a value can be attributed to the data bar as a percentage of the overall height value of the graph. As the shadow region or 3D extrusion is typically a different colour than the data bar, this approach ignores such embellishments and only factors in the original data bar. Once each data bar has been assigned a value, the data is mapped to audio feedback. At present, the audio feedback is mapped in a similar fashion to previous work by Mansur [10], with higher data bars being assigned a higher pitch.

The web developer can now test the automatically generated sonification themselves by selecting either the Auto Play mode (an audio event is generated per pixel), or User Mode, where the mouse can be used to explore the region. These current options for sonification are not ideal however as the per-pixel feedback takes too long to gain an overall impression of the graph. Also, due to errors induced by the graph creator such as gaps between data bars, errors may result during sonification. For these reasons, several extensions to the application were introduced.

3.3 Optimizations

In order to improve the audio feedback, several main types of enhancements are provided; compression, enhancing and time scaling.

Compression

In order to improve the audio feedback so that only one event occurs per data bar (rather than generating an event for every pixel), block compression ensures that repeated notes are omitted within each step size. A further enhancement to this is the option to suppress zero values which occur in runs less than the step size. For example, in the graph shown in Figure 5, the small gaps in between some of the data bars may cause erroneous values to be calculated, however as they are shorter than the step size, the suppress zeros option removes them from the calculations.

Enhancing Regions

If the web designer wishes to highlight a particular region of the graph to the end user, for example, a peak, trough or average value, then the Enhance Area function should be used. Once the Enhance Area button is clicked, the web designer then clicks on the designated area of interest, which prompts a dialog box to appear containing two values to be set. The first of these values is the volume that the data bar value is to played at (0-127), the second value is the duration the audio event should play for (in milliseconds).

Time Scaling

Using the time scaling options, the web designer can set the time it takes to play each audio event. They can also set different times for the lowest and highest value in the graph, giving them a distinct characteristic for the user. The web developer can assign key bindings for three types of feedback which the user will use whilst on the web page. There are Auto Play, Low Detail and High Detail. These features will be discussed in the User Experience section. The defaults for these key bindings are 1, 2 and 3 on the top row of the keyboard.

4 Applet Embedding

The menu option Data Output, allows the web developer to automatically generate the necessary code to export the graph and associated meta-data to an HTML file. This option automatically generates a new folder containing the HMTL page for the graph only (and meta-data), however if the graph is to be embedded on a more complex web page, the web designer can simply paste in the automatically generated code into whatever web page they are designing. Figure 7 shows the associated meta-data for Figure 2.

As the meta-data is extremely small (less than 1KB in this example), this technique works well for users even on dial-up Internet access.

5 User Experience

From the user's perspective, the graph appears just as any other on a web page. However, whenever the user's mouse enters the graph region, they will receive audio feedback on the data bar which the cursor is over. In the case of Figure

```
<APPLET CODE="JaStApplet.class" HEIGHT=420 WIDTH=590>
<PARAM NAME=Image VALUE="stats.gif">
<PARAM NAME=Orient VALUE="true">
<PARAM NAME=Rect VALUE="43;39;540;299;">
<PARAM NAME=Step VALUE="27">
<PARAM NAME=Colors VALUE="1">
<PARAM NAME=Duration VALUE="80">
<PARAM NAME=Velocity VALUE="70">
<PARAM NAME=MaxLength VALUE="20">
<PARAM NAME=KeyChar VALUE="1">
<PARAM NAME=FastChar VALUE="2">
<PARAM NAME=SlowChar VALUE="3">
<PARAM NAME=Cmp VALUE="0">
<PARAM NAME=Instruments VALUE="0;">
<PARAM NAME=Vels
VALUE="20-70-70-70-70-70-70-70-70-70-70-70-70-70-70-70-70-70-70-70-70-">
<PARAM NAME=Durs
VALUE="80-80-200-80-80-80-80-200-80-80-80-80-80-80-200-80-80-80-80-80-">
<PARAM NAME=LowDisp VALUE="40">
<PARAM NAME=HighDisp VALUE="40">
<PARAM NAME=Audio
VALUE="20-40-40-43-40-40-40-46-49-43-46-49-59-59-83-96-76-73-79-53-40-">
</APPLET>
```

Fig. 7. Meta-data embedded into an HTML page

2, the y-position of the user's cursor does not matter, only the x-position as the applet calculates which data bar is associated with the current x-position and renders the audio feedback (effectively the audio event associated with the height of the bar graph). Similarly, if the graph has data bars which are oriented horizontally, only the y-position of the mouse is used to determine the audio feedback. This allows the user to explore the data bars without the mouse cursor having to travel in a straight line. Previous research by Yu et. al. [8] has shown that it is possible (and advantageous) for blind people to use a mouse for navigation and exploration.

In addition to free mouse exploration, the user can use the three keys assigned by the web developer. The first key allows the user to play back sonification automatically (identical to the Auto Play feature the web developer has access to), or they can request a slow overview version of the graph, or a rapid overview version. The rapid overview version of the graph assigns a small millisecond value per audio event, resulting in fast playback. The slow overview version assigns a larger millisecond value per audio event, effectively slowing down the sonification process, thus allowing the user more time to process the information.

Figure 8 shows the graph running as an applet in a web browser.

The user experiences the graph sonification via standard MIDI notes assigned to the data bars. The instrument is rendered according to the choice the web developer made during the markup stage. In the case of a double bar graph, each

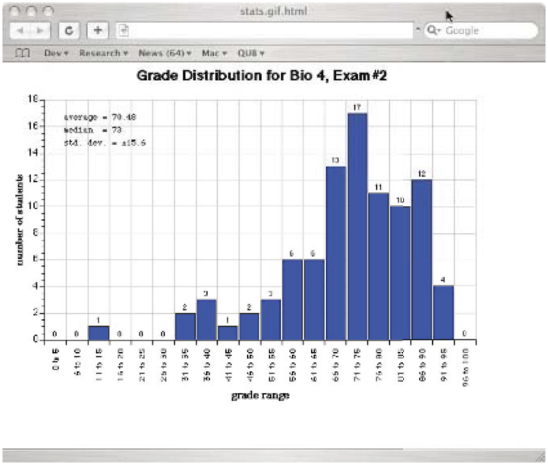


Fig. 8. Accessible bitmapped graph embedded in a web browser

data bar is rendered using its assigned audio timbre. For stacked bars, the data bars are rendered as a chord, however this approach will need further refinement in the future by using approaches suggested by Brown et. al. [9]

6 Conclusions and Future Work

The accessible bitmapped graph creation application presented offers a number of significant advantages and new features over previous research. Firstly, the tool uses the strengths of the web designer (their human visual system) and the efficiency of computer algorithms to allow for very fast markup of bitmapped graphs. A typical time for a web designer to markup an image and generate the necessary applet code is 30 seconds. Given the time taken to design a complete web page, this time only represents a small fraction. Secondly, the tool can cater for a wider variety of graphs than before. Not only can ordinary bar graphs be made accessible, but also bar graphs with multiple data bars, stacked data bars, bars with 3D extrusions and shadows, or any combination of these types. The tool also offers a range of features that enable the web designer to highlight areas of significant interest on a graph. This helps to ensure that the important aspects of the graph are not glossed over, and they can help inform the visually impaired user of the graph creators original intended meaning. The application requires no additional hardware, any computer with a sound card which is MIDI compatible has access to all the features discussed. Finally, the tool allows the user to browse the graph with a mouse, allowing them to explore at their own pace (rather than enforced automatic playback).

Research and development are still ongoing with the tool. In order to obtain accurate data values, the tool should be coupled with a screen reader, allowing visually impaired users to gain an overall impression of the graph and also obtain individual values. The variety of graphs also needs to be extended to include line and pie charts, each presents their own difficulties, even using the web designer's HVS. The next stage will also involve extensive user evaluation for both the web developer and end user, to ensure that the application is performing as intended.

Acknowledgements

This research is supported by ENABLED (Enhanced Network Accessibility for the Blind and Visually Impaired), an Integrated Project funded by the EU Framework 6 Programme, IST (FP6-2003-IST-2-004778).

References

1. McCoy, K. F., Carberry, S. M., Roper T., Green, N.: Towards Generating Textual Summaries of Graphs. Proceedings of HCI (2001), 695-699
2. Herman, I., Darailler, D.: SVG Linearization and Accessibility. Computer Graphics Forum, **21**(4) (2002) 777-786
3. Rotard, M., Otte, K, Erti, T.: Exploring Scalable Vector Graphs for Visually Impaired Users. Springer-Verlag Lecture Notes in Computer Science, **3118** (2004) 725–730
4. Ebina, T., Igi, S., Miyake, T., Takahashi, H.: Graph Access System for the Visually Impaired. Proceedings of APCHI (1998), 254-259
5. Yu, W., Brewster, S. A.: Comparing Two Haptic Interfaces for Multimodal Graph Rendering. Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, (2002) 3-9
6. Stockman, T., Nickerson, L. V., Hind, G.: Auditory Graphs: A Summary of Current Experience and Towards a Research Agenda. Proceedings of ICAD (2005), Limerick, Ireland
7. Franklin, K. M., Roberts, J. C.: A Path Based Model for Sonification. Information Visualization (2004), 865-870
8. Yu, W., McAllister, G., Murphy, E., Kuber, R., Strain, P.: Developing Multimodal Interfaces for Visually Impaired People to Access the Internet. Proceedings of HCI International (2005), Las Vegas, USA
9. Brown, M. L., Brewster, S. A., Ramloll, R., Burton, M., Riedel, B.: Design Guidelines for Audio Presentation of Graphs and Tables. International Conference on Auditory Displays (2003), 284-287
10. Mansur, D. L., Blattner, M. M., Joy, K. I.: Sound-Graphs: a Numerical Data Analysis Method for the Blind. Journal of Medical Systems, 9(**3**) (1985) 163-174

Supporting Cross-Modal Collaboration: Adding a Social Dimension to Accessibility

Fredrik Winberg

Department of Human-Computer Interaction
School of Computer Science and Communication
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
fredrikw@csc.kth.se

Abstract. This paper presents a study of cross-modal collaboration, where blind and sighted persons collaboratively solve two different tasks using a prototype that has one auditory and one graphical interface. The results shows the importance of context and the design of tasks for the accessibility of cross-modal collaborative settings, as well as the importance of supporting the participation in a working division of labour.

1 Introduction

Research on interfaces for blind users has mainly focused on single user interfaces, where one person interacts with a computer. Much focus has been on support for access to the *functionality* of the graphical user interface, providing the same means for *manipulating* the interface objects, support for *exploration* of the screen contents, and a representation that is *coherent* or similar in terms of completeness of the information presented, which objects are presented, how they can be manipulated, and the relationship between the objects [2]. This has successfully been demonstrated in a number of novel applications (see for example [3, 4, 6, 10, 11, 12]). More recent work has focused on applications for specific tasks, such as graphs [8] and mathematics [7], where more focus has been on spatial relations and interactive exploration.

1.1 Collaboration

Even though most researchers and practitioners today agree that collaboration is an integral part of everyday work practice, and that blind users in particular often work with sighted co-workers, studies that actually investigate collaboration are very few. Instead, it is sometimes assumed that *if* a non-visual interface is successful in enabling information manipulation, exploration and provides access to functionality in ways which are coherent, *then* effective collaboration can emerge as a natural benefit. That is, if the same relationships between manipulable interface elements are available to both sighted and blind users, then there are no barriers to effective collaboration.

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 102 – 110, 2006.

Mynatt and Weber [3, p168] make the very strong statement that, if mutually coherent visual and non-visual interfaces are available, “cooperation is thus ensured”. Petrie et al. [4, p429] argue that using a multi-modal display to represent spatial arrangements through tactile and auditory information will “enable successful collaboration” with other users who benefit from standard GUIs. Savidis et al. [6, p118] remark concerning the design of non-visual interfaces which parallel the display relationships of visual ones: “this will inherently lead to equal computer access opportunities for blind users”. Wall and Brewster [ref, p1131] make more cautious claims about collaborative benefits, talking about analogous spatial frame of reference to “potentially support collaboration”.

Previous research has however pointed out that an accessible single user interface does not necessarily imply support for collaboration [9], and it can even be detrimental to accessibility in a collaborative setting to strive for complete functional similarity [9, p340). What seems more important is to support the participation in a *working division of labour* where the blind user’s work is integrated *in an accountable way through a collectively developed, negotiated and evolving knowledge and practice* [1, p117). This means focusing not on the individual’s ability to cope with all possible situations that may arise, but rather the group’s ability to interdependently solve the task at hand, and make collective corrections when problems arise.

Designing alternative presentations for supporting cross-modal collaboration is a question of designing a presentation that *provides means to collaborate, solve problems, and participate in a way that is beneficial and meaningful for the group and for the individual.* This means that the alternative presentation should be sufficient to let the blind user take an active part in the collaboration and be part of the problem solving. The blind participant’s work should also provide an important contribution to the group’s work, and the contribution should be meaningful for the individual as well.

Evaluating interfaces based on the above properties requires not only a conclusion that one interface works, or that one particular solution is better than another, but rather a qualitative understanding of *how collaboration takes place*, and what aspects in the interface that *makes collaboration at all possible*, hence the focus on a qualitative case study methodology rather than a traditional comparative experimental study.

2 Auditory Drag and Drop

The prototype used in the study described in this paper was designed to support drag and drop, which involves movement of objects by positioning a pointer on the object to be moved, picking it up, dragging it to the desired location and dropping it there. In order to do this, the interface must support getting an overview of all objects, locating a specific object, and interacting with that object. The prototype is a simplified version of the interaction common in graphical user interfaces, where a number of objects are located on a two dimensional workspace. The user interacts using a graphics tablet and a pair of headphones. A previous single user study showed how this interface successfully provided a blind user access to the drag and drop features described above [12], but the collaborative aspects of the interface remained to be evaluated.

2.1 Sonification Model

Below is a description of the sonification model used in this study, and the necessary adjustments that were done in order to make it support collaboration.

Zoomed View. The zoomed view gives the user detailed information about a subset of the display, the quadrant in which the pointer is located. All objects placed in the same quadrant as the pointer are audible. The volume of each object depends on the distance to the pointer; the closer an object is, the higher the volume. The objects are presented one by one in a rapid sequence. This enables the user to get a quick overview of the whole workspace by quickly positioning the pointer in each quadrant and listening to the short sequence of objects.

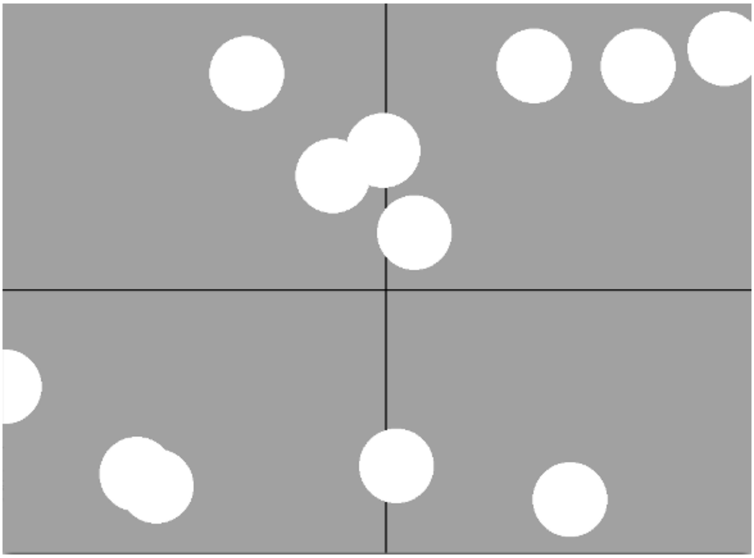


Fig. 1. A graphical presentation of the auditory drag and drop prototype

Objects. All objects have separate sounds. The sound changes depending on where the object is located with respect to the pointer. A guiding tone is added to the sound (a high, low or middle pitched tone representing above, below or at the same vertical level). The distance between the pointer and the object is presented in two different ways. The intensity (volume) of the object sound and the guide tone is mapped to distance, the closer the pointer is the louder these sounds will play. The guide tone has also a repeating pattern that changes with distance. The total time is always constant, but the number of repetitions increases when the distance decreases. This means that the closer the pointer gets to an object, the faster the guide tone will repeat itself.

The horizontal location is represented using stereo panning (left, right or middle representing left, right or the same horizontal level).

Interaction. The user interacts with the objects using a pen stylus on a graphics tablet. This is used in order to have absolute positioning of the pointer, as opposed to the mouse whose relative positioning makes it harder to use sound as the only output device when the complexity of the display is large [cf. 5]. Additionally, using the mouse requires sonification of the position of the cursor, which limits the auditory bandwidth left for sonifying other components [cf. 11].

There are also event driven sounds that give the user feedback on specific actions, in order to emphasize the directness and physical nature of the interaction. These actions include hitting, picking up, dragging, and dropping an object.

Collaboration. For the purpose of using this prototype to study cross-modal collaboration, we added a graphical representation in order to have both an auditory and visual representation of the workspace (see figure 1). Also, additional event driven sounds like the ones described above was added to support awareness of when the user of the visual interface picks up, drags and drops objects.

As described above the screen was divided into quadrants, and these were used to define two private workspaces. The auditory interface just showed the objects placed in the top-left and bottom-right quadrant, and the visual interface only showed the objects located in the top-right and bottom-left quadrant. The presentations were mutually exclusive, which means that there was no shared presentation at all, either you used an audio-only or a visual-only interface.

3 The Study

A collaborative study was performed, where one blind and one sighted person were to solve a number of tasks using this prototype. This study is a follow up study to the Towers of Hanoi collaboration study presented in [9]. In the first study a limited interaction space was available for the users, only three discrete horizontal locations was used for interacting with the auditory interface. In the study presented in this paper the users were able to move objects almost continuously in two dimension, and the question were whether the results from the first study would be valid when interacting with a more complex interaction space.

3.1 Setup

The subjects were sitting next to each other at a table, the blind subject with a pair of headphones and a graphics tablet, and the sighted with a small monitor and a regular computer mouse (see figure 2 below). The sessions were video taped for later analysis.

3.2 Procedure

The subjects were given two different tasks, sorting and handover.

In the *sorting task* an unknown number of objects was randomly scattered on the workspace, and the goal was for the subjects to place an equal number of objects in each quadrant. This task focused on the interplay between individual perception of the objects placed in the private workspace, and the mutual agreement and coordination on what needed to be done to sort the objects. The subjects tried this with 8 and 12 objects.

In the *handover task* three objects were placed in separate quadrants, and the goal was to “move” the empty quadrant clockwise by moving the objects, until the empty quadrant had travelled four laps, while making sure that no quadrant ever contained more than one object. This task focused on the coordinated movement of objects and mutual understanding of an agreed strategy.

Fig. 2. Two subjects solving the handover task. The sighted subject to the left and the blind subject to the right. The small screen at the bottom of the screen was present for monitoring purposes.

For both of these tasks the subjects were told what they had to do, and that solving the tasks required them to collaborate since they could not see or hear the objects located in the other subject’s private workspace. They were told they could move objects and drop them in the other’s workspace, but not picking up objects (by randomly clicking in areas they know objects are placed. When the subjects agreed that they were finished with a task they were supposed to tell the session leader about this.

The session ended with an informal interview with both subjects. In this interview issues like collaboration, what might have caused possible problems, the division of work, and possible changes to the software were discussed. General issues about inclusion were also asked, such as for example whether the blind subject contributed to the problem solving in a meaningful way and if the blind subject only followed the sighted subject's instructions.

3.3 Participants

Two pairs participated in this study; both pairs consisted of one blind and one sighted subject. The first pair consisted of one blind female, 45, with basic computer knowledge, and one sighted male, 37, with intermediate computer experience. The second pair consisted of one sighted female, 29, with intermediate computer knowledge, and one blind male, 30, with basic computer knowledge. The subjects were recruited both at the department (the sighted male and the sighted female) and via contacts from previous studies (the blind female and the blind male).

4 Results

The session was analysed using the recorded video from the sessions (see figure 2 for a screen shot from one of the sessions). During the review of the recorded material specific attention was on how the subjects achieved a *working division of labour* [1] when performing the tasks [cf. 9].

4.1 Sorting Task

The sorting task involved two separate activities: establishing the total number of objects and re-distributing the objects.

To establish the total number of objects the subjects needed to count the number of objects placed in "their" quadrants, add them together, and decide how many objects that needed to be relocated. For the sighted subject this was an easy task and something that was done directly. For the blind subject this involved a bit more attention and took more time. This task was done interactively, which meant that the subjects were talking to each other throughout the whole process, discussing how to solve the problem. Even though it took the blind subject more time to get an overview of the quadrants and the sighted subject often had to wait for the blind subject to find out how many objects were located in the blind subject's quadrants, this was not experienced as a major problem by either participant in the post interview.

One thing that caused some initial confusion was when one or several objects were located at the border between two quadrants (as can be seen in figure 1 above). For the user of the auditory interface an object was either located inside or outside, there was no representation of "on the border", as was the case in the visual interface. This led to an ambiguity when it was time to count the total number of objects since it was not clear whether an object that was in the border was audible or not for the blind user. After a while the subjects realized that the easiest way to deal with this situation was to simply grab these ambiguous objects and place them well within the boundaries of one quadrant. In the post interview this delay in clearing this ambiguity was partly explained by the fact that one of the sighted subjects initially was not sure that she could move any objects at all.

Re-distribution of the objects started when the subjects had agreed on a strategy, which in this case meant that they had decided that a certain number of objects should be re-located to another quadrant. This second part was done very quickly and without any delays. Both subjects had agreed what to do and the movement was done individually.

4.2 Handover Task

The handover task involved the same basic activities as the sorting task: establishing which quadrant was empty and movement of the objects to complete four laps.

Establishing which quadrant was empty was done quickly and no apparent difference between the subjects was noticeable, nor was this brought up in the post interview.

Moving around required some afterthought. In order to “move” the empty quadrant clockwise the objects needed to be moved anticlockwise. After a few seconds of deliberation they all figured it out and started to move. The movement was accompanied by talk all the time, where both subjects were thinking aloud and announcing their moves, as well as acknowledging what the other one was doing. Occasionally one of the subjects would get lost and start to hesitate, but this was quickly alleviated by the other subject and the movement was picked up again. This momentary confusion happened to both subjects.

4.3 Collaborative Issues

The auditory interface made it possible for the blind subject to take part in the problem solving, both by *active inquiries* in the interface as well as *repairing breakdowns.* The auditory interface made it possible for the blind subject to take part in the problem solving, both by *active inquiries* in the interface (exploration of the auditory space) as well as *repairing breakdowns* (realizing when an error has been made and taking the necessary steps to correct this). Interaction with the interface objects, albeit slower than for the sighted user, was unproblematic and did not cause problems for the collective problem solving. *Monitoring* the other person’s actions was indirectly supported by the assembly of resources [9] provided by the manipulation in the interface and the social interaction. In these respects, the interface supports the blind subject participation in a *working division of labour* [cf. 1], where each participant at every instance has a job to do, as well as resources for monitoring the activities of the other participant. Also for both tasks the auditory presentation made it possible to *collaborate, solve problems,* and *participate* in a way that was *beneficial and meaningful for the group and the individual.*

5 Discussion

5.1 Accessibility

When talking about accessibility this is often thought of as something absolute and definite, either we have it or we do not. However, in the post interview one of the blind subjects expressed that even though he lacked the quick overview you get when using a visual interface and sometimes needed to ask the sighted subject about things, he thought that this prototype was fully accessible in the sense that he took part in the problem solving, and that an equal amount of work was done by both subjects. When asked further about this the subject said it did not matter so much as long as it made collaboration at all possible.

5.2 Design of Tasks

The main difference between the two tasks was that in the first case (sorting task) the focus was on getting a correct initial overview and the movement was secondary, whereas in the second case (handover task) taking momentary decisions and making the moves was in focus and the initial overview was secondary. In the first task the perceptual properties of the interface was the focus and in the second the problem solving (albeit on a quite low level). This led to a difference in the relationship between the sighted and the blind subject. Since getting a proper and exact overview takes more time for the blind subject, every instance where this is crucial will leave more time for the sighted subject to work with the actual problem solving, hence creating an imbalance in how the initiative shifted between the participants. This imbalance was present in the first task but not in the second.

This stresses the importance of the design of tasks. How the tasks are designed, and what resources are available, both through manipulating and getting feedback from the interface, and how verbal and non-verbal communication is supported, will not only influence the demands on the multi-modal system, but also to a great extent the division of labour.

5.3 The Importance of Context

The results emphasize the importance of acknowledging the context of use, which encompasses issues such as for example task, location, motivation and the experiences of the people interacting with the system. A fundamental concept like *accessibility* can involve different things depending on what contexts it is applied to, and more importantly an application that is believed to be accessible and to support collaboration in one context might not do so in another.

One crucial question is of course whether a study of an artificial task in a laboratory setting like the one described here really takes into consideration the different aspects of context outlined in the previous paragraph. Studying a real task *in situ* would of course yield the most relevant data *for that specific context*, but there are still good reasons to expect the basic issues of cross-modal collaboration to be present as long as the context includes collaboration of some sort that takes place in different modalities between participants of different abilities [cf. 9, p339].

Future studies of cross-modal collaboration must take these issues into account, and also find a way to capture the specific context that is of interest, as well as the social interaction between the participants.

Acknowledgments

I would like to thank the people who participated in this study, without their contribution this study would never have been possible. Thanks to Kerstin Severinson-Eklundh and Eva-Lotta Sallnäs whose feedback both in planning the study and in writing this paper has been very valuable. Finally, a special thanks to Sten-Olof Hellström for his work on the sonification model used in this study.

This work has been funded by the EU/IST eInclusion project MICOLE (IST-2003-511592 STP).

References

1. Hughes, J., Randall, D., Shapiro, D.: Faltering from ethnography to design. In: Proceedings of CSCW 92. ACM Press (1992)
2. Mynatt, E.: Transforming graphical interface into auditory interfaces for blind users. Human-Computer Interaction, 12 (1997) 7-45
3. Mynatt, E., Weber, G.: Nonvisual presentation of graphical user interfaces. In: Proceedings of CHI ’94. ACM Press (1994)
4. Petrie, H., Morley, S., Weber, G.: Tactile-Based Direct Manipulation in GUIs for Blind Users. In: Conference companion to CHI '95. ACM Press (1995) 428-429
5. Pitt, I. J., Edwards, A. D. N.: Pointing in an Auditory Interface for Blind Users. In: IEEE International Conference on Systems, Man and Cybernetics. IEEE (1995) 280-285
6. Savidis, A., Stephanidis, C., Korte, A., Crispien, K., Fellbaum, K.: A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction. In: Proceedings of Assets ’96. ACM Press (1996)
7. Stöger, B., Miesenberger, K., Batusic, M.: Mathematical Working Environment for the Blind Motivation and Basic Ideas. In: Proceedings of ICCHP 2004. 656-663
8. Wall, S., Brewster, S.: Feeling What You Hear: Tactile Feedback for Navigation of Audio Graphs. In: Proceedings of CHI 2006. ACM Press (2006)
9. Winberg, F., Bowers, J.: Assembling the Senses: Towards the Design of Cooperative Interfaces for Visually Impaired Users. In: Proceedings of CSCW ’04. ACM Press (2004)
10. Winberg, F., Hellström, S.-O.: The quest for auditory direct manipulation: the sonified Towers of Hanoi. In: Sharkey, P., Cesarani, A., Pugnetti, L., Rizzo, A. (eds.): Proceedings of the 3rd International Conference on Disability, Virtual Reality and Associated Technologies. (2000) 75-81
11. Winberg, F., Hellström, S.-O.: Qualitative aspects of auditory direct manipulation. In: Proceedings of the 7th International Conference on Auditory Display (2001)
12. Winberg, F., Hellström, S.-O.: Designing Accessible Auditory Drag and Drop. In: Proceedings of the 2003 conference on Universal usability. ACM Press (2003) 152-153

Non Visual Haptic Audio Tools for Virtual Environments

Charlotte Magnusson[1], Henrik Danielsson[2], and Kirsten Rassmus-Gröhn[1]

[1] Certec, Division of Rehabilitation Engineering Research, Department of Design Sciences,
Lund University, PO Box 118, SE-221 00 Lund, Sweden
`{Charlotte.Magnusson, Kirre}@certec.lth.se`
`http://www.certec.lth.se`
[2] Department of Behavioural Sciences, Linköpings universitet
SE-581 83 Linköping, Sweden
`henda@ibv.liu.se`

Abstract. This paper reports the results of a test involving twelve users of different haptic audio navigational tools for non-visual virtual environments. Analysis of the test results confirms the usefulness of a constant attractive force as well as of haptic fixtures to help users locate objects in a virtual environment. The 3D audio turned out to be less useful due to the design of the environment. However, user comments indicate that this type of sound feedback helps spatial understanding. Contrary to expectations, no significant tool effects were seen on spatial memory.

1 Introduction

With one point haptic interaction in a non-visual setting, it is easy to miss objects or get lost in haptic space [1]. Some navigational tools have been suggested, such as "magnets", "crosses" (allowing the user to feel if he or she is aligned with an object) or a "ball" (to feel things from a distance) [2]. The attractive force in particular has been used and found to be helpful in many circumstances (e.g. [3, 4] and is included as a standard tool in the current OpenHaptics software from SensAble). For graph exploration, Roberts et al. [5] and more recently Pokluda and Sochor [6] presented different versions of guided tours, while Wall and Brewster [7] tested the use of external memory aids, so called "beacons", which the users could place on a surface and which then could be activated to drag the user back to this particular location. Text labels have been used extensively to help users obtain an overview of maps [8] or traffic environments, for example [9].

Other suggested ways to help the user with navigation/learning are automatic guiding constraints, referred to as "fixtures", which have been used for tele-operation, shared control tasks, tracking and training, often in a medical context [10], or to have the user cancel forces generated by the haptic device [11].

If we look at the combination of audio and haptic feedback, we see that for 3D (VR) type environments, there is still not much work being done on designs involving both these modalities. In this paper, we will discuss results from a study performed at

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 111–120, 2006.

Certec, Lund University in the autumn of 2005. It examines a subset of the implementations suggested in our previous pilot studies [12, 13] which investigated several different navigational tools utilising 3D audio together with haptics.

2 Navigational Tools Test

In two previous pilot studies [12, 13] we investigated several different haptic audio navigational tools. We concluded that with the suggested tool designs, the presence of a haptic search tool shortened task completion times. Two different types of attractive forces were tested, and it turned out that the users preferred a constant force (which the user could resist) to a gravity well type force (which forced the hand of the user). It is important to note that the combination of audio and haptic feedback utilised here makes it possible for the users to use tools such as an attractive force or a fixture more effectively. Instead of having distracting forces coming from all objects as in [3], forces were now only activated for one object at a time (on the basis of the sound information). Furthermore, these pilot tests pointed to a possible conflict between speed/tool use and memory/spatial understanding. The tests showed that the usefulness of different tools was not independent of the task – in the first pilot study, "tabbing through objects" was preferred (the task was to find one specific object), while in the second, nobody liked this interaction technique (here the task was to play a spatial memory game).

The "ears in hand" interaction technique introduced proved to be fruitful, but it was not clear how size was perceived (the audio environment was scaled with respect to the haptic environment to achieve a more distinct spatial sound distribution). It should be pointed out that the "ears in hand" technique is intimately tied to the active exploratory actions performed with the hand. Passive input does not produce the same type of spatial experience.

The aim of this study was to further test the tools most popular in the pilot studies, as well as examine possible influences on spatial perception by different navigational tools. Audio feedback (using the ears in hand metaphor) together with haptic feedback in the shape of either a constant attractive force or a linear fixture was investigated. A task of locating three targets and then reproducing their positions was chosen to test effects on spatial memory.

2.1 Implementation

The targets to be located were small boxes. The size of the side of the cubic box was 5 mm to make it virtually impossible to find objects by chance. Two different types of objects were included in the environment. To determine the identity of an object, the user had to press the PHANToM stylus against the side of the cube. This press/click type action generated either a frog or a ping sound. The design was motivated by a desire to force the user to actually locate the targets. The navigational tools were designed in such a way that they always pointed to the object closest in space. Three different navigational tools were tested:

3D audio using the ears in hand metaphor. This implies 3D radial audio sound sources placed at different object locations in the virtual space, while the "ears of the user" are placed at the PHANToM stylus position. Thus the user can explore the resulting 3D soundscape by moving the stylus around. In contrast to the previ-ous tests, this audio feedback did not contain any information about the nature of the object. A short musical loop was used for navigational feedback. The looped sound enabled users to "hear" borders between areas close to different objects since the loop would restart each time the object the sound led to changed.

Linear fixture. This tool was designed essentially the same way as in [13], except that it used a stronger force to attract the user to the line (-400 ρ êρ vs -200 ρ êρ) and that it was toggled on/off by a keyboard press.

Constant radial force. This force was weaker than in [13] (-0.5 êr N vs -1.0 êr N) to allow the user to easily resist the pull. This way it was easy to move about within the environment without being disturbed by the force. When users wanted guid-ance, they just relaxed their grip on the PHANToM stylus, and were moved to-wards the target object by the attractive force. This force was also toggled on/off by pressing a key on the keyboard.

These tools were all tested separately. In addition, the two haptic tools were tested in combination with audio feedback.

During the first part of the test, the task was to find all three objects in the environment and to count the number of "frogs". No visual feedback was available (i.e. the objects were invisible and the PHANToM pointer was not shown graphically on the screen). Fig. 1 shows a visual representation of this environment.

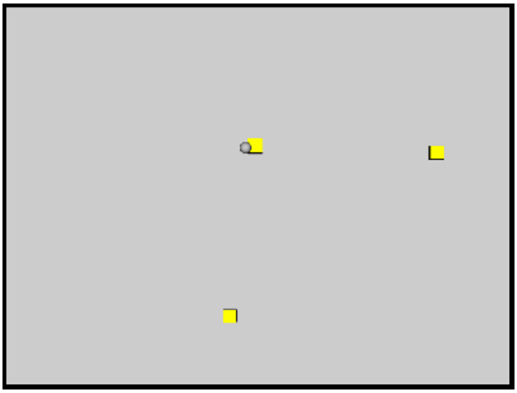


Fig. 1. A visual representation of the test environment. To identify an object such as a "frog" or a "ping" the user had to move the PHANToM pointer to the object and press it. This pressing action activated the sound file identifying the object.

When the user felt confident that all objects had been found, he or she informed the test leader, and the test person then entered the second part of the test. The user was instructed to put the PHANToM pointer at the remembered position of each object and click the button on the PHANToM stylus. This would place an object at this position. The type of object could be changed by pressing a key on the keyboard. This enabled the user to build a model of the test environment encountered in the first part of the test.

Finally, when the user was satisfied with object positions and types, the result was displayed visually on the screen as shown in Fig. 2.

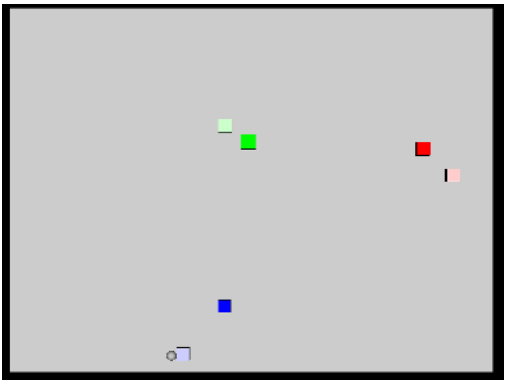


Fig. 2. The visually displayed test result. The intensely coloured boxes are the originals, while the positions assigned by the user are shown in a lighter shade. The computer assigned the object pairing by minimising the total difference in distance between test object positions and as-signed object positions.

2.2 Technical Detail

The PHANToM premium with the ReachIn API were used for the haptics. Zalman ZM-RS6F 5.1 Surround Headphones with Direct3DSound were used for the sound feedback. The following set of sound parameters was used:

Scaling factor (from haptic size to audio size): 100
Rolloff: 1.0
Minimum distance: scaling factor 0.0025 m
Maximum distance: scaling factor 200.0 m

The haptic world was enclosed by a limiting box of 0.2x0.15x0.08 m^3. The sides of the cubic boxes were 0.005 m. A constant force of $-0.5\hat{e}_r$ N was used for the attractive force to make it easy to resist. It should be noted that the strength of such a force needs to be adapted to the haptic device used, since it should be strong enough to move the stylus to the target. The linear fixture was implemented as a spring force attracting the PHANToM tip to a line towards the target. No force was applied along the line – the user had to move actively to reach the target. The force used to attract the PHANToM tip to the line was $-400.0 \rho \hat{e}_\rho$ N where ρ is the perpendicular distance from the line.

2.3 Test Users

Eleven sighted persons and one visually impaired person performed the test. Their age and gender are summarised in the Table 1.

Table 1. The test users

Age	Gender (F/M)	PHANToM experience
37	F	Expert
44	M	Intermediate
58	M	Never used
25	F	Never used
51	M	Few times
53	M	Few times
49	F	Few times
30	F	Never used
43	F	Expert
40	M	Expert
29	M	Few times
35	M	Never used

2.4 Test Setup

This test consisted of two phases. In phase one the user was asked to locate and iden-tify the three objects found in the environment. In phase two the user was asked to build a copy of the environment encountered in phase one. Each test person carried out this task three times for each navigational tool combination (audio only, fixture only, force only, fixture + audio and force + audio). The order of the test tasks was varied to minimise the learning effect. The users received no visual feedback from the environment (from objects or from the PHANToM pointer) except after the test when they were allowed to see how well they had managed to reproduce the initial phase one environment. After the test, each user was asked about preferences and was en-couraged to comment on the experience and the different navigational tools.

The test program logged PHANToM position, object positions, object types as well as toggle actions (fixture and force tool), object presses, elapsed time and the time at which different events occurred.

2.5 Results

The user preferences are summarised in Table 2. The attractive force was a clear winner, while it was unclear whether the 3D sound helped. A summary of the results for the different navigational tools is presented in Table 3.

The results were analysed using five different analyses of variance (ANOVA) with the measures, 1) Time to complete, 2) Distance (total difference between assigned positions and actual positions), 3) Correct number of frogs/pings (even if the sound was assigned to the wrong object), 4) Object clicks per second, and 5) Fully correct object assignments, as dependent variables.

Table 2. Preferred navigational tool

User nr	Preferred tool	Comments
1	Force only	The strength of the force is just right. The audio feedback is confused with the sound tags of the objects (harder to remember them).
2	Force (with or without sound)	Possibly better without sound, since the sound may be disturbing.
3	Fixture with sound	The line is more fun – you get to do something by yourself (the force is automatic).
4	Force only	The sound is not necessary.
5	Fixture with sound	Easiest.
6	Force with sound	Better to use two senses.
7	Force with sound	Easiest.
8	Force with sound	Much faster. The sound helped you feel sure.
9	Force (with or without sound)	The fixture with sound somehow helped with the relative positions, but in a complex environment, I believe the force will be better.
10	Force only	Force more intuitive. But the fixtures were good too, once you learnt to use them. The force works just as well without sound (in contrast to the fixture).
11	Force only	The sound is distracting (harder to remember the object sounds).
12	Force with sound	But this depends on the application – I often neglected the sound.

The independent within-group variable is the navigation tool with five conditions: Fixture, Sound, Force, Fixture with sound, Force with sound. Post hoc analyses were carried out using the Tukey test. The significance level was set to 0.05 throughout the analyses.

The ANOVAs on time to complete ($F(4, 44) = 28.9$, $p < .05$), correct number of frogs/pings ($F(4, 44) = 3.18$, $p < .05$), and object clicks per second ($F(4, 44) = 21.6$, $p < .05$) revealed significant differences. For time to complete, the post hoc test showed sound took significantly more time compared to all other conditions (fixture $Q(5, 44) = 10.8$, $p < .05$, force $Q(5, 44) = 12.0$, $p < .05$, fixture with sound $Q(5, 44) = 11.9$, $p < .05$, and force with sound $Q(5, 44) = 12.9$, $p < .05$). For correct number of frogs/pings, the post hoc test showed no significant difference. Force, however, tended to have more correct numbers of frogs/pings than fixture $Q(5, 44) = 3.93$, $p < .10$ and sound $Q(5, 44) = 3.93$, $p < .10$. Sound had significantly fewer object clicks per second than all other conditions (fixture $Q(5, 44) = 9.85$, $p < .05$, force $Q(5, 44) = 10.1$, $p < .05$, fixture with sound Q(5, 44) = 8.67, $p < .05$, and force with sound Q(5, 44) = 11.7, $p < .05$). No other differences reached significance, including the ANOVAs on distance and fully correct object assignments.

Table 3. Results for the different navigational tools on the five dependent measures

Measure	Navigation tool	Fixture	Sound	Force	Fixture with sound	Force with sound
Time to complete (s)	Mean	133	388	104	107	83
	Standard deviation	59	133	87	90	50
Distance (mm)	Mean	46	44	38	41	40
	Standard deviation	17	15	18	16	13
Correct nr of frogs/pings	Mean	2.69	2.69	2.94	2.78	2.89
	Standard deviation	0.10	0.08	0.04	0.07	0.06
Object clicks per second	Mean	0.178	0.033	0.182	0.161	0.206
	Standard deviation	0.086	0.020	0.092	0.070	0.087
Fully correct object assign-ments	Mean	2.50	2.47	2.78	2.50	2.78
	Standard deviation	0.522	0.521	0.296	0.503	0.296

The user comments are listed below. The items are grouped by content. Each list item is from a different user.

User comments:

The sound is somewhat confusing – one tends to confuse it with the object identification sounds. The navigational sound actually makes you somehow forget the object sounds.

The navigational sound somehow made it harder to remember the object identifica-tion sounds.

The 3D property of the sound is not very good – it is more like stereo + feedback from your moves (the volume/stereo changes as you move).

The 3D sound had good stereo, but up/down and back/front is hard. The sound also makes the object identification sounds harder to remember.

It is harder to remember the object identities (frog or ping) than to remember the positions.

I would like object specific navigational sounds.

The sound loop restarting every time you cross the border to an area close to a new object is a really good clue.

> The borders where the sound loop restarts are really important! They tell you that you are approaching a new object.
> I tried to listen for the restart of the sound loop – this tells you it is a new object.
> Had to visit the objects several times to know where they were. Thought hearing was more demanding – the object positions were somehow easier to remember.
> The sound makes you aware of the room – this could be used for theoretical training of spatial ability (visually impaired user).
> With the sound you really notice the space of the room – I did not notice it that much before. It is really first now that I understand how to move my hand.
> The sound gets better if you close your eyes.

3 Discussion

The results of these tests confirm the usefulness of the constant, weak, radial attractive force (on its own or with 3D audio). For the fixture, which was also considered useful, the sound may have been more important since it provided directional feedback. As for spatial memory, there is really no significant difference between the tools. The force showed a tendency to give better results than the fixture on the recall of the number of frogs/pings. This may be interpreted as if the force interfered less with the recall of the audio object type feedback. However, since no effect was seen for the fully correct object assignments (which also includes object type feedback), it is not clear if this effect is real. No effect was seen for the fully correct assignments or for the distances. If this effect exists, it does not have anything to do with spatial memory. In previous tests, there was a tendency to remember the environment better if you spent a longer time in it, but here, spending a long time in the environment did not appear to help. Another possible effect that would tend to influence the results in the opposite direction is the number of times you can “check back” or rehearse the object positions. Since no really significant effect on spatial recall was seen, it is possible that these two effects cancel each other out with the present test design. Even though the sound with this set up generated significantly longer completion times, the user comments indicate that the 3D sound (ears in hand) may enhance the spatial under-standing – it seems as if this sound feedback may heighten the sense of immersion (we cannot say anything definite on this point though, since immersion was not tested). A factor that may influence the results was that we used the same navigational sound for all objects. We chose this design because we wanted to force the users to actually locate the targets. One of the advantages of sound, however, is that it can be heard from a distance (i.e. it provides a possibility for accessing object information before actually reaching the object). This test also indicates that navigational feedback may interfere with the actual task, although this most likely depends on design as well as modality.

An accidental artefact in the design was the restarting of the audio loop at the borders between spaces close to different objects. This artefact turned out to be quite useful, and implies that when the same navigational sound is used, it should be possible to hear this type of border.

4 Conclusion

A weak constant attractive force has been shown to be useful. This force should be weak enough to allow the user to resist it, while at the same time strong enough to attract the user to the target once the grip is released. With this design, the type of problems associated with attractive forces reported in previous studies [3] did not seem to cause problems in the present set up (we used only three objects, but these would, on occasion, be located quite close to each other). It is, however, important to note that the strength of the force needs to be adjusted to the type of hardware used – the present results were obtained with the PHANToM premium. The use of a fixture to restrict the user to a path leading to the object was also useful, although there was a possible tendency for the users to perform better on the recall of the audio object type feedback with the attractive force. More users preferred the attractive force, but one user liked the fact that with the fixture he had to perform the movement himself. 3D sound feedback with the ears of the listener attached to the PHANToM position (“ears in hand”) is a type of feedback which may help users to gain an understanding of a spatial environment and which may also increase the sense of immersion within the environment (this, however was not tested). This type of audio is also a tool for navigation, but in this kind of environment (with few, small objects) it is less effective, and thus it also limits the ability of the user to “check-back” to rehearse object positions. User comments (as well as results from the earlier pilot tests [12, 13]) show that if possible, it is useful if the sound feedback allows object identification from a distance. The borders between different object spaces provide important information, and it is useful if the sound feedback gives this type of information. In the case of sound identi-fication from a distance, this is provided automatically since the sound will change as the object changes, but in the case of a general navigation sound, this is something that needs to be considered. This test also highlights the possibility that navigational feed-back may interfere with the actual task, which indicates that particular care needs to be taken in the design of navigational tools to avoid such interference.

Furthermore, this test, in contrast to what had been suggested in the pilot tests [12, 13], did not show any significant tool effects on spatial memory. This may be due to the test design, and further investigation is needed to resolve this issue. Another issue that remains to be investigated is the effect of a turning of the ears in the virtual envi-ronment – so far we have always used a fixed facing-forward avatar orientation.

Acknowledgements

The study was carried out with the financial assistance of the European Commission which co-funded the NoE ENACTIVE - “ENACTIVE Interfaces”, FP6-2002-IST-1, No IST-2002-002114, the IP “ENABLED”, FP6 - 2003 - IST - 2, No 004778 and the STREP “MICOLE”, FP6 - 2003 - IST - 2, No 511592. The authors are grateful to the EU for the support given in carrying out these activities.

References

1. Colwell, C., Petrie, H., Hardwick A. and Furner, S. (1998) Haptic Virtual Reality for Blind Computer Users, Proceedings of the ACM Conference on Assistive Technology - ASSETS 98, pp 92-99
2. Sjöström, C. (1999) "The IT Potential of Haptics – touch access for people with disabilities", Licentiate thesis, http://www.certec.lth.se/doc/touchaccess/
3. Wall, S.A., Paynter, K., Shillito, A.M, Wright, M., Scali, S. (2002) The Effect of Haptic Feedback and Stereo Graphics in a 3D Target Acquision Task, In Proc. Eurohaptics 2002, Univeristy of Edinburgh, 8-10th July, 2002, pp. 23-29
4. Langdon, P., Hwang, F., Keates, S., Clarkson, P. J., and Robinson, P. (2002), Investigating haptic assistive interfaces for motion-impaired users: Force-channels and competitive attractive-basins. In Proceedings of Eurohaptics 2002 (Edinburgh UK, 2002), pp. 122-127
5. Roberts, J.C., Franklin, K., Cullinane, J. (2002), Virtual Haptic Exploratory Visualization of Line Graphs and Charts. In Mark T. Bolas, editor, The Engineering Reality of Virtual Reality 2002, volume 4660B of Electronic Imaging Conference, page 10. IS&T/SPIE, January 2002.
6. Pokluda, L., Sochor, J., (2005) Spatial Orientation in Buildings Using Models with Haptic Feedback, Proceedings of WorldHaptics 2005 (Pisa, Italy, 2005), pp 523-524
7. Wall, S., Brewster, S. (2004), Providing External Memory Aids in Haptic Visualisations for Blind Computer Users, In Proceedings of International Conference Series on Disability, Virtual Reality and Associated Technologies (ICDVRAT), New College, Oxford, UK, 20-22 September 2004, pp. 157-164.
8. Wood, J., Magennis, M., Arias, E.F.C., Helen Graupp, H., Gutierrez, T., Bergamasco, M. (2003), The Design and Evaluation of a Computer Game for the Blind in the GRAB Haptic Audio Virtual Environment, EuroHaptics 2003, pp 147-158
9. Magnusson, C., Rassmus – Gröhn, K. (2005), A Virtual Traffic Environment for People with Visual Impairment, Visual Impairment Research, Vol 7, nr 1, pp 1-12
10. Prada, R., Payandeh, S. (2005), A Study on Design and Analysis of Virtual Fixtures for Cutting in Training Environments, Proceedings of WorldHaptics 2005 (Pisa, Italy, 2005), pp 375-380
11. Saga, S., Kawakami, N., Tachi, S. (2005), Haptic Teaching using Opposite Force Presentation, Poster presentation (CD proceedings) WorldHaptics2005
12. Magnusson, C. (2005), Audio haptic navigational tools for non-visual environments, First ENACTIVE Workshop, Pisa, March 21-22, 2005
13. Magnusson, C., Rassmus-Gröhn, K. (2005), Audio haptic tools for navigation in non visual environments, ENACTIVE'05, 2nd International Conference on Enactive Interfaces, Genoa, November 17-18, 2005

A Semiotic Approach to the Design of Non-speech Sounds

Emma Murphy[1], Antti Pirhonen[2], Graham McAllister[1], and Wai Yu[1]

[1] Sonic Arts Research Centre, Queen's University of Belfast,
Belfast BT7 1NN, Northern Ireland
{e.murphy, g.mcallister, w.yu}@qub.ac.uk
[2] Department of Computer Science & Information Systems
P.O.Box 35, 40014 University.of Jyväskylä, Finland
pianta@cc.jyu.fi

Abstract. In the field of auditory display there is currently a lack of theoretical support for the design of non-speech sounds as elements of a user interface. Sound design methods are often based on ad hoc choices or the personal preferences of the designer. A method is proposed in this paper based on a semiotic approach to the design of non-speech sounds. In this approach, the design process is conceptualised by referring to structural semiotics, acknowledging the unique qualities of non-speech sounds, as a mode of conveying information. This method is based on a rich use scenario presented to a design panel. A case study where the design method has been applied is presented and evaluated. Finally recommendations for a practical design method are presented supported by this empirical investigation.

1 Introduction

While there are many general techniques and methods available to user interface (UI) design, there is a lack of theoretical support for the design of individual non-speech sound elements. Entire UI-design projects can often be based on one clear, verified method or methodology. For example, complete methods [1] aim to support all phases of design, from requirement capture to the final evaluation of the completed implementation. However in many application design techniques, the implementation of single UI elements seldom involves a readily available procedure. This stage is usually determined by the personal preferences and experience of the individual designer.

The lack of support for the detailed design of individual UI elements is particularly relevant to non-speech sound. At present most sound design methods for non-speech sounds in auditory interfaces are based on empirical knowledge, often resulting in sounds derived from random selection or the personal preference of the sound designer. A more theoretical design background is required which will create a framework that can be integrated with a practical approach to create the required results. Existing guidelines concerning non-speech sounds mainly relate to the delivery of simple messages like warning sounds [2]. However, in many application areas, the use of non-speech sounds for more complicated communication is necessary. For example, in applications, where the auditory display is the main communication device, the ability to use sounds to convey complex meanings is

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 121–132, 2006.

essential. Furthermore in designing multimodal interaction it is important to be able to use sounds to construct coherent multimodal wholes where visuals, sounds and other modalities can support each other.

Due to the complicated nature and structure of non-speech sounds, it is difficult to provide a purely analytical design approach or framework. There have been attempts to provide such an analytical framework through sound design guidelines [e.g., 3]. However the complex nature of sound makes it difficult to define rules to control all physical and sensory parameters of sound. Even assuming it were possible to assign semantic meanings to every dimension and parameter of non-speech sound, the interaction among different properties of sounds within the interface further complicates the design problem from a purely analytical perspective.

An alternative approach to sound design is proposed in this paper where the creativity of participants in a design panel is exploited. Using design panels as an application of cooperative design is not a new strategy in UI design. For instance, paper prototyping with groups of users at a design stage, is a useful technique for designing visual layouts. Due to the complexity and divergence of sound, visual design techniques cannot be directly transferred to the context of sound design. The design framework selected and presented in this paper is based on a semiotic approach to the design of non-speech sounds. In this approach, the design process is conceptualised by referring to structural semiotics, taking into account the unique qualities of non-speech sounds, as a mode of conveying information. The underlying idea is to understand and conceptualiseUI sounds as elements of a larger whole, rather than designing individual sounds in isolation.

2 Concepts and Theory

2.1 Auditory Icons and Earcons

Non-speech sound is a broad concept in interface design, which is difficult to define clearly. In the context of human-computer interaction, it is usually referred to when dealing with relatively short sound signals with identifiable meaning.

Non-speech audio signs are usually divided into two categories; auditory icons [4] and earcons [5]. Auditory icons are based on real life events while earcons are symbolic in nature. As the focus of this study relates to sound design methodology, the unambiguous categorisation of non-speech sounds is not the primary task. For the purpose of this study

1) all non-speech audio signs are either auditory icons or earcons and
2) the borderline between these two is subjective. A single sign may be perceived as an auditory icon to one person and as an earcon to another, depending on how important the resemblance to a real-world entity is to the meaning of the sound. (see [6] about this phenomena analysed from the point of view of metaphors).

2.2 Audio Signs: A Semiotic Perspective

In this study, a semiotic approach to the design of non-speech sounds is proposed. Semiotics, often defined as “the study of signs”, provides an appropriate conceptual basis to support the design approach in this study.

Semiotics has numerous branches and schools, but the main branch referred to in this study is structuralism, which originates in the linguistic studies of Ferdinand de Saussure [7]. Despite its theoretical foundations in human language, this approach is well suited to sound design as structurally earcons are highly symbolical and can be considered syntactical. Therefore, the structural analysis originally created for the needs of analysing human language, is largely applicable to the analysis of communication through non-speech sounds.

Existing literature concerning non-speech sounds in UIs has a strong emphasis on the recognition and semantics of sounds (for example in the work of Edworthy [8]). Most of the research is focused on the design, analysis and empirical evaluations of individual sounds. However, from a semiotic perspective, context of use is as important as the properties of individual sound. In structural semiotics, semantic analysis is performed across two different dimensions: *syntagmatic* and *paradigmatic* (e.g., [9, p. 195]).

In a *paradigmatic* dimension, the relationships within a class of signifiers are analysed. A paradigmatic choice is to choose one signifier from a class of signifiers. Each member of the class meets the structural requirements, but the choice has an influence on the meaning of the whole. A typical example is word choice when constructing a sentence. *Syntagmatic* analysis refers to the analysis among constituents of a meaningful whole, e.g. relationships among the words of a sentence. In syntagmatic analysis, the semantic value of an individual signifier is dependent on its relation to the whole (syntagma). Therefore the individual signifier and the whole are reciprocally dependent on each other.

In order to conceptualise sound design within the framework of structural semiotics, it is necessary to define an appropriate syntagma. In other words, we need to conceptualise the whole, part of which each sound is. In workstation application with graphical user-interface (GUI), the GUI itself can be seen to constitute the syntagma (the whole). Individual GUI elements, including sounds, would then get their semantic value in relation to the whole GUI. However, when dealing with applications which mainly rely on auditory display, the definition of syntagma is more complicated. This is the case with small portable devices, which are intended to be used in the move, without constantly using the small visual display. Similarly, applications for visually impaired users are another example, in which audio display has a major role and defining the syntagma is a challenge.

The design method presented here is based on the assumption that the interpretation of an individual sound is highly dependent on the context of use. Therefore, the basis for designing a sound should involve understanding and recognising the relevant environmental (e.g., physical, social and psychological) context. By acknowledging the importance of the context of the individual sounds a designer can create effective communication in the interface. In this study a rich use scenario is proposed as a syntagma for sound design.

2.3 Rich Use Scenarios – A Way to Understand the User

In designing a UI an individual designer will usually have a certain vision about the actual usage of the application. To enable understanding and discussion of this usage, it is necessary to make the designer's internal vision explicit. This is often achieved through the creation of a use scenario.

Typically a formal description of the use of an application can be described as a use scenario. The most common way of formulating a use scenario [10] is heavily based on demographic information. This strategy involves collecting information about the intended user group in order to describe a few generic characters [11] and create a vision on how these characters would use the application. The aim is to encompass as wide a user population as possible with the character description and to describe typical usages of the application. This kind of use scenarios could be effective to concretise a wide variety of usages. The intention in this study is to provide a use scenario to create an overall syntagma for individual sounds, rather than to detail an application. As sounds are strong, multidimensional entities, the syntagma needs to be articulated at the same level. The syntagma, i.e., the use scenario, has to be a deep and vivid framework which provides an appropriate framework for working out individual sounds.

In the proposed sound design method, the purpose of the use scenario is not to cover all possible usages. The story, or rich use scenario, describes a person in a unique situation, although the use of an application is obviously in the focus of the story. In appropriate points of the story, there are places for sound effects. These are the sounds to be designed.

The actual design is proposed to take place in design panels. The participants do not need to be experts in sound design, nor familiar with the application or its usage. The rich use scenario, not the application, is the basis for design. The rich use scenario can be introduced to the design panel as a radio play with sound effects. The task of the panel is to design these effects. The purpose of the rich use scenario is to support creative group work and to encourage panellists to discuss imaginative and divergent sound ideas.

3 The Design Method – A Practical Application

In order to evaluate the design theory described above, three design panel sessions were arranged, similar to that proposed in the previous version of the design method [12]. Participants were different each time, and the nature of the session varied according to the phase of design. Each session involved four or five panellists and the duration of each session was approximately one hour. All panel sessions were videoed for analysis purposes.

The initial aim of the sequence of sessions was to create an iterative design pattern. The first panel was intended as a brainstorming session – no sounds were implemented and panellists were encouraged to describe effective sounds for the use scenario. In the second session, sound samples were presented to participants based on the ideas presented in the first session. These draft sounds were designed with the intention that they would be elaborated on during the second session. Finally these further developed sounds were discussed and evaluated by a third panel. Thus each user panel had different role in the development of sounds. The first panel began with event descriptions, giving initial sonic ideas and principles, while the second session involved working with the first versions of the sounds. The final session involved making assessments and choices of modified sounds.

The use scenario implemented in the user panels was the same in all sessions, although it was slightly shortened in the third session on the basis of feedback from one panellist. The principle in the preparation of the story was to find a balance between inspiration and the details of application use. The intention in the creation of this use scenario was to avoid dull descriptions of a potential sequence of user actions, and include inspiring details about very everyday things that might have an impact in the characters mood and actions. Also, it was considered relevant and important to evoke a mental image of a real, living and credible person, to provide an object for panellists to identify with. The only explicit context related condition was that the story included gaps in the use scenario for sound events to be designed.

3.1 Design Case: Auditory Interface to Convey Spatial Information to Visually Impaired Internet Users

An auditory interface specifically designed for a Multi-modal Browser plug-in [13] to convey spatial information on a web page to visually impaired Internet users was used as a case study to test the design method. It is intended that this system will convey spatial information on a web page in terms of the location of images, links and other web objects such as web forms on a web page through both audio and haptic feedback. The spatial information conveyed using the multi-modal browser should enable blind and sighted users to work together navigating and describing the same space on a web page using spatial directions and descriptions which is currently not possible with current screen reading technology. The interface was considered particularly relevant to this design method as it involved designing non-speech sounds for a specific user group; visually impaired Internet users. An obvious part of design for such an interface is in understanding the user's needs and also the way that a visually impaired user works which can be difficult to understand from a sighted perspective. The use scenario for this design study was written from the perspective of a visually impaired user.

Use Scenario Description

The use scenario applied in this study describes a visually impaired character buying a music file online using the multimodal plug-in. The character is introduced as a young visually impaired student and the scenario describes the character's mood, the technology he is using and the sounds in the environment around him.

> It was one of those mornings, which Kenny would have preferred to skip…..He did not have a screen reader or any other special tools designed for the blind, but had coped reasonably well to date with the help of little plug-ins his former girlfriend had installed on his computer. These plug-ins provided haptic and audio cues to help find and locate graphical objects in a user-interface…. as he scrolled through the playlist to cover the dull sounds of Monday morning:
>
> Doors, toilets, showers and all kinds of household appliances produced in a block of flats create an enormous symphony when people wake up and leave for to work.

The scenario describes the user's movements in the process of finding and buying a music file using the multi-modal plug-in. Task descriptions are punctuated with spaces for possible sounds.

> Kenny typed in the address of the site and soon heard the sound [sound 1] that indicates that the page had successfully opened. At the start page of the online shop, called Cheaphits, Kenny moved the mouse across the page, from left to right, then down. He was already familiar with the tactile and audio feedback and soon got an overall impression of the page. He was especially happy with the sound that guided him towards the links [sound 2] – it attracted the mind like a magnet and made the hand move the mouse towards the link area. A similar kind of magnetic effect led the mouse towards images [sound 3]….
>
> In this page, all ads appeared to be images, so it was quite easy to distinguish them from the useful information because of the clear sound [sound 4] indicating that the mouse was on the image. Another sound [sound 5] told Kenny that he had reached the link area.

User Panel 1 – (Belfast)

Setup. Four people took part in the panel, two visually impaired females and two fully sighted males. All four participants were involved in audio or music related research. Two researchers acted as moderators to read the use scenario and chair the discussion.

One researcher read the use scenario to the group all the way through and engaged the group in discussion with a question about the user's character. The discussion was then focused on possible sound design solutions based on the participants' understanding of the tasks and requirements of the visually impaired character in the user scenario. One of the researchers led the discussion to deal with each section of the scenario that required a sound idea without leading the group towards any sound design solutions.

Sound suggestions. Participants described their ideas for sounds by referring to specific examples of distinctive timbres or sounds from television or software programmes, which are listed in Table 1.

Table 1. Sound suggestions of Panel 1

Task Description	Sound Description Suggestions
Opening a web page	- short sound success sound - bell "you won" - tv programme "ta da" sound
Sound to indicate page loading	-a background sound to know whether a page is downloading - 80 percent etc. to hear what's happening - could be mapped to pitch or a sound filling
"Magnetic" sound to draw users towards hyperlinks	-Sound of a spinning lid falling - the closer you are to the centre the faster and more intense it sounds -Metallic sound
"Magnetic" sound to draw users towards images	Participants did not consider that this location cue was as significant considering that the use scenario character was visually impaired images should have a sound
Sound for cursor over link	-Either an unobtrusive swelling sound or speech audio with link text, metallic sound
Web Form	Keyboard typing - Sound of a pen on paper

The ideas presented in the first panel session had two dominating features. Firstly, as illustrated in Table 1, many of the panellists' ideas clearly fall into the category of auditory icons. Environmental sounds seemed most intuitive when ordinary computer users were asked to spontaneously suggest computer sounds. However, the panel additionally agreed that more abstract symbolic sounds (rather than figurative ideas) could also be effective, but would require learning. The second dominating feature in the discussion was the panel's preference for unobtrusive sounds. The participants spoke about a desirable "swelling" quality and were concerned about the prolonged use of harsh timbres and sharp attacks.

One design idea, in the first panel was quite different from the typical auditory icon or environmental sound metaphor previously discussed. When asked for a sound to combine with a haptic spring effect towards a hot-spot of a web-page, one panellist suggested the sound of a falling lid. She imitated the accelerating rhythm of a round object, spinning on a flat surface. Her idea was that the rhythm would accelerate as the cursor moves towards a hot-spot. So it was not an auditory icon since the origin of the sound (falling lid) had nothing to do with the intended use context (decreasing proximity). Yet the idea could be defined as an environmental sound because of the original association with a real-world object.

User Panel 2 – (Glasgow)

This panel consisted of five sighted researchers in the University of Glasgow, two female and three male participants with one moderator present. This session used the same use scenario and design case as the first user panel. However, in addition to the verbal part of the use scenario, participants were presented with sounds, which were implemented on the basis of the ideas from the first panel. These sounds were purposely unpolished in quality. The reason for this was to give an impression of rapidly constructed drafts, intended to provoke comments and the expression of alternative design ideas. The reactions of the panel are presented in Table 2.

User Panel 3 – (Belfast)

Four people took part in the third and final panel, two visually impaired participants, (male and female) and two fully sighted participants (male and female). This session used the same use scenario and design case as the first two user panels. Similar to the second user panel, participants were presented with implemented sounds designed using the comments and recommendations from the previous sessions.

In this session the use scenario was read to the panel by one researcher, and during this first reading, the researcher played sounds to punctuate the story with sound effects. These were the same sounds had been presented to the second user panel. On completing the use scenario the researcher went back over the tasks with sound effects and gave the panellists a choice of sounds to choose from. Participants were firstly presented with an original sonification of the sound idea proposed by panel one and presented to panel two. They were also presented with a second sound idea, which was a version of the first sound that had been modified according to the recommendations of the second panel.

Participants were asked to comment on the sounds, and to select the most effective sound giving reasons for their choice. Participants were also encouraged to suggest further modifications to the chosen sounds if they did not consider either appropriate. Reactions of the third user panel are presented in Table 2.

Table 2. Reactions of Panels 2&3

Sound Event	**Panel 2 Reaction**	**Panel 3 Reaction**
Opening a web page	A very short pitched "pop" sound was created to signify opening a web page as suggested in panel 2. Panel 2 conceded that the shortness of the sound was important to signify the event but that the sound was too high-pitch to be heard regularly in this context.	In accordance with the second user session, participants rejected the short "pop" sound designed to signify a page loading based on the high frequency. In the final panel, high pitch was associated with a negative sound event.
Sound to indicate page loading	An auditory icon of a glass filling with water was implemented based on the first panel's suggestion for a page loading. A second rhythmic sound was also created and presented to the panel to signify the event of a web page loading. The sound began with a slow rhythmic pace and a soft dynamic and increased towards the end of the page analogous to the nearer the page had loaded.	The rhythmic sound was previously presented to the second panel to signify the event of a web page loading was again played to the third group. A new loading sound was also created which was a mixture between the rhythmic sound and the filling glass sound. Participants preferred the first rhythmic sound, to convey the amount that the page had downloaded.
"Magnetic" sound to draw users towards images	This sound event had and task description had been considered redundant by visually-impaired participants in panel one as although they could use images as navigation cues for spatial cognition, they did not want to be directed towards them in the same way that they wanted to be directed to links.	-
"Magnetic" sound to draw users towards hyperlinks	The implementation of the magnetic "swirling" sound to depict the cursor's movement towards a link was created by synthesizing a sound that began to swirl in frequency and intensity slowly and became faster toward the end of the sound. Participants in panel two responded positively to both the idea and timbre of this sound event.	Participants in panel three responded in a similar way to panel two for this sound. They were positive about timbre of this sound the intended relationship to the task event.
Sound for cursor over link	Participants were presented with a short "swelling" sound with no attack or decay to represent a link. Another sonification of this even was a metallic hitting or "clinking" sound. Panellists agreed that these sounds were effective to convey a web link. Although one user felt that the metallic sound was too harsh and suggested a more subtle clicking sound instead.	Panellists were presented with both subtle clicking sound suggested by Panel 2 and the metallic "clinking" sound. The third panel gave preference to the metallic sound, as they preferred the timbre of this sound.
Web Form Sound	Generally sounds with environmental origin such as the camera sound for an image and the typing sound for a web form were received positively. Although when presented with the sound of a pen writing as an alternative sonification of a web form, panellists reacted negatively as the texture reminded them of an unpleasant scratching noise.	Panellists were presented with both subtle clicking sound suggested by Panel 2 and the metallic "clinking" sound. The third panel gave preference to the metallic sound, as they preferred the timbre of this sound.

3.2 Recommendations for Future Practical Design Panels

User Panel 1. Although at this stage of the process, panellists are only required to brainstorm ideas, it would be useful during this session to include a tool for testing sound ideas at this stage to help users to create specific sound ideas. It can be difficult for panellists to describe sound ideas in words especially if they are not familiar with relevant vocabulary to describe audio parameters. It could be useful to create synthesis instruments (using MAX/MSP or Pure Data) with scalable controls for parameters such as frequency, timbre and intensity so that panellists or even the chairing researcher could create sounds during the session. The introduction of real sounds could focus the panellist's ideas and sounds descriptions.

User Panel 2. As previously discussed by Pirhonen et al. [12], the initial idea of presenting intentional draft-quality sounds in order to encourage the panellists to suggest changes, was not successful. While the approach has proven appropriate in the design of visual lay-outs [14], in this study, such an analogy between visual and audio design is not strong enough to justify taking the same approach in sound design. When draft-quality sounds were presented to the panel, they were immediately criticised by participants. In the discussions it became clear that the main trigger for the negative reaction was not caused by the basic idea of each sound (formulated in the first panel). The negative reaction transpired to be an emotional one - the sounds were simply unpleasant to listen to. This negative reaction prevented the panellists from elaborating on the sounds constructively. Instead, they presented new, alternative ideas as a basis for subsequent sound versions, rather than building from those suggested by the first design panel. Therefore, in order to enable constructive, iterative development of each sound idea, implementations of first versions of sounds should be high quality polished examples.

User Panel 3. The experience of implementing the third user panel revealed that it was quite difficult for the researcher to co-ordinate playing sound choices while simultaneously attempting to chair the sound design discussion. A solution to this could be to create a dramatised recording of the use scenario with sounds included as effects. In this idea the use scenario could resemble a radio-play. Various versions of the recorded scenario could include different sound solutions that the panellists could choose from. Furthermore by recording the use scenario, a more dramatic story could be realised to encourage the users to identify with the user character in the scenario.

3.2 Proposed Design Method

1. Prepare a task description about the sound functions of the application
2. Prepare a user description, based on a vision of a plausible user. The fictitious character should not be designed to hold generic qualities, as the aim is not to cover as many users as possible but to create a lively, inspiring character.
3. On the basis of stages 1 and 2, write a short story in which the interaction among the character and the application is in important role. The perspective is the one of the character – remember that there are many other things than just the application in his/her life and mind. In the story, leave blanks or pauses for the audio effects (the sounds to be designed).

4. Organise a design panel session with 4-5 panellists. Start the session by reading the use scenario, keeping a brief pause in the place of each sound. Having read the story, discuss the story at a general level. Then return to the story, and read again the sentence that includes a blank space for a sound effect. Ask the panellists to try to describe, what kind of sound would be appropriate. Do the same with each sound to be designed. Record the session.
5. Implement the panellist's ideas of the appropriate sounds.
6. Organise a second session with different people. This time, use the draft sounds (implemented in stage 5) when reading the story. In other respects, follow the steps of the first session.
7. Analyse the reactions and new ideas of the second session. Modify original sounds and create new ones as suggested by the second panel.
8. Organise a third session with a different set of participants. Using sounds from 7) record the use scenario and present it to the third panel as a radio play and ask participants to choose sounds at points in the story.

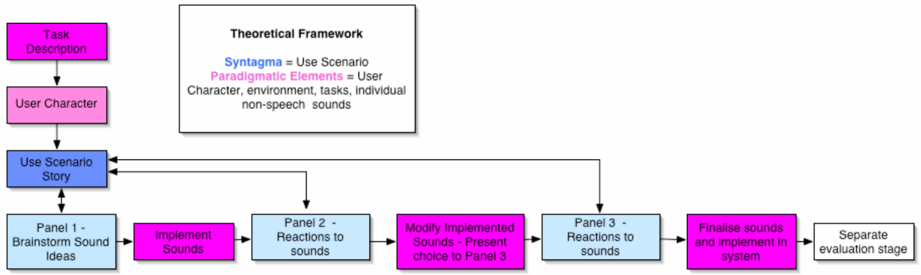


Fig. 1. The sequence of stages of the proposed method

General remarks: The central challenge in the use of the proposed method is to generate a suitable creative motivation among panellists. Otherwise the default expectation of the panellists seems to be that they are supposed to design sounds for an application. However, as stressed above, the task of the panellists, in this approach is not to design sounds for an application but rather to design (or choose, in the 3rd session) sound effects for a radio-play format use-scenario.

In order to generate the abstract creative motivation for the design method the following questions should be considered:

- The invitation to the panel. How should the task be described without stressing the application?

- The panellists: What is the background of the panellists? Are they computer scientists experienced in application development who might find it difficult design effects for a radio-play?

- The venue: Is the panel taking place in the department of computing or other similar place which evoke expectations about a technical task? Do the panellists know you and your design/research interests?

Furthermore, the content of the story itself is fundamental to the design approach. Not only is it important that the orientation of the panellists is story-centred, but the story should refer to all user actions which contain the sounds to be designed. However, it is important to make these references a fluent part of the story, so that the panellists do not become distracted with the technicalities of the application but keep the character and the story central.

4 Conclusions

In this design method the panel described above should not be considered as a design team, but as a source for brainstorming for the designer in the first session. Later in the second and third sessions the role of the panel is to confirm sound ideas of the previous panellists and also the implementations created by the designer. The final implementation of sound sketches rest with the sound designer. For example throughout the three design panels users tended to refer to sounds from real world experiences. Describing auditory icons are a simple way to create for metaphors. However it is up to designer to transform these ideas into more complex sound ideas in the interface.

The main aim of the method is to trigger creativity and achieve a certain amount of group confirmation and finally to provide the designer with a good outlet to get ideas and test sounds early in the development of an auditory interface.

This method provides concrete criteria for the quality of individual sounds: Each sound is assessed according to its suitability to the use scenario. This obviously raises question about the suitability of the same sound in different kind of context other than the specific use scenario described to the panel. However, the aim of this design method is not to optimise the versatility of the resulting earcons. The method is constructed to support creative earcon design, to bridge the gap between an analytical approach and a physical sound.

Acknowledgments

Initially, our approach may have appeared unorthodox in the eyes of many of our colleagues and the participants in the design panels. We appreciate their patience and support.

This work is funded by National Technology Agency of Finland (www.tekes.fi) and Eduserv (www.eduserv.org.uk).

References

1. Wixon, D. & Wilson, C. 1997. The usability engineering framework for product design and evaluation. In: M. G. Helander, T. K. Landauer & P. V. Prabhu (eds.): Handbook of human-computer interaction. 2nd edn. Elsevier, Amsterdam (1997) 653-688
2. Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft. Paper No. 82017. Civil Aviation Authority, London (1982)
3. Brewster, S., Wright, P & Edwards, A. Experimentally derived guidelines for the creation of earcons. In Adjunct Proceedings of HCI'95 (Huddersfield, UK), (1995) 155-159

4. Gaver, W. W. Auditory icons: Using sound in computer interfaces. Human-Computer Interaction 2, 2 (1986) 167-177
5. Blattner, M., Sumikawa, D. & Greenberg, R. Earcons and icons: Their structure and common design principles. Human-Computer Interaction 4, 1 (1989) 11-44
6. 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. Springer Verlag, London (2004) 105-123
7. de Saussure, F. Course in general linguistics. R. Harris, (trans.), Duckworth, London (1983 / 1990)
8. Edworthy, J., Hellier, E. & Hards, R. The semantic associations of acoustic parameter commonly used in the design of auditory information and warning signals. Ergonomics 38, 11 (1995) 2341 – 2361
9. Nöth, W. Handbook of semiotics. Indiana University Press, Bloomington, IN (1990)
10. Carroll, J. M. Making use: Scenario based design of human-computer interactions. MIT Press, Cambridge, MA (2000)
11. Cooper, A. The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity. Sams, Indianapolis, IN (2004)
12. Pirhonen, A, Murphy, E, McAllister, G. & Yu, W. Non-speech sounds as elements of a use scenario: a semiotic perspective. In: Proceedings of ICAD 2006 (in press)
13. Yu, W., Kuber, R., Murphy, E., Strain, P. & McAllister, G. A novel multimodal interface for improving visually impaired people's web accessibility. Virtual Reality 9, 2 (Special issue on haptic interfaces and application) (2005) 133-148
14. Beyer, H. & Holzblatt, K. Contextual design: Defining customer-centered systems. Morgan Kaufmann, San Francisco, CA (1998)

Listen to This – Using Ethnography to Inform the Design of Auditory Interfaces

Graeme W. Coleman, Catriona Macaulay, and Alan F. Newell

School of Computing, University of Dundee, Dundee, DD1 4HN, Scotland
{gcoleman, catriona, afn}@computing.dundee.ac.uk

Abstract. Within the wider Human-Computer Interaction community, many researchers have turned to ethnography to inform systems design. However, such approaches have yet to be fully utilized within auditory interface research, a field hitherto driven by technology-inspired design work and the addressing of specific cognitive issues. It is proposed that the time has come to investigate the role ethnographic methods have to play within auditory interface design. We begin by discussing "traditional" ethnographic methods by presenting our experiences conducting a field study with a major UK-based computer games developer, highlighting issues pertinent to the design of auditory interfaces, before suggesting ways in which such techniques could be expanded to consider the role sound plays in people's lived experiences and thus merit further research.

1 Introduction

Significant steps have been made to identify many of the technical issues associated with auditory interface design. Concurrently, the wider interaction design community has seen increasing interest in developing a descriptive understanding of human behaviour and thus designing appropriate artifacts which support the activities described [1]. Additionally, there is growing recognition of the migration of the computer from the desktop into everyday objects and into our everyday living environments [2], the result being an acknowledgment that new technologies should no longer be seen purely as functional items for accomplishing certain tasks [3], but also to appreciate their *presence* within everyday life [4]. In other words, as well as considering how a new technology shall be *used*, it is necessary to consider whether this technology shall be *accepted*.

Ethnography is one of the methods that Human-Computer Interaction (HCI), Information Systems (IS) and Computer-Supported Cooperative Work (CSCW) researchers have turned to for insight into context and situated practice [5, 6]. A work practice study, for example, can convey the importance of the *sociality* of workplaces, identifying the many complex actions and interactions that take place [7]. Yet, with regards to auditory interface design, such approaches have received scant attention. In this paper, we highlight the potential roles these approaches offer within this domain by firstly presenting a summary of a traditional ethnographic study with a major UK-based computer games developer, highlighting the discoveries that we believe are pertinent for the design of sound for other forms of interaction design. Following this,

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 133–144, 2006.

drawing upon literature on soundscapes and sound classification tools, we suggest how traditional ethnographic tools can be expanded upon to understand the day-to-day role that sound plays in people's lived experiences. As such, it is an explicit attempt to open the discussion as to the problems and benefits using such techniques can bring to auditory interface design.

It is important to note that we do *not* see such techniques as a replacement for the traditional cognitive approach to HCI and related research. Indeed, it would be foolish to dismiss all that the cognitive approach has to offer [8]. At the same time, however, it would be foolish not to understand the context of use of computer systems within a social and cultural context [9]. We therefore see the techniques described here as a means of *supporting*, as opposed to *replacing*, the traditional cognitive approach.

2 The Case for Ethnographic Research Methods

The past two decades has seen discussion within the wider research community as to whether traditional experimental psychology can provide a broad enough frame to understand human activity in order to develop appropriate computer based systems [10, 11]. As a result, there has been what Hughes *et al* [12] describe as a "turn to the social": a requirement for "richer stories" [13] about the role technology plays in everyday life to act as a preliminary for design.

Contemporary ethnographic studies attempt to understand why humans within particular contexts act in a particular manner under particular circumstances [14], and to understand how participants "collide and mix" in changing circumstances in order to find solutions to problems [15]. The basis of ethnographic research involves *fieldwork*, i.e. becoming intimately familiar with the participants and their social activities. Traditionally, ethnographers would carry out research within the natural setting of their participants (the "field") using methods such as observation, interviews, "desk research" (research carried out using documentation and records kept by participants) and surveys. *Participant observation*, whereby researchers directly take part in the culture and lives of the participants while keeping a professional distance to allow adequate observation and data collection [14], is a useful method to adopt. The boundaries between "participating" and "observing" are somewhat narrow, although Wolcott suggests that the key to successful participant observation is "to participate more and to play the role of the aloof observer less" [16]. From this, unexpected discoveries beyond initial preconceptions and frameworks can be found [17]. Hence, the ability of ethnography to make this "real world" visible, and to describe social settings as perceived by its inhabitants (i.e. potential users), underpins its appeal within wider HCI research [18]. Additionally, ethnographic methods can be employed to highlight requirements which could otherwise prove difficult for potential users to articulate; valuable data which may never have been discovered through interviews, and certainly not through the use of walkthrough tests, can be discovered [19].

Yet, despite the use of such methods within more general HCI research, they remain largely overlooked to specifically inform auditory interface design, an exception being Macaulay and Crerar's study of work practice and the uses (current and potential) of the "soundscape" of the offices of a national daily newspaper [9]. In this paper, we highlight reasons as to why this could be the case and suggest how potential problems could be tackled. We begin, however, by turning our attention to a traditional ethnographic approach in which the work practices of a sound department within a major UK-based computer games developer were examined.

3 In the Field

In this section, we present a summary of a work practice study with VIS Entertainment Ltd, a UK-based computer games developer formed in the mid-1990s and with a workforce of around 120 employees, and additional interviews held with sound designers and their colleagues from other developers based both in the UK and in the US. Due to space restrictions, we provide only a brief summary of our experiences in this paper; for a more complete description, see [20].

3.1 The Case for Computer Games

In 1997, Gaver [21] noted that research into the effective use of sound within multimedia and computer games design could greatly inform the design of more general auditory interfaces. Yet, while studies exist to show the more general features of computer games, such as user performance and satisfaction which could help to improve application design satisfaction (for example, [22] and [23]), the specific characteristics of computer game audio have been largely overlooked. Despite this, there are many opportunities for further research, which we briefly provide here.

To begin with, Rollings and Adams [24] suggest that, while sound is seen to be in "third place" behind visual and interactive elements of a computer game, many games are deemed to be unplayable without it. While discussing whether or not the use of audio improves game playability falls beyond the scope of this paper, it is conceivable that the increasingly advanced audio capabilities of games consoles, and the vast majority of commercially available games which take advantage of such capabilities, demonstrate what sound has to offer. Furthermore, given increasing recognition within the wider community towards human factors such as pleasure [25], emotion [26] and fun [27], it is arguable that the ability of computer games, and consequently computer game sound, to induce such factors in the player offers useful insights to designers of other forms of interactive technologies. Indeed, not only should computer game sound be *functional*, i.e. to direct the user towards undertaking certain tasks, but it should immerse the player within the environment and, hopefully, encourage them to return to the game again and again. Hence, designers of auditory interfaces could gain insights into how games are more playable through the use of sound.

Naturally, there are differences between computer games and more general applications. To begin with, computer games by their very nature are required to *challenge* the player; a game perceived as too easy may not be attractive to an experienced games player, whereas easy navigation may be perceived to be the goal within other forms of interaction design. Consequently, it is arguable that computer games serve an altogether different purpose. Federoff [28] suggests that computer games are purchased on a voluntary basis, whilst other forms of software will be purchased to perform necessary tasks. Furthermore, Pausch *et al* [29] suggest that users of application software are generally motivated to overcome poor design to complete a task, whereas the lack of external motivation while playing a computer game can lead to games with inadequate interfaces failing in the marketplace. With those thoughts in mind, the computer game industry appeared to be an appropriate domain to study.

3.2 Methods

During our field study at VIS, we concentrated primarily upon the audio department, made up of three members of staff: a Head of Audio and two sound designers. We observed their working methods and carried out unstructured interviews with those staff members and, where possible, other members of the development team. Visits took place once a week, depending upon the availability of participants, and lasted three to four hours at a time. Extensive field notes were taken from the observations and interviews, and were then "coded", a method whereby data are broken down to find themes. In addition to the work practice study, we made contact with a small number of participants employed by other computer games developers in order to carry out semi-structured face-to-face and e-mail interviews. These participants were made up of a programmer and a sound designer based at two separate computer games developers in Dundee, and a freelance sound designer based in San Francisco.

3.3 Outcome and Discussion

While this was a relatively modest study, we discovered that a study of this nature offered numerous insights for the auditory interface designer. To begin with, not only do we feel that auditory interface designers can learn about the *creative* ways in which sound is designed and implemented within computer games industry, but we also feel that a study of this nature also highlights transferable working methods. For example, outside of computer game development, there is still an assumption that sound design and implementation is carried out towards the end of the development process [30]. However, while this may be true of some game developers, this is not the case at VIS. The sound designers communicated regularly with other members of the development team to ensure that implementation problems can be prevented at an early stage. This is facilitated through formal documentation and procedures such as milestones, regular meetings, the "game document" (a 50-200 page long non-technical document detailing the creative, conceptual and functional aspects of the game), as well as informal communication such as e-mail. Of course, as Back [31] notes, sound design cannot exist in a vacuum. Back suggests that sound design differs from music, as the sound designer must take into account the design gestalt of the entire game, not just the sound itself. Therefore, from a more general auditory interface design perspective, how could collaboration between the user interface designers and those working on audio be increased?

A qualitative study of this nature occasionally had some drawbacks. Firstly, as there were only a few staff in the audio department, they were normally extremely busy, thus only one visit per week was possible. This meant that the study was what Hughes *et al* term a “quick and dirty ethnography” [32], i.e. one in which the study is short and focused in order to gain a general picture of the setting, and thus concentrating upon the portions of the work setting of particular importance in informing design. Similarly, because of the competitive nature of the computer games industry, it became clear that sound designers often appeared reticent about providing too much information about “what they do” for fear of breaking their non-disclosure agreements. Finally, we were rarely able to speak to and observe the other members of the development team at VIS, simply because the audio department was located in an entirely different building from the other employees. Nevertheless, we believe that a study of this nature highlights the opportunities ethnographic methods have towards auditory interface design.

4 “Observing” Sound?

Current ethnographic techniques offer little insight into understanding the day-to-day role sound plays in people’s lived experiences. This can be attributed partly to the visual bias of much ethnographic work and method. Studying the importance of birds to the Kaluli of Papua New Guinea, Feld observed that “when asked direct questions that include the name of a bird, the response ‘it sounds like X’ is universally presented by a Kaluli before any sort of ‘it looks like X’ statement” [33]. Despite being a musician himself, Feld recounts his struggle to get past his own visual bias and understand that, to the Kaluli, the primary sense through which the world is conceptualized and related to is auditory, not visual. Hence, other than the relatively small community of anthropologists concerned with “the anthropology of the senses”, of which Feld is a good example, sound as a phenomenon and listening as an activity receive scant attention.

Similarly, the design skills of sound designers are not easily transferred [34], thus scoping and probing sound related fieldwork can be difficult. Jordan [35] notes that the majority of the participants involved in the sound design process may not be what he terms “sound professionals”. In addition to sound designers, product designers, market researchers, marketing personnel and product managers may be involved, all with sound related interests in mind, but without a full understanding of the formal properties of sound such as pitch, timbre and volume. Additionally, given the general problem facing design ethnography as a whole, that is successfully bridging the gap between the expansive texts of the fieldworker and the comparatively sparse depictions seeking to focus upon the “key elements” of the designer [18], how can we apply the resultant findings of a sound related field study into an appropriate sound design?

4.1 Soundscapes and Sound Classification Tools

In our work, we are examining the use of methods traditionally associated with the classification of the *soundscape* to expand upon traditional ethnographic techniques and thus to understand how potential users interact within their current auditory environment. The term “soundscape” is derived from “landscape”, and is described by Schafer as “any portion of the sonic environment regarded as a field for study” [36], and by Truax as "how [the sonic] environment is understood by those living within it" [37]. Schafer introduced the notion of soundscapes as a response to what he felt was the dominance of the visual modality in (predominately Western) society, to such an extent that he believed the ability of children to listen to their environment was deteriorating (discussed further in [38]). As a result, Schafer [36] formalizes a terminology for analyzing soundscapes. The most commonly used terms are as follows:

- Keynote sounds – these are sounds heard frequently and continuously enough to form a background against which other sounds can be perceived. For example, the soundscape of a beach is likely to contain the sound of the ocean as a keynote sound;
- Sound signals – these are foreground sounds listened to consciously, that is, those designed specifically to attract attention. For example, warning sounds such as car alarms are identifiable as sound signals;
- Soundmarks – these are "community sounds" which are unique to specific environments, for example, waterfalls or the sounds of traditional activities within the context.

Of course, given the polysemic nature of sound semantics, and given the earlier point that those involved in the design process may not all be sound professionals, tools are required to help *describe* the sonic events heard. To do this, there are numerous soundscape classification schemes we can turn to. One such scheme is Gaver's framework for everyday listening [39] in which everyday sound events are classified into three groups of "interacting materials". He classified those materials into *vibrating objects* (subcategorized into impacts, scraping and "others"), *aerodynamic sounds* (explosions and continuous) and *liquid sounds* (dripping and splashing). This concept is similar to Chion's "causal listening", i.e. "listening to a sound in order to gather information about its cause (or source)" [40]. More recent classification schemes which can be referred to include Özcan and van Egmond's perceptual framework [41], in which descriptions for domestic product sounds are classed into eleven semantically different groups such as *action* (for example, opening a door), *meaning* ("Wake up!") and *onomatopoeia* ("rattling"). In a separate study, Jordan and Engelen [42] describe twenty-five pairs of "sound personality descriptors" to describe the phenomenological properties of sounds to act as a means of communication between the designer and other stakeholders. Examples include feminine/masculine, strong/weak and dirty/clean sounds. Participants in their evaluation session indicated that they found those terms meaningful, and there was a high level of agreement about the relationship between the descriptions and particular sounds. Furthermore, as such descriptors relate to the experiential, rather than formal, properties of sounds, they can be a useful means of communication between individuals who may not be sound professionals.

4.2 Applying Soundscapes to Ethnography

To apply the concepts behind soundscapes to ethnography, Macaulay and Crerar [9] developed a "mapping" tool, as shown in Fig. 1, to assist fieldworkers to both describe the soundscape of a particular environment and to enable them to think of the ways in which any future sound design could "fit" into this context.

Fig. 1. A soundscape mapping tool developed by Macaulay and Crerar [9]

Using this tool, the soundscape can be characterized into three dimensions: *sound type* (broken down into music, speech, non-speech abstract and non-speech everyday sounds), *acoustical information* (broken down into background, foreground and contextual) and *informational* (broken down into visible, hidden and imagined entities/events, the passing of time, position in space, patterns in entities/events and emotions). As an example, if this tool were to be used in an office environment, it can supply information about:

- Visible entities/events – a ringing telephone on a desk;
- Hidden entities/events – distant doors opening and closing;
- Imagined entities/events – this is a busy period for the office;
- Patterns of entities/events – the constant whir of the photocopier printing;
- Passing of time – it is nearly 5pm, so everyone is hurriedly completing their daily tasks;
- Emotions – it is a stressful time of the day;
- Position in space – the ringing telephone is behind me.

How, therefore, could this tool be used to consider how a *future* sound design could fit into a particular context? McGregor *et al* [34] suggest a number of potential uses. Firstly, such a tool can contribute to the design of “intelligent noise maps”, in which sounds that are necessary or desirable and those that are background or redundant by those who inhabit the environment can be identified. Furthermore, the use of such a tool could be used to identify how *additional* sound generating artifacts could affect the existing auditory environment; for example, it could help to identify sounds it has to compete with. Of course, as McGregor *et al*

[34] note, no two different time periods will ever have a completely identical soundscape. However, it is arguable that there will be numerous "fixed" sounds (or events) that remain relatively constant during consecutive periods (for example, traffic noise heard in downtown offices).

4.3 Discussion and Future Research

In our research, we are attempting to expand upon traditional ethnographic techniques to develop pragmatic approaches to capture and interpret soundscapes in a way that is intuitive and accessible to a wide range of designers who may or may not have formal expertise with sound. To do this, it is necessary to consider aspects of the environment inhabited by potential users that we need to focus attention upon. We open this discussion by providing an example list of aspects requiring such attention:

- Overall Context: What is the nature of the context in which our sound designs will eventually be used – a "serious" working environment, a home environment, or a mobile environment? How will additional sound affect the context? These are essential issues to address, as the optimal goal is to complement and enhance the existing soundscape rather than to disrupt it.
- Individuals: Who inhabits the environment? Who are we actually designing for – children, office workers, older users? Do they have visual impairments that could be supported through sound?
- Activities: What are the activities undertaken within this environment? How are such activities undertaken? What tools/technologies are used? Is sound currently employed within those tools to aid individuals and, if so, how? How useful is it? Are there any "breakdowns" (a term used by Beyer and Holtzblatt [43] within contextual design to refer to problems in communication or coordination) which are caused by the sound, or could these breakdowns be overcome through the use of alternative auditory cues? Here, ways in which sound can be used to support *existing* activities can be identified at an early stage. Additionally, the ways in which sounds *are currently* employed within specific artifacts can act as a basis for discussion between the fieldworker and participants as to their usefulness, or even how participants perceive the artifact as a whole.

The next stage of our research is therefore to examine practical methods to tackle some of these issues. Our intention is to ensure such tools are intuitive and accessible to designers, that they can bridge the gap between the data collected and eventual sound design requirements, and that they can allow for studies to be undertaken within a relatively short period of time. At the time of writing, we are still examining ways in which this could be achieved, although there are numerous examples in the literature, from both within and outside of auditory interface research. For example, Baillie [44] created a method for investigating the use of technologies within the home known as the "technology tour", in which families took the researcher on a tour of their homes, thus providing details as to how the technologies were used, problematic situations they had experienced with the technologies, and to understand how the technologies fitted into the lives of the participants. A "soundscape tour", asking participants to describe how they use sound within their environments, may provide equally useful insights for the design of auditory interfaces. Additionally, everyday stories created by participants can act as a starting point for sound design. The use of "earwitness" accounts in soundscape literature can often act as a basis for discussion, for example, Berendts "summer experience" in which he describes an August afternoon by a lake *without* turning to visual impressions [45]. Within auditory interface design, the use of such accounts is not particularly new; indeed, the EarBenders approach developed by Barrass [46] is an approach using short stories about everyday listening to inform design. We believe that such approaches encourage participants to listen and to interpret the soundscape around them and what it means to them, thus highlighting the ways sound permeates their everyday lives. Such stories could confirm attitudes towards sound producing events, as well as additional phenomenological descriptions; for example, participants may find birdsong "relaxing" while the sounding of car horns may be seen as "threatening". Furthermore, while auditory interface designers can use these stories as a starting point for sound design, borrowing elements where required, participants (who may be eventual users of the system) are encourage to think closely about how they "use" sound, and thus are able to better communicate sound design requirements. Finally, the development of possible "future use scenarios" between users and auditory interface designers can encourage both parties to think not only about *how* sound could be used (for example, which functions it should support), but *what it should sound like*, thus allowing both parties to envisage new and interesting ideas for sound design.

5 Conclusions

In our research, we are examining appropriate and intuitive ways to incorporate ethnographic research into the design of auditory interfaces. Little attention has been given to the use of such methods within this domain; however, given that technologies to enable auditory interaction have matured to a level that design issues must be considered, we believe the time is right to widen the scope of the research agenda to consider the role sound plays in people's everyday lives. Here, we suggest how traditional ethnographic approaches can be employed and expanded to specifically identify the role sound has to play in everyday life. We believe an acknowledgement of such approaches should lead us to our shared ultimate goal: to ensure sound can contribute to more usable and useful auditory interfaces.

Acknowledgments

The authors would like to thank the participants within this study for their time, cooperation and patience.

References

1. Blomberg, J., J. Giacomi, A. Mosher, and P. Swenton-Hall, *Ethnographic Field Methods and their Relation to Design*, in *Participatory Design: Principles and Practices*, A. Namioka and D. Schuler, Editors. 1993, Lawrence Erlbaum Associates: Hillsdale, NJ, USA. p. 123 - 154.

2. Norman, D.A., *The Invisible Computer*. 1998, Cambridge, Massachusetts, USA: MIT Press.
3. Bannon, L.J., *A Human-Centred Perspective on Interaction Design*, in *Future Interaction Design*, A. Pirhonen, et al., Editors. 2005, Springer-Verlag: London. p. 31 - 51.
4. Hallnäs, L. and J. Redström, *From Use to Presence: On the Expressions and Aesthetics of Everyday Computational Things.* ACM Transactions on Computer-Human Interaction (TOCHI), 2002. **9**(2): p. 106 - 124.
5. Brown, J.S. and P. Duguid, *Borderline Issues: Social and Material Aspects of Design.* Human-Computer Interaction, 1994. **9**: p. 3 - 36.
6. Cooper, G., C. Hine, J. Rachel, and S. Woolgar, *Ethnography and Human-Computer Interaction*, in *The Social and Interactional Dimensions of Human-Computer Interaction*, P. Thomas, Editor. 1995, Cambridge University Press: Cambridge. p. 11 - 36.
7. Plowman, L., Y. Rogers, and M. Ramage. *What are Workplace Studies For?* in *ECSCW '95: Proceedings of the 4th European Conference on Computer-Supported Cooperative Work*. 1995. Stockholm, Sweden.
8. Green, T.R.G., S.P. Davies, and D.J. Gilmore, *Delivering Cognitive Psychology to HCI: The Problems of Common Language and of Knowledge Transfer.* Interacting with Computers, 1996. **8**(1): p. 89 - 111.
9. Macaulay, C. and A. Crerar. *'Observing' the Workplace Soundscape: Ethnography and Auditory Interface Design*. in *Proceedings of the 1998 International Conference on Auditory Display*. 1998. Glasgow, UK: ICAD.
10. Bannon, L.J. and S. Bødker, *Beyond the Interface: Encountering Artifacts in Use*, in *Designing Interaction: Psychology at the Human-Computer Interface*, J.M. Carroll, Editor. 1991, Cambridge University Press: New York, N.Y., USA. p. 227 - 253.
11. Bannon, L., *Activity Theory*. 1997, University of Limerick: Limerick, Ireland, http://www-sv.cict.fr/cotcos/pjs/TheoreticalApproaches/Actvity/ActivitypaperBannon.htm.
12. Hughes, J., V. King, T. Rodden, and H. Andersen. *Moving Out from the Control Room: Ethnography in System Design*. in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. 1994. Chapel Hill, North Carolina, USA: ACM Press.
13. Macaulay, C., *Inside the Palimpsest: A Study of Newsroom Information Gathering*. 2000, PhD thesis, Napier University, Edinburgh, Scotland.
14. Boyle, J.S., *Styles of Ethnography*, in *Critical Issues in Qualitative Research Methods*, J.M. Morse, Editor. 1994, Sage Publications Inc.: Thousand Oaks, California, USA. p. 159 - 85.
15. Chambers, E., *Applied Ethnography*, in *The Handbook of Qualitative Research*, N.K. Denzin and Y.S. Lincoln, Editors. 2000, SAGE Publications Inc.: Thousand Oaks, California, USA. p. 851 - 869.
16. Wolcott, H.F., *The Art of Fieldwork*. 1995, Walnut Creek: Altamira Press.
17. Miles, M.B. and A.M. Huberman, *Qualitative Data Analysis*. 1984, Beverly Hills, California, USA: SAGE Publications Inc.
18. Hughes, J.A., J. O'Brien, T. Rodden, M. Rouncefield, and S. Blythin. *Designing with Ethnography: A Presentation Framework for Design*. in *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques*. 1997. Amsterdam, The Netherlands: ACM Press.
19. Burke, J. and A. Kirk, *Ethnographic Methods*. 2001, University of Maryland: Maryland, USA, http://www.otal.umd.edu/hci-rm/ethno.html.
20. Coleman, G.W., C. Hand, C. Macaulay, and A.F. Newell. *Approaches to Auditory Interface Design - Lessons from Computer Games*. in *Proceedings of the 2005 International Conference on Auditory Display*. 2005. Limerick, Ireland.

21. Gaver, W.W., *Auditory Interfaces*, in *Handbook of Human Computer Interaction*, M.G. Helander, T.K. Landauer, and P.V. Prabhu, Editors. 1997, Elsevier Science B.V.: Amsterdam, The Netherlands. p. 1003 - 1035.
22. Malone, T.W. *Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games*. in *The 1982 Conference on Human Factors in Computing Systems*. 1982. Gaithersburg, Maryland, USA: ACM Press.
23. Dyck, J., D. Pinelle, B. Brown, and C. Gutwin. *Learning from Games: HCI Design Innovations in Entertainment Software*. in *The 2003 Conference on Graphics Interface*. 2003. Halifax, Canada.
24. Rollings, A. and E. Adams, *Andrew Rollings and Ernest Adams on Games Design*. 2003, Boston, Massachusetts, USA: Pearson Professional Education.
25. Green, W.S. and P.W. Jordan, eds. *Pleasure with Products: Beyond Usability*. 2002, Taylor & Francis: London.
26. Norman, D.A., *Emotional Design: Why We Love (Or Hate) Everyday Things*. 2004, New York: Basic Books.
27. Blythe, M.A., K. Overbeeke, A.F. Monk, and P.C. Wright, *Funology: From Usability to Enjoyment*. 2004, London: Kluwer Academic Publishing.
28. Federoff, M.A., *Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games*. 2002: MSc Thesis, Department of Telecommunications, Indiana University, Bloomington, Indiana, USA.
29. Pausch, R., R. Gold, T. Skelly, and D. Thiel. *What HCI Designers Can Learn from Video Game Designers*. in *CHI '94 Conference Companion*. 1994. Boston, Massachusetts, USA.
30. Deutsch, S., *Music for Interactive Moving Pictures*, in *Soundscape: The School of Sound Lectures 1998 - 2001*, L. Sider, D. Freeman, and J. Sider, Editors. 2003, Wallflower Press: London, UK. p. 28 - 34.
31. Back, M.J., *Designing the Auditory Narrative: A Methodology for Sound Design*, in *Audio Anecdotes II: Tools Tips and Techniques for Digital Audio*, K. Greenebaum and R. Barzel, Editors. 2004, A K Peters, Ltd.: Wellesley, MA, USA. p. 305 - 328.
32. Hughes, J., V. King, T. Rodden, and H. Andersen, *The Role of Ethnography in Interactive Systems Design.* Interactions, 1995. **2**(2): p. 56 - 65.
33. Feld, S., *Sound and Sentiment: Birds, Weeping, Poetics and Song in Kaluli Expression*. 2nd Edition ed. 1990, Philadelphia: University of Pennsylvania Press.
34. McGregor, I., A. Crerar, D. Benyon, and C. Macaulay. *Soundfields and Soundscapes: Reifying Auditory Communities*. in *Proceedings of the 2002 International Conference on Auditory Display*. 2002. Kyoto, Japan.
35. Jordan, P.W., *Designing Pleasurable Products*. 2000, London, UK: Taylor & Francis.
36. Schafer, R.M., *The Soundscape: Our Sonic Environment and the Tuning of the World.* 1994, Rochester, Vermont, USA: Destiny Books.
37. Truax, B., *Acoustic Communication*. 2001, Westport, CT, USA: Ablex Publishing.
38. Wrightson, K., *An Introduction to Acoustic Ecology.* Soundscape, 2000. **1**(1): p. 10 - 13.
39. Gaver, W.W., *What in the World do we Hear? An Ecological Approach to Auditory Event Perception.* Ecological Psychology, 1993. **5**(1): p. 1 - 29.
40. Chion, M., *Audio-Vision: Sound On Screen*. 1994, New York, USA: Columbia University Press.
41. Ozcan, E. and van-Egmond. *Characterizing Descriptions of Product Sounds*. in *Proceedings of the 2005 International Conference on Auditory Display*. 2005. Limerick, Ireland.
42. Jordan, P.W. and H. Engelen. *Sound Design for Consumer Products*. in *Hey, Listen!* 1998. Stockholm, Sweden: Royal Swedish Music Academy.

43. Beyer, H. and K. Holtzblatt, *Contextual Design: Defining Customer-Centered Systems*. 1998, San Diego, CA: Academic Press.
44. Baillie, L. *The Home Workshop*. in *Home-Orientated Informatics and Telematics International Working Conference*. 2003. California, USA.
45. Berendt, J.E., *Third Ear: On Listening to the World*. 1988, Dorset, UK: Element Books Ltd.
46. Barrass, S., *Auditory Information Design*. 1997, PhD thesis, Australian National University, Canberra, Australia.

An Activity Classification for Vibrotactile Phenomena

Conor O'Sullivan[1] and Angela Chang[2]

[1] 600 North US Highway 45, DS-175, Libertyville, IL 60048, USA
conor.o'sullvan@motorola.com
[2] 20 Ames St. Cambridge, MA 02139 USA
anjchang@media.mit.edu

Abstract. We observe that the recent availability of *audio-haptic* actuators allow richer vibration content to be available in commercial devices. However, we note that consumers are unable to take advantage of these rich experiences, mainly due to the lack of a descriptive language for vibration. We analyze the current methods for classifying vibrations. We propose a new framework for describing vibrotactile haptic phenomena, based on an organizing the media based on content activity. We describe this naming system, based on Russolo's families of noise, and address other pertinent issues to introducing vibration content into commercial devices.

1 Introduction

Today there is a gap in the community and industry of haptic actuation technology (specifically vibrotactile sensations) that needs to be addressed to allow a more general adoption of and dedicated engagement with the technology. In various forums of research, presentation and informal discussion it was found that although the average consumer or target adopter of such technology seems comfortable or even enthusiastic with their experience of haptics, the extent to which they are able to communicate and describe this experience seems very limited. The language of describing haptic phenomena in consumer electronic devices seems thus far limited to a small selection of words such as 'buzz' or 'vibrate', even as the devices are evolving to allow richer experiences (such as gesture input, vibration enhanced ringtones, and UI feedback).

To understand how a general population can talk about haptics they need to understand how to create and distinguish among haptic effects. The correlation between audio and haptics in the case of the mobile device has already been shown [6]. Audio-haptics (driving vibrations using audio signals [27]) for consumer devices can be generated by audio manipulation and synthesis [15], but are perceived tactilely as a more advanced form of vibration. It is of course sufficient to speak of vibe or buzzing in the general case of vibration alert mechanisms. Given the range and depth of current and future haptic actuation technologies however, it is imperative that a more user-friendly, sensorial descriptions are available.

This will allow the general consumer to become knowledgeable, interested and proficient in the language of haptic description. A consumer armed with such proficiency

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 145–156, 2006.

and enthusiasm can only serve to drive interest and consequently propagate the development of the technology and of design. This type of language can develop from new or existing forms of haptic engagement, but the industry may be better served in the medium term to associate with 'real world' or ecologically familiar constructs.

2 Motivation for Describing Vibrotactile Phenomena

Haptics is the science of touch, encompassing both kinesthetic and tactile phenomena. The study of haptics is a wide field, some areas of research are: 3D force-feedback actuation technologies, perception of touch (psychophysics), tactile display of shapes and textures, temperature sensing, and awareness and location relationships between body parts. Because of the broad applications of haptics, terminology between different subsections of haptic research are usually specific to their sub disciplines, and may not apply to other areas of research (e.g. perception measurements of vibrotactile displays might not apply to 3D force feedback displays).

For the purposes of this paper, we refer to the specific phenomenon of *vibrotactile stimulation* (or *vibes*), tactile stimulation using vibration. We address this particular area due to the recent availability of better vibration actuators in consumer products that allow for greater control over tactile sensation [6, 19]. Increased resolution and variation enables more dynamic range in the vibration content. This new expressiveness could be used as a selling point for differentiating haptic products. The featured improvement of haptic resolution could be appreciated, providing consumers could differentiate among complex vibrotactile stimuli.

2.1 Toward a Generalized Knowledge of Haptics

The combination of temporal and spatial information provided by haptics to reinforce perception and communication is ubiquitous. For many years, vibration in commercially available mobile devices, such as pagers and mobile phones, alerted people of calls or messages. The ubiquity of vibrating alerts in mobile devices has also misled many consumers to believe haptics is equivalent to vibration. Furthermore, when the authors demonstrated a vibration-enhanced mobile phone to novice users, general comments received about haptics provide an illustrative example of the lack of public haptic awareness and vocabulary.

There appeared to be a large vocabulary gap between the scientific and HCI community and haptics understanding in the common usage. Quality in consumer products often relates directly to touch (weight, surface finish, and contours). However, there is little vocabulary to describe haptic sensations, in particularly vibrations. When asked to describe haptic quality, a surprising number of people would relate haptic vibrations to audio [6]. Many people simply said "it feels different", but were unable to say why.

This lack of expressiveness about haptics is surprisingly minimal in contrast with that of the sense of smell. Before the days of aromatherapy, there was little conscious awareness of the cause and effect of smell. In fact, until marketing began to become popular, body odor was not considered a problem [18]. People did not typically describe smells, nor did they obsess about altering them. Nowadays, people can identify common scents in products (e.g. cinnamon and vanilla) and they can also understand smells can be controlled and distinguish differences between smells [1]. The common vocabulary for smells has increased from merely "strong or noxious" to a myriad of descriptive phrases "lavender, rosemary, bergamot, etc". Furthermore, there is a vast amount of consumer interest in creating scents (one can easily find recipes online for air fresheners, relaxation, and medical therapy). The large perfume market gives evidence to the fact that "smell is a commodity" [18, p. 62]. We are equally interested in how to create enlightened tactile consumers, a *tacterati,* so to speak.

2.2 Raise Haptic Consciousness

One way to raise consciousness is to create objects that stimulate the haptic senses more directly. Various force-feedback and vibration gaming joysticks have incorporated haptics sensations for years. Certain phones include haptics as a feature inherent in audio quality. Digitally controlled haptics is subtly integrating into consumer products, but it is still very hard for consumers to distinguish and disseminate haptic sensations. Could this lack of definition be due to the dearth of vocabulary to describe haptic qualities in our culture?

The visualization of haptics might be a challenge because of the lack of awareness of haptic phenomenon in popular culture. It is rare to find language for distinguishing the roughness of textures from one another (e.g. dotted pattern versus serrated rows). There are few common ways to describe vibration profiles (sawtooth versus sine waves). For example, try to describe the difference in tremors in the floor due to heating systems "oscillations", traffic "rumblings" and earthquakes "shockwaves". Even among researchers in haptics, there are few *generalized* terms to describe force profiles or vibration envelopes.

A quick survey of the consumer marketplace finds that many haptic experiences are targeted to the high-end of consumer product lines. Force feedback joysticks are high-priced and geared toward the gaming public. Devices touted for their tactile qualities, such as high-end audio equipment, are more expensive than offerings for the general public. In most cases, devices displaying greater vibration control are targeted to special-needs users [2]. The target market for haptic devices seems to be very specialized and narrow.

Perhaps social consciousness will be raised when product literature incorporates descriptive expressions for haptic information that can be repeated and discussed by the general population. Once the public can express the value of haptic sensations, haptic visualization may be more common, and haptics might become a commodity by which differentiates products.

2.3 Haptics in Infancy Parallels Sonic Development and Language

We propose a lexicon beyond buzzes, pulses, and shaking of the pager motor era. Given that the lexicon of haptic description is in its infancy, a correlation may be drawn with the industry of olfaction around aromatherapy (as previously described) or perhaps general electronic audio in the first half of the last century:

2.3.1 Composers and Theorists

As early instruments such as the Telharmonium began to appear, people such as Ferrucio Busoni saw a role for such machinery in the future synthesis of music and sound. This was acted on more expressly by the Futurist movement and artists such as Luigi Russolo, who realized the musical, sonic and social value of noise [20]. Russolo began to develop a language of sound and of noise, expressing his categorization of noise into six categories. This will be looked at later, as an example of a basis for haptic expression.

Edgard Varese was another artist who approached music as an instantiation of sound and consequently looked to expand the range of techniques of synthesis, crying out for electronic synthesis technology until it became available practically, to realize his auditory design. The work of such pioneers and others including Schaeffer and Stockhausen, allowed the language and technique of electronic audio to emerge [26].

2.3.2 Auditory Interface Design

More recently work has taken place that looks at auditory interface design and how the perception and expression of such feedback can be realized. Gaver and others speak of ecological acoustic perception, to enable a user to associate sounds in the natural environment to the presentation of information from a computer interface [21, 22]. Blattner, Brewster and others have looked at earcons, or more musical motives, as a means of presenting information that has associated meaning either inherently or learned over time [23, 24, 25].

2.3.3 Haptic Vibrations Are Easy to Make, but Not Easy to Talk About

Currently in mobile devices the types of auditory feedback we experience includes voice, ringtones, music playback, user interface (UI) sounds and gaming. Among these types and even among UI sounds, we hear examples of both earconography and auditory iconography. Haptics in today's mobile devices where it occurs beyond basic alerting vibe mechanism is often mapped to the rhythm of the sound and so tends to exist as derivatives of both categories of sound feedback. It has been shown previously how haptics can be generated with and from sounds. Haptics therefore can be relatively straightforward to produce but as we've seen are not always so easy to discuss.

2.4 Study of Classifications for Tactile Phenomena

We present two examples of the benefits of tactile classification below, haptic exploratory procedures and tactile language displays.

2.4.1 Haptic Exploratory Procedures

The haptic exploratory procedures (EPs) demonstrate how touch can be both an active and passive way of interpreting information [13]. This classification of EPs has allowed researchers to generalize haptic procedures, and better understand the mechanical actions performed by our hands. Additionally, by organizing the knowledge gained from touch activities, we can compare haptics with other information-gathering modalities. Table 1 summarizes some qualities of haptics relevant to information gathering.

Table 1. Comparison of haptics attributes to audition and vision

	Touch	Audition	Vision
Information given relative to person	Useful for details smaller than a person	Useful for gauging distance relative to a person	Useful for gauging size and distance relative to a person
Temporal aspects	Identification by parts. Information is gathered by sensing information over time.		Vision uses both part and whole images to identify whole things
Active and Passive Identification	Active and Passive methods for information gathering (e.g. applying pressure or echolocation)		Passive only
Spatial aspects	One-to-one only, a close proximity sensory experience	Can be close or remote, can be broadcast	
Illusory Possibilities	Touch is hard to fake	Easy to represent, replicate, reproduce digitally	

By understanding the classification of EPs, we can describe how physical details are found by certain hand motions

2.4.2 Tactile Language Displays

Existing tactile languages can be subdivided into two classes of languages, alphabetic and symbolic. The first class, alphabetic language uses the representation of the alphanumeric letters to form words. Examples are chording keyboards, Braille, Moon, and telegraphs. The second class, symbolic language, represents higher-level concepts that are not mapped directly to words, but rather, abstract ideas and expressive emotions. Examples of symbolic language are facial expressions, hand gestures and body language for expressing interest and emotional state. Symbolic and alphabetic language can be combined in a language, such as in Morse code and fingerspelling. The combined methods enable users to transmit information more quickly than working with either method alone. Fingerspelling is more expressive than Morse code, using gestures made by one hand to rest on another hand, making full use of the EPs. Fingerspelling has faster transmission rates than Morse code. Sign language displays, when used with vision, are commonly used in public speeches because of the broadcast nature and quick speed of transmission.

In terms of vibrotactile displays, *Tadoma* allows users to receive intonation information directly from the vibration of a speaker's throat [16]. Vibrotactile devices, e.g. Tactaid, remap ambient audio frequencies to vibration channels. Combining vibration signals with visual cues allow faster transmission of information than the hand signing or alphabetic methods. Vibrotactile displays allow for quick transmission, but are private rather than broadcast. The unobtrusiveness of these displays has been employed in vibrotactile arrays to provide navigational and geographic information [11] and entertainment [9].

Classification schemes give information for *comparison* among groups of related items, *visibility* of differences between items, and give us a way to *control* communications about different subsets [28]. By classifying the tactile languages, comparisons can be made between the transfer rates of alphabetic languages and symbolic languages. Among researchers, there have been previous approaches to describe vibrotactile content. The need for a flexible, yet structured classification scheme becomes even more necessary as vibes are now mapped to buttons, ringtones, and other UI media in commercial devices. We present our experience using the following classification methods while working with audio-haptics for commercial products.

3 Example Descriptive Frameworks for Vibrotactile Phenomena

3.1 Descriptions of Vibes Using Compositional Elements

Like sound, the compositional elements of any vibration can be described by frequency, intensity, duration, waveform and space. The first method we tried for describing vibrations was to create vibration files from the compositional elements of frequency, envelope, and waveform shapes (sine, triangle, and square waves). The skin is most perceptive of intensity, duration and waveform, and many variations of these elements were created to create a library of haptic waveforms.

From testing of elements in the library, we confirmed much of the psychophysics testing that there was little correlation between a perceived "pitch" and frequency (unlike in audio). High frequencies (> 250 Hz) have higher pitches and feel smooth, while lower frequencies (around 100-150 Hz) sound like buzzes and feel rough. In general, a sine wave was perceived as "smoother" than a square wave. Gunther compares the variations of waveform to the "texture" of tactile stimuli [9]. Varying the waveform envelope was compared to adjusting the volume of a sound. The intensity of entrance and exit of the waveform, respectively called the *attack* and *release* angle. Sharper attacks were perceived as pulses, while smoother attacks eased the vibe into the users' focus of attention. Similarly, smooth releases were perceived as "fading away".

In all, although many hundreds of combinations were created in the laboratory for testing, it was not clear whether users were able to name a particular sensation. The differences between two similar compositions (e.g. 100Hz vs. 110Hz) were oftentimes indistinguishable to the end user. Users were able to distinguish waveforms of different envelopes much easier than the other compositional elements, but with names like "100 Hz sine wave fading for 0.5 seconds", though helpful for a sound designer, were too technical for an end-user.

3.1.2 Rhythm and Tempo Descriptors

Brewster, et.al, has experimented with applying musical techniques for varying the waveform envelopes in tactile icon design [4,24]. In their work, subtlety and variation are emphasized, along with integration with a user interface. However, the problem with assigning too many vibrations to an interface is that the skin can become overloaded, much like the ears. One way to distinguish vibrations was by varying rhythm and tempo. References to vibrations with "slow soft tempo" or "dot-dot-dash" using the Morse code-like descriptions are common in literature on vibration devices (e.g. haptic mouse textures[29]).

Rhythm and tempo descriptions are a more user-friendly way of describing haptic actuation than describing the compositional elements. One drawback of this technique is these descriptions are still abstract to the end-user. A user might not be familiar with the terminology of musical or rhythmic classifications, and further, the terminology is still symbolic of the vibration media. Issues of interest to us were "how does a car crash *feel*?" or "what should a button press *experience* be?"

3.1.3 Form Factor Based Descriptions

One main problem of vibrotactile simulations is that there can be a variation of form factors and arrangements of actuators. Various researchers have explored a vast number of ergonomic shapes (e.g. knobs, mice, phones, joysticks) and surfaces (e.g. rollers, vests, chairs, human bodies) for tactile stimulation [3, 7, 8, 12, 14, 17, 30]. Some additional physical language needs to be used to describe vibrations from devices with specific form factors. If more than single-point vibration is used, spatially oriented words such as "diagonal" "across" or "up down" are appropriate. One difficulty with describing vibrations on these form-factor descriptors is that there is a critical reliance on the exact implementation of the vibration.

3.1.4 Descriptions Used When Combined with Other Modalities

When combined with other modalities, it is usual for the descriptions of the haptic phenomena to be associated with the other modality. When spoken conversation is augmented with a vibrotactile channel, there are tactile gestures that mimic spoken interactions [5]. By sampling vibrations off of surface textures [14] or existing audio, users were able to attribute the surface or audio information to the vibration. For example, users can say, "I like the feel of that rough surface". When testing a vibration-enhanced motorcycle driving game, designers sampled the audio from a real motorcycle, the result was that the "motorcycle sound feels realistic".

In ActiveTV, O'Modhrain found that there was ambiguity of what character was associated with haptic content when there were many actors on a screen[31]. For example, if an animal kicked a ball into a tree, would the vibration represent the animal's foot kicking the ball, the ball bouncing, or the impact between the ball and the tree? These lessons suggested to us that vibrations might be described by activity.

3.1.4 Design Criteria for a New Vibration Classification Scheme

Though technically descriptive the prior methods described technical information (such as compositional elements and musical terms), or relied heavily on a particular form factor, which was not transferable to other devices, resulting in problems scaling the media across different hardware platforms. The last method of using language from another modality seemed to amplify the awareness of other modalities, but did not make users more aware of how to describe their haptic sensations. We believed that in order to bring vibration awareness to the forefront, one possibility was to use more activity-driven language to describe these vibrations.

4 Proposed Activity Classification for Vibes

4.1 Translating Sound and Vibe

Since haptics and sound co-exist on our mobile devices today it may be useful to look at the development of the language of sound and see how this can be applied to a new development of a language in haptics. One way to do this is to observe some of the

Table 2. Russolo's families of noise and a haptic activity classification derivative

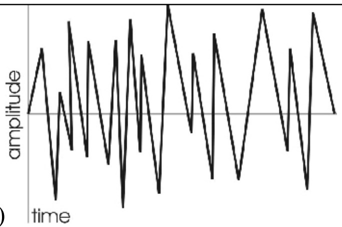
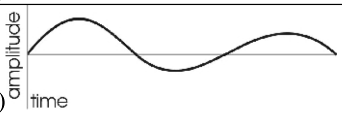
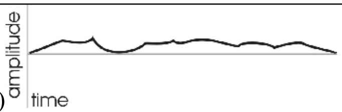
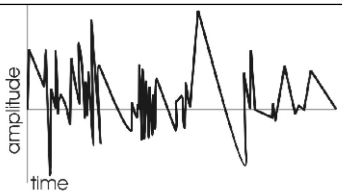
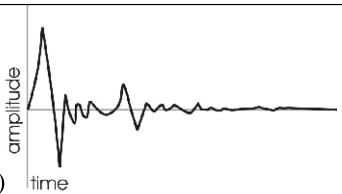
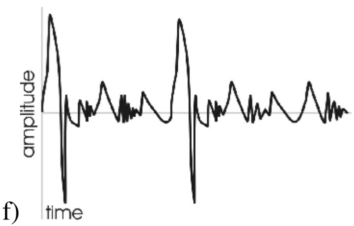
Russolo's families		**Haptic Textural Families (and example profiles)**	
1	Roars Thunderings Explosions Hissing roars Bangs Booms	**Quakes** Rumbles Coarse Grainy Calamity	a) 
2	Whistling Hissing Puffing	**Smooth** Slides, Buzzes Sheer Fades	b) 
3	Whispers Murmurs Mumbling Muttering Gurgling	Surface **Grazing** Speckles Taps, Tickles Light Touches Feathery	c) 
4	Screeching Creaking Rustling Humming Crackling Rubbing	**Complex** Textures Scrapes Strokes Kneading Squashing Folding	d) 
5	Beating on… Metals Woods Skins Stones Pottery	**Beats** Thumps Bangs Pulses Resonant Cavities Drops	e) 
6	Voices of animals and people Shouts, Shrieks Screams Wails, Sobs Hoots, Howls Death rattles	**Living** textures Scratch Shakes Claps Breath Chest resonance Howl, Interjection	f) 

Fig. 1(a-f). Typical sound profiles for the 6 classes of activities

thinking around the beginnings of the electronic sound movement and see what parallels, if any, can be drawn. Luigi Russolo, one of the early proponents of music as an instance of noise, published his manifesto in 1913 describing families of noise he considered groups that were fundamental in noise [20]. Here we look at these and show how a similar grouping categorization can be derived and applied to haptic textures. This is intended to serve as a basic framework by which a haptic language could develop. It is not intended to be a definitive list of haptic textures, but rather a starting point for thought and further structure to grow. Table 1 shows our proposed classification method for haptic vibrations.

4.2 Sample Vibe Classification and Waveform Shapes

To explore more fully some ecological examples of these haptic textures, it is useful to apply real world examples such as

Group One, Quakes: Riding a bicycle over tracks
Group Two, Smooth: Ice Skating
Group Three, Grazing: Goosebumps
Group Four, Complex: Picking up an object, interacting with the object
Group Five, Beats: Bang on the drum, a tap on the shoulder
Group Six, Living: Feeling resonance in your chest

The textures of groups 4 and 6 are related and could perhaps together form a wider group. They comprise the group of textures that make up a large part of our daily haptic experiences. All these groups can be represented visually as audio-haptic waveform representations as shown in figures 1a-1e. Quakes typically have coarse textures, large amplitudes, and sustained energy oscillations (fig. 1a). Smooth textures tend to have small changes in amplitude (fig. 1b). Grazing vibes are characterized by low amplitudes, perhaps at the thresholds of perception (fig. 1c). Complex textures encompass the wide range of activities where impulses and noise compete for attention. Beats are impulse-driven, quickly damped resonances (fig.1e). Living vibes are characterized by their periodic nature (fig.1f).

By using these classifications, we can lump different activities with typical profiles, and find that within the groups, substitutions in waveforms can be made. In answer to the previous question, “what should a button press experience be?” our answer in to use media from group 5, highly damped impulses. When a vibe is needed for moving a pointer across a screen, we can select from smooth (group 2) when there is nothing of interest and then use grazing textures (group 3) when they are over a button. Similarly, when the system has a notification message, it may be appropriate to either use a beat texture (group 5) or a start-up or power-down quake (group 1). For some ringtones, or subtle alerts, we may use living textures under audio (group 6) or quakes, if the situation calls for increased stimulation.

4.3 Ecological and Musical Considerations

Gaver lays out a framework for everyday listening that can provide a similar method for categorizing haptic textures. Gaver groups sounds firstly by Liquid, Solid and Gas categories and their Hybrid crossovers [21]. There are then further sub-groups given within each category. Some of these sub-groups seem immediately suited to haptic groups, such as deformation, impact, scraping and indeed we see these sound types have similarities to some of the haptic groups above. Similarly, a mapping of the earcon motive elements to the specific case of haptic feedback may be given by looking at features such as rhythm, timbre, and dynamics. These features can contribute to a haptic language by exploring their equivalents (or comparables) like rhythm, texture and strength.

4.4 Tools and Methods for Creating Haptic and Describing Phenomenon

It has been shown previously that textures can be explored and created for consumer electronic devices, such as the mobile device, that exist in the marketplace today. Though the resolution required to synthesize the feeling of some of these textures described above doesn't exist currently in mass market products, a broad range of effects and even a good sample set from each of these families can be created using tools already employed by any sound designer.

In our experience, having an activity based classification scheme works a bit better than the previous schemes. When developing vibes among a group of haptic novices, we have found that discussing "rumbles, slides, taps, scrapes, pulses, and echoes" helps make it easier for users, marketing people, and content designers to distinguish among groups of profiles. Classes of content (e.g. beats) are used among different form factors and UI elements (e.g. different buttons) because the profiles provide transparency into the types of sensations within a class. The descriptors seem to give visibility to the vibrations better than prior methods.

We are aware that with any classification scheme, there is a tradeoff between being too vague and too specific. Although technical information is not specified, the activity vibration classification seems to give developers and users some control to manage content relating to the different profile groups. Of course, more refinement may be needed as people become more familiarity with the sensations.

5 Conclusion

A general issue for haptic visualization is that there is currently little consumer understanding of haptic phenomenon. Further work will have to be done to raise the public consciousness and educate the world about understanding and describing the sense of touch. In this paper we proposed a new approach to classify vibrations, an activity classification scheme, influenced by the classification schemes of music. This approach is a more sensorial description and we hope, will be more descriptive to end-users.

Of course, the real test of the approach to framing a common language for haptic expression is its ease and success of adoption by the general 'haptic consumer'. The approach described herein seems useful and intuitive to the authors, but it remains to be seen how the consumer adapts to the taxonomy. The burden is on the promoters of the science not only to adopt this or similar methods but to reinforce the terminology and language so that it is readily associated with the senses. This can be achieved through consistent use over time, persistence and the penetrative force of marketing.

The ability to discuss, distinguish, and judge haptic phenomenon introduces users to the potential richness of these sensations. We also hope that once a framework for describing haptic effects is commonly employed, there might be interest and opportunity for users to create and share their own haptic effects.

References

[1] Ackerman, D; *A natural history of the senses*, New York: Random House, 1990, p.80.

[2] Auer, E.T., Jr, and Bernstein, L.E. Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals. *Journal of Acoustical of Society of America. 104, 4.* October1998, 2477-2489.

[3] Brave, S. and Dahley, A., inTouch: A Medium for Haptic Interpersonal Communication, Extended *Abstracts of CHI 1997*, ACM Press (1997), 363-364.

[4] Brewster, S. A., Wright, P. C., and Edwards, A. D. N., "Experimentally derived guidelines for the creation of earcons," in Proc. of HCI'95, Huddersfield, UK, 1995.Burdea, Grigore C. *Force and Touch Feedback for Virtual Reality*, John Wiley & Sons; New York, 1996.

[5] Chang, A., O'Modhrain, M. S., Jacob, R. J., Gunther, E., Ishii, H. ComTouch: design of a vibrotactile communication device. *Proceedings of Designing Interactive Systems 2002,* ACM Press(2002), 312-320.

[6] Chang, A. and O'Sullivan, C. Audio-Haptic Feedback in Mobile Phones. *CHI Extended Abstracts 2005*, ACM Press (2005), 1264-1267.

[7] Geldard, F., and Sherrick, C., The cutaneous rabbit: A perceptual illusion," Science, 178, 1972, 178-179.

[8] Geldard, F.A. Body English, Random House, 1967.

[9] Gunther, E. (2001). Skinscape: A Tool for Composition in the Tactile Modality. MIT MSEE Thesis, 2001.

[10] Ishii, H. and Ullmer, B. Tangible Bits: Toward Seamless Interfaces between People, Bits and Atoms, *Proc. of CHI 1997,* ACM Press (1997), 234-241.

[11] Jones, L., Nakamura, M. and Lockyer, B. Development of a tactile vest, *Proc. of Haptic Symposium 2004,* IEEE Press (2004), 82-89.

[12] Lécuyer, A., Burkhardt, J.M., and Etienne, L. Feeling bumps and holes without a haptic interface: the perception of pseudo-haptic textures, *Proc. of CHI 2004,* ACM Press (2004), 239-246.

[13] Lederman, S. J. and Klatzky, R. L. Hand Movements: A Window into Haptic Object Recognition, *Journal of Cognitive Psychology*; **19**, **3**, pp. 342-368, 1987.

[14] Minsky, M. D. R. Computational Haptics: The Sandpaper System for Synthesizing Texture for with a Force-Feedback Haptic Display. MIT PhD Thesis, 1995.

[15] O'Sullivan, C. and Chang, A. Dimensional Design; Explorations of the Auditory and Haptic Correlate for the Mobile Device, *Proceedings of International Conference on Auditory Display (ICAD 2005),* Limerick, Ireland, July 2005.

[16] Reed, C.M., Delhorne, L, and Durlach, N. A Study of the Tactual and Visual Reception of Fingerspelling, *Journal of Speech and Hearing Research*, **33**, December 1990, 786-797.

[17] Tan, H. Z., Perceptual user interfaces: haptic interfaces; *Communications of the ACM* **43**, **3** (Mar. 2000), 40 – 41.

[18] Twitchell, J.B. *Twenty Ads that Shook the World.* Crown Publishing Group, New York, NY, USA, 2001, pp.62,164-167.

[19] Verrillo, R.T. and Gescheider, G.A. Perception via the Sense of Touch: Chapter 1 from *Tactile Aids for the Hearing Impaired*, Ed. Summers, Ian). Whurr Publishers, London, 1992.

[20] Russolo, L. *The Art Of Noises*, Pendragon Press; New York, c1986.

[21] Gaver, W., Auditory Interface: Chapter from Handbook of Human-Computer Interaction, Eds. Helander, M.G., Landauer, T.K. and Prabhu, P., Elsevier Science; Amsterdam, 1997.

[22] Casey, M.A. (1998). Auditory Group Theory with Applications to Statistical Basis Methods for Structured Audio. MIT PhD Thesis, 1998.

[23] Leplatre, G., Brewster, S.A. (1998). *An Investigation of Using Music to Provide Navigation Cues*. In Proceedings of ICAD'98 (Glasgow, UK), British Computer Society.

[24] Brewster, S.A. *Chapter 12: Non-speech auditory output*. In Jacko, J. and Sears, A. (Eds.) The Human Computer Interaction Handbook (Lawrence Erlbaum Associates, USA), 220-239

[25] Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1)

[26] Kostka, S. *Materials and Techniques of Twentieth-Century Music*, Prentice Hall; New Jersey, NJ, 1999.

[27] Yano, H.; Igawa, H.; Kameda, T.; Muzutani, K.; Iwata, H., AudioHaptics: audio and haptic rendering based on a physical model, *Proceedings of the Intl. Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'04)*, 250- 257.

[28] Bowker, G.,Star, S. *Sorting Things Out: Classification and Its Consequences (Inside Technology)*, MIT Press, Cambridge, MA, 2000, 253.

[29] www.logitech.com

[30] MacLean, K. E. (2000). "Designing with Haptic Feedback", in *Proceedings of IEEE Robotics and Automatio*n (ICRA'2000), San Francisco, CA, 2000.

[31] O'Modhrain, S.,Oakley, I. Adding Interactivity: Active Touch in Broadcast Media, *Proceedings of the International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'04)*, 293- 294.

Haptic-Audio Narrative: From Physical Simulation to Imaginative Stimulation

Stephen Barrass

Sonic Communications Research Group, University of Canberra,
ACT Australia 2601
`stephen.barrass@canberra.edu.au`

Abstract. This paper describes the design and development of an interactive narrative for the 'Experimenta Vanishing Point' media arts exhibition in 2005. The Cocktail Party Effect tells the story of the imminent extinction of Great Apes in the wild using touch and sound in the absence of visual elements. The narrative is driven by haptic-audio exploration of a virtual cocktail glass which functions as a heterodiegetic narrator, and the traversal of cut-up conversations that make up the story within. The interface was developed through a series of prototypes that explored the perception and mental imagery of a haptic-audio simulation of the invisible glass. These experiments also developed narrative functions of the haptic-audio interface beyond conventional iconic metonyms to include grammatical and dramatic special effects. Observations during the exhibition show promising narrative engagement with the piece but identify problems with the clarity of the sounds, and a conflict between the narrator and the story content.

1 Introduction

Haptic interfaces allow museum visitors to touch virtual models of rare or fragile objects that are usually kept out of reach, such as a collection of teapots [McLaughlin et. al. 2000]. The tangibility of haptic objects is enhanced by simulations of the acoustics produced by interactions with virtual surfaces [DiFranco et. al. 1997] [DiFilippo and Pai, 2000] [Barrass and Adcock 2002]. The narrative functions of sound have been developed in cinema and computer games over the past century, but haptic narrative is still very much in its infancy. The combination of sound with haptics has potential to allow new forms of storytelling which do not necessarily rely on visual or verbal modes.

The Cocktail Party Effect is an experiment in haptic-audio storytelling designed for the 'Experimenta Vanishing Point' exhibition in September 2005, [Experimenta 2005]. The work tests the maxim that 'seeing is believing' through a haptic-audio simulation of a cocktail glass that has no visual component. Can the mental image of the curvy margarita glass be successfully conveyed without visual cues? It also explores the possibility that haptic-audio interfaces can be used for interactive narratives. The story of the imminent extinction of Gorillas in the wild is told through conversations at a party, unfolding in response to the haptic explorations of the object. What are the possible narrative functions of haptic-audio? How well can audiences understand a story told in this manner?

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 157–165, 2006.

The next section provides background on narrative theory, narrative sound, and haptic storytelling, and semiotics of multi-modal signs. The section after that describes the technical apparatus and iterative development of a series of prototypes for the Cocktail Party Effect. Finally the discussion presents observations from the Vanishing Point exhibition, and reflections on designing haptic-audio narratives.

2 Background

Narratives are either plot or character driven. A plot-driven narrative has a logical structure of cause and effect for sequential events. A character-driven narrative is organised by interactions between the motives and emotions of the characters. The storyteller or narrator may be outside the story (heterodiegetic) or may be a character in the story (homodiegetic). Narrative functions for sounds have developed during the past century of cinema. Diegetic sounds are sounds heard by the characters as well as the audience, and include the voices, objects and events in the story. Non-diegetic sounds are not heard by the characters and may include the narrator's voice, mood music, or dramatic effects. Recognisably characteristic soundscapes mark particular scenes or places in a grammatical manner, and are used in flashbacks. On-screen sounds support the suspension of disbelief, while off-screen sounds expand the spatial environment. Narrative sounds are also used in computer games to enhance on-screen and off-screen information, and for diegetic, non-diegetic, and grammatical purposes.

The dextrous hands of humans and apes are used to understand objects, manipulate tools and not least of all for interpersonal communication and expression. Although haptic interfaces have been developed for understanding and manipulation, for example in training surgical techniques [Gun et. al. 2005], the expressive and communicative possibilities are largely unexplored. There is a development of haptic narrative in Computer games, where the recoil of a gun, or a bump in the road, cause a vibration in a rumble pack or resistance in a steering wheel. The narrative function of the haptics is to refer to events, and the haptic icon is often synchronised with an audio icon that signifies the same event. These 'metonymic' icons refer to part of the event to represent the whole – just as an icon of a rabbit's ear may signify the whole rabbit. Synchronous haptic and audio metonyms strengthen the reference to the event, in a complementary manner.

The narrative functions of haptics were developed further in a story driven by the haptic handling of a virtual object, presented in a stereographic, haptic and audio interface [Barrass and Adcock 2003], shown in Figure 1. In this story, titled 'A Trifling Token', a model of a silver snuffbox takes on the role of a homodiegetic (within the story) narrator. The sounds of tapping and scratching on the snuffbox have the same function as metonymic sounds in computer games, but in this case continuous sounds are produced using modal and granular synthesis algorithms [Castle et. al. 2002]. The lid of the snuffbox has an engraving of a castle on it, which can be felt and heard by tapping and scratching with the stylus. Opening the lid causes a model of a ship to sail across a map from Britain to Australia, and the names of the passengers to emerge from inside. These floating nametags can be picked up for closer inspection of details about that passenger. Under the lid is another engraving "A Trifling Token of Esteem from the Emigrants on the Lady Ann to Captain W. Maxton, 1854".

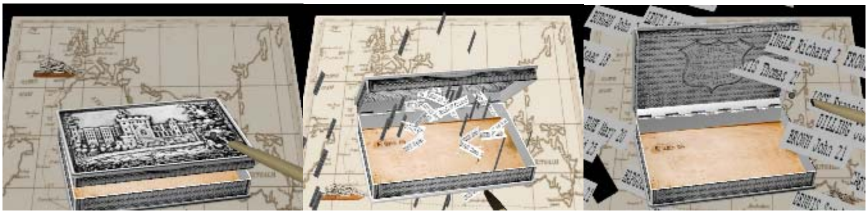


Fig. 1. ‘A Trifling Token’ – the haptic snuffbox exhibit

3 The Cocktail Party Effect

Trifling Token extended the narrative function of haptics to the role of a homodiegetic narrator. This section describes the further development of this approach but this time using haptic-audio without the conventional support of graphics or text. The Cocktail Party Effect was commissioned by Experimenta as part of the New Visions Commissions program in 2004 for the Experimenta Vanishing Point media arts exhibition in 2005. This section describes the exhibition brief, the initial proposal of a haptic-audio installation, the technical apparatus, a series of prototypes, and the final installation.

3.1 Exhibition Brief

The theme for the 2004 Experimenta New Visions Commissions, and Experimenta's 2005 exhibition program is Illusions. The concept of illusion, in all its forms, has long been a source of fascination. Throughout time, the creative exploration of illusions has not only inspired possibilities for the future, but provided a much needed alternative perspective to mainstream medias. What illusions exist in our world today? How do they affect our lives? What do they reveal about our current existence? What illusionary experiences can be created through media art works? Applicants for the New Visions Commissions are encouraged to explore the potential of illusion in all its forms - optical, sensory, aural - in order to create magic, wonder and mystery, and to uncover issues of political and social relevance. Selected artworks will animate untold stories, offer an alternative to our everyday lives, and allow us to uncover the invisible, psychological and emotional reality that exists beyond the surface.”

Note: the theme ‘Illusions’ was updated to ‘Experimenta Vanishing Point’ later on.

3.2 Proposal

The proposal for this brief was to provide the illusion of interaction with an invisible object through the senses of touch and sound. The illusion will consist of invisible cocktail glasses that have recorded conversations at a party. The immaterial glasses can be felt and heard when tapped with a spoon but are otherwise invisible. The spoon will vibrate in response to the laughter, and react to other aspects of the conversation. Each glass has a different memory of the party that can be heard by stirring. The apparatus consists of a SensAble Omni force feedback device on a table with stools around it. The Omni, with a spoon in place of the stylus, is positioned in the centre of the table, and cabled through a hole in the table top to a Pentium 4 Dual Processor PC running Windows XP hidden underneath. The PC has a Creative Labs Audigy soundcard connected to a speaker hidden under the table-top. The initial mock-up of the piece, shown in Figure 2, used wire-frames shaped as outlines of three cocktail glasses. Experiments showed the stylus bumped against the wire-frames and pushed them out of shape. It got trapped in awkward positions between the glasses, and the spatial range was too small for 3 full scale cocktail glasses.

Fig. 2. Mockup of the Cocktail Party Effect

The first interactive prototype was simplified to a single martini glass modeled by an inverted cone attached to a cylindrical stem in OpenGL rendered using the OpenHaptics HLAPI [OpenHaptics 2005]. Tapping sounds were synthesised using the glass harmonica from the STK Synthesis Toolkit [Cook and Scavone, 1999]. Experiments with the prototype showed up disconcerting haptic discontinuities at geometric intersections, so the model meshes were replaced with mathematically smooth implicit surfaces [Bloomenthal 1988], programmed with the lower level OpenHaptics HDAPI. The wire-frame visual references made it apparent that the outer haptic surface had a smaller radius than the inner, due to the elasticity of the stylus proxy spring. The stylus exists in the virtual haptic space as a single point so the handle does not cause contacts that are expected when it passes through the (invisible) sides of the glass shown by the wire-frames. The wire-frames were replaced with a mist-making machine that poured a stream of fog onto the table ass a cue for the location of a glass. However the mist dissipated in the wake of movement of the spoon, and introduced confusion in the signification of a glass. In another experiment a photographic slide of the shadow cast by a margarita glass was projected onto the tabletop, to show both the shape and the region of interaction. However it was difficult to calibrate the haptic contacts between the spoon and glass with visual contacts between the shadows of the spoon and the glass. The interplay between the position of the user and their shadow cast by the projector also proved problematic.

This led to the question of whether haptics and audio could be sufficient without any visual icons? Five subjects were asked to identify the unknown object using only the haptic-audio interface. They exhibited behaviours such as sideways sweeping, circling, and poking that produced occasional isolated contacts and tapping sounds. None identified the shape as a cocktail glass, and the general impression was of a wind-chime. The problem was that the surface of the invisible glass occupied a relatively small proportion of the volume of interaction, and the thin stem was difficult to find or follow. These observations lead to the design of a larger, more shapely, margarita glass with a top bowl, a smaller lower bowl, and three overlapping ellipsoids making up the stem. The ratio of haptic surface to interaction volume was increased by constraining the spoon to the inside of the glass. The literal context for interpreting a cocktail glass was developed though the replacement of the spoon with a swizzle stick, and placement of a coaster at the location of the glass. The exhibition designer proposed that the installation would be decorated like a cocktail bar. The context of a cocktail party was also extended to the haptic and audio components of the installation. The simulated glass was half-filled with a simulated liquid that resisted the movement of the swizzle stick. An audio recording of a party was played when the swizzle was inside the liquid. The introduction of this audio recording transformed the simulated glass into a heterodiegetic (external) narrator, telling the story of the party it contained.

The narrative prototype was evaluated by five practicing authors in the narrative and interactive arts. The Omni was placed on a table with the swizzle resting on a coaster, and each was asked to talk aloud as they explored it. They were distracted by the look of the Omni, e.g. ‘it looks like R2D2’ (a robot character in Star Wars), or ‘is it some kind of kitchen appliance?’ Although it should have been familiar the participants held the swizzle—stick at different locations, and grasped it like a pen, a shovel, or a fork. One rotated the swizzle-stick on the joint with the robot arm without actually displacing it, so no haptic or audio effects were produced at all. One participant accidentally removed the swizzle stick from the arm, but then gripped the arm directly and continued to interact. In this configuration the affordance of the tool was much more constrained, and the point of haptic contact at the end of the arm was much clearer. Nevertheless, drawings of the object by the participants in Figure 3 show a general agreement about the shape of the glass.

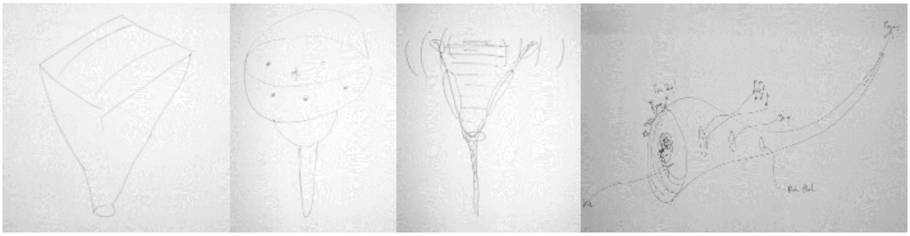


Fig. 3. Drawings from the evaluation

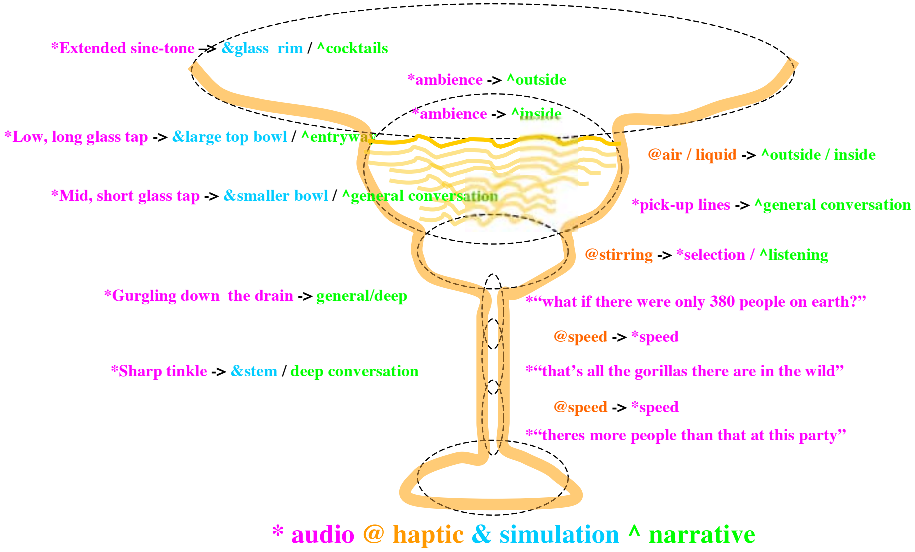


Fig. 4. Narrative structure organised by spatial layout and haptic-audio effects

Several participants in this evaluation suggested that the conversations at the party could be developed into a story that might increase audience engagement with the piece. One way to generate narratives from recorded documentary material is to juxtapose elements in new relations in a technique known as cut-up or montage. The recording of the party contained many conversations that could provide material for an audio montage, such as a light-hearted conversation about 'pick-up' lines, and a deeper discussion about a seminar on the Great Apes Survival Project [GRASP 2005]. Sentences from these conversations were re-recorded in a quiet room with a female voice actor. The components of these conversations were arranged in the margarita glass, as shown in Figure 4. The empty space at the top is the entrance with a grammatical soundscape of beats and muffled conversations. The clarity and loudness increases dramatically on the boundary with the simulated liquid half way up the top bowl, signifying entry into the party space. Moving down into the lower bowl enters into the cut-up general conversation of 12 pick-up lines played over the party soundscape depending on the location of the haptic contact. Moving down into the stem enters into the deep personal conversation with three connected statements at three depths - 'what if there were only 380 people left on earth?', 'that's how many Gorillas there are in the wild', and finally 'there are more people than that at this party'. The narrative regions were further delineated by changes in the tapping sounds at the boundaries of each region. Tapping the top bowl produces a longer, lower pitched tone to indicate larger size. Tapping the smaller bowl produces a shorter higher sound, while tapping in the stem produces a short high tinkle. Special haptic-audio effects were introduced to stimulate curiosity and imagination. Rubbing the rim of the glass produced a sine-tone simulation of the tone produced when the rim of a real crystal glass is rubbed. The rate of stirring in stem caused the speed of the voice to speed up. The recording of water gurgling down a drain was placed at the intersection the lower bowl and the stem that also separates the general and deep conversations.

This narrative object can be explored in an arbitrary manner, but should convey the mental image of the curvy margarita glass that is the narrator, the story of a cocktail party, and the understanding that Gorillas have reached the vanishing point. This version of the Cocktail Party Effect without any visual elements received an honourable mention in the SensAable 3D-Touch Challenge at Siggraph in 2005 [SensAable 2005]. It was one of only two interfaces that had sounds in them, and the only one that did not have a visual interface.

4 Discussion

The final version of the Cocktail Party Effect, shown in Figure 5, was one of twenty interactive installations at Experimenta Vanishing Point at the Black Box Gallery in Melbourne in September 2005. In this version the Omni was tilted at a 45 degree angle and positioned behind a façade so that the arm could be seen but the body was hidden. The rotation joint was removed leaving the end of the arm bare, leaving only one way to hold it. Mirrors on the wall signified a cocktail bar.

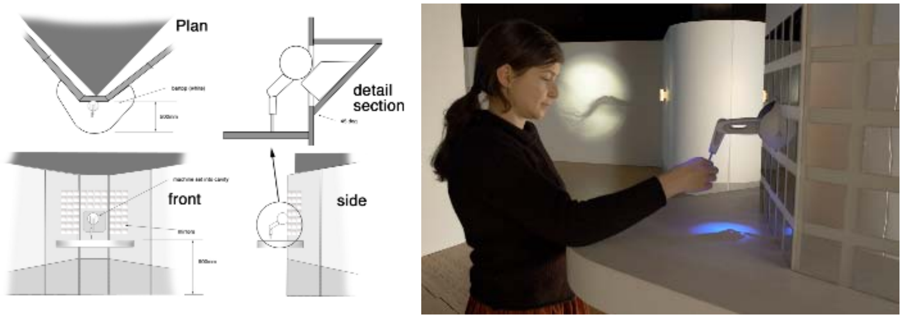


Fig. 5. Experimenta Vanishing Point, installation view. Courtesy Experimenta. The Cocktail Party Effect is an Experimenta New Visions Commission.

An artists statement next to the installation provided the following directions ...

“Use the swizzle stick to feel an invisible cocktail glass and overhear conversations at a party. If you to walk away with a mental image of the invisible glass, you will have reached the ‘Vanishing Point’.”

The visitor responses to the Cocktail Party Effect were observed on the opening night, and over the four week course of the exhibition. The noise and queues on the opening masked the sounds and did not allow for extended explorations. Visitors were cautious and unsure of the forces that could be expected or exerted, saying “Its jerky”, “I think its broken”, “What is it?” A short demonstration almost always produced more positive remarks such as “Oh, I feel it”, “There’s something there”, “You should try this!” On the following days visitors spent up to fifteen minutes exploring the installation, indicating narrative engagement. However many had trouble understanding the spoken phrases due to muffling of the audio by placement of the speaker in the tabletop. Most of the interaction consisted of tapping the glass near the top to hear the ringing tone, or sliding in and out of the stem where there was the most significant change in shape, and where the individual phrases about gorillas were most clearly discernable. Most articulated response was to the simulation of the glass, while the understanding that Gorillas are nearly extinct went largely uncommented on. The curators and gallery attendants commented that a simple simulation may have been more effective for audiences in this exhibition. The consistent voice of the female actor may have produced confusion between the glass as a heterodiegetic narrator outside the story, and the voice as a homodiegetic narrator wiithin the story. In the third week a cable inside the Omni was broken, and the installation ceased to function.

5 Summary

The Cocktail Party Effect developed the narrative functions of haptic-audio for interactive storytelling. It also explores haptic-audio simulation as a means to convey the mental image of an invisible object. In this piece a simulated cocktail glass takes the role of a heterodiegetic narrator telling the story of the imminent extinction of Gorillas in the wild. The piece uses curiosity-driven haptic exploration of a virtual cocktail glass to drive a narrative made from cut-up conversations, grammatical soundscapes, and dramatic haptic-audio special effects. These new narrative functions for haptic-audio were developed in three cycles of iterative prototypes and evaluations, culminating in a month long public exhibition. Observations from the exhibition indicate encouraging engagement and the potential of haptic-audio for interactive storytelling.

6 Future Work

A/Prof. Jen Webb suggested Lyric Poetry that portrays personal feelings, states of mind, and perceptions, rather than characters and events, could be used to create poetic haptic-audio installations. The next work in this series is a lyrically poetic haptic-audio teapot which will be ‘invisibly’ exhibited among a collection of computer graphic teapots at Siggraph 2006 in Boston http://www.siggraph.org/s2006/main.php?f=conference&p=teapot

Acknowledgements

Chris Gunn.
Matt Adcock.
Experimenta Media Arts.
Lance Davis, Exhibition Design.
Artistic Works-in-Progress seminars at the University of Canberra.
Students of Narrative and Interactive Practice at the University of Canberra.

References

[Barrass and Adcock 2002] Barrass S and Adcock A (2002) Interactive Granular Synthesis of Haptic Contact Sounds, AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, Helsinki University of Technology, Espoo, Finland, 15th-17th June 2002.

[Barrass and Adcock 2003] Barrass S. and Adcock M. (2003) Haptic Snuffbox, GRAPHITE 2003 Art Show, SPAN Galleries, Melbourne, 7-14 February, 2003.

[Bloomenthal 1988] Jules Bloomenthal, (1988) Natural Shapes and Implicit Models, Proceedings of the Third International Conference on Engineering Graphics and Descriptive Geometry; Vienna, Austria; 11-16 July 1988, S. Slaby and H. Stachel, eds.

[Castle et. al. 2002] Castle G, Adcock M and Barrass S (2002) Modal and Granular Synthesis of Haptic Tapping and Scraping Sounds, Eurohaptics International Conference, University of Edinburgh, July 8-10, 2002.

[Cook and Scavone, 1999] Cook, P. R. and Scavone, G. (1999) "The Synthesis ToolKit (STK)", Proceedings of the International Computer Music Conference, Beijing, 1999.

[DiFranco et. al. 1997] D.E. DiFranco, G.L.Beauregard, and M.A. Srinivasan, (1997) “The Effect of Auditory Cues on the Haptic Perception of Stiffness in Virtual Environments”, in Proc. 1997 of the ASME Dynamic Systems and Control Division, Ed. G. Rizzoni, ASME, 1997, pp 17-22.

[DiFilippo and Pai, 2000] D. DiFilippo and D. K. Pai, (2000) “The AHI: An Audio and Haptic Interface for Contact Interactions”, in Proc. 2000 Annual ACM Symposium on User Interface Software and Technology, San Diego, USA, 2000, November 5-8.

[Experimenta 2005] Experimenta Vanishing Point (2005) Black Box Gallery, The Arts Centre, Melbourne, 1-30 September, 2005.

[GRASP 2005] Great Apes Survival Project, http://www.unep.org/grasp/

[Gunn et. al. 2005] Gunn, C., Hutchins, M., Stevenson, D., Adcock, M., Youngblood, P., Using Collaborative Haptics in Remote Surgical Training, Conference Proceedings, WorldHaptics, Pisa, Italy, March 2005.

[McLaughlin et. al. 2000] McLaughlin, M.L., Sukhatme, G., Shahabi, C., Medioni, G., and Jaskowiak, J., (2000) “The Haptic Museum,” Proceedings of the EVA 2000 Conference on Electronic Imaging and the Visual Arts, Florence, Italy (2000).

[Omni 2005] SensAble Technologies, Omni Haptics Device http://www.sensable.com/products/phantom_ghost/phantom-omni.asp

[OpenHaptics 2005] SensAble Technologies, OpenHaptics, http://sensable.com/products/phantom_ghost/HapticLibraryAPI.asp

[SensAble 2005] SensAble Technologies (2005), 3D Developers Touch Challenge, Siggraph 2005, http://www.sensable.com/newsevents/pressreleases/pr-080205A.asp

Author Index