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

Volume I

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

Information Science
REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey · New York

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

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

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

Library of Congress Cataloging-in-Publication Data

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

p. cm.

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

Includes bibliographical references and index.

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

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

QA76.59.H36 2008

004.165--dc22

2007024493

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

If a library purchased a print copy of this publication, please go to http://www.igi-global.com/reference/assets/IGR-eAccess-agreement.pdf for information on activating the library's complimentary electronic access to this publication.

Chapter XXIX

Model-Based Target Sonification in Small Screen Devices:

Perception and Action

Parisa Eslambolchilar
University of Wales, UK

Andrew Crossan
University of Glasgow, UK

Roderick Murray-Smith
University of Glasgow, UK
Hamilton Institute, NUI Maynooth, Ireland

Sara Dalzel-Job
University of Glasgow, UK

Frank Pollick
University of Glasgow, UK

ABSTRACT

In this work, we investigate the use of audio and haptic feedback to augment the display of a mobile device controlled by tilt input. The questions we answer in this work are: How do people begin searching in unfamiliar spaces? What patterns do users follow and which techniques are employed to accomplish the experimental task? What effect does a prediction of the future state in the audio space, based on a model of the human operator, have on subjects' behaviour? In the pilot study we studied subjects' navigation in a state space with seven randomly placed audio sources, displayed via audio and vibrotactile modalities. In the main study, we compared only the efficiency of different forms of audio feedback. We ran these experiments on a Pocket PC instrumented with an accelerometer and a headset. The accuracy *of selecting, exploration density, and orientation of each target was measured. The results quantified the changes brought by predictive or "quickened" sonified displays in mobile, gestural interaction. Also, they highlighted subjects' search patterns and the effect of a combination of independent variables and each individual variable in the navigation patterns.*

INTRODUCTION

One of the main goals of interaction design is to make the interfaces as intuitive as possible. In our everyday environments, humans receive a variety of stimuli playing upon all senses, including aural, tactile, and visual, and we respond to these stimuli. Even though hearing and vision are our two primary senses, most of today's interfaces are mainly visual.

Visual interfaces have crucial limitations in small-screen devices. These devices have a limited amount of screen space on which to display information. Designing interfaces for mobile computers/phones is problematic, as there is a very limited amount of screen resource on which to display information, and users' need to focus on the environment rather than the interface (so that they can look where they are going) so output is limited (Blattner, Papp, & Glinert, 1992; Brewster, 1997; Brewster & Murray, 1998; Johnson, Brewster, Leplatre, & Crease, 1998; Kramer, Walker, Bonebright, Cook, Flowers, Miner, 1999; Rinott, 2004; Smith & Walker, 2005; Walker & Lindsay, 2006); also, low graphics resolution and further constrain the freedom of interface designers. In new generations of mobile phones (e.g., iPhone) with high graphics resolution, power consumption for graphics rendering is high, which can adversely affect battery life; also, large screens can lead to physical robustness issues, as well as being very demanding of user attention in mobile scenarios.

One way around these problems would be sonically enhanced interfaces that require less or no visual attention and therefore, the size of the visual display and the portable device can be decreased; also, auditory interfaces potentially interfere less in the main activity in which the user is engaged. Consequently, the user may be able to perform more than one task at a time, such as driving a car while using a telephone or grabbing a cup of coffee while waiting for a mobile phone to finish downloading an image. Auditory feedback can often be a necessary complement, but also a useful alternative to visual feedback. When designing a mobile electronic device, it is difficult to predict all possible scenarios when it might be used. Obviously, visual feedback is preferred in many situations such as in noisy environments or when the user has to concentrate on a listening task. However, as there might be numerous occasions when a user cannot look at a display, versatile devices such as mobile phones or handheld computers benefit from having flexible interfaces.

Novel Interaction and Continuous Control

In the past 10 years many researchers have focused on tilt-based inputs, and audio and haptic outputs in mobile HCIs (Dong, Watters, & Duffy, 2005; Fallman, 2002a, 2002b; Harrison & Fishkin, 1998; Hinckley, Pierce, Horvitz, & Sinclair, 2005; Oakley, Ängeslevä, Hughes, & O'Modhrain, 2004; Partridge, Chatterjee, Sazawal, Borriello, & Want, 2002; Rekimoto, 1996; Sazawal, Want, & Borriello, 2002; Wigdor & Balakrishnan, 2003). The results of these researches have proved one-handed control of a small screen device needs less visual attention than two-handed control and multimodality in the interaction can compensate for the lack of screen space. So these novel interaction techniques, that is, gesture recognition, and audio and haptic devices, are characterised by the significance of the temporal aspect of interaction and in such an emerging environment, the interaction is no longer based on a series of discrete steps, but on a continuous input/output exchange of information that occurs over a period of time at a relatively high rate, somewhat akin to vision based or audio/haptic interfaces, which we may not model appropriately as a series of discrete events (Doherty & Massink, 1999; Faconti & Massink, 2001).

Novel interaction techniques with computers and handheld devices are examples of interactive dynamic systems, and development of these systems explores a range of possible solutions for overcoming some problems of development on computing devices, including the limited source of input/output devices, adaptability, predictability, disturbances, and individual differences. We explicitly include dynamics because we experience our environment in the way we want it by our actions or behaviour. Thus, we control what we perceive and while, in principle, interaction with handheld devices is rich in the variety of tasks supported, from computation and information storage to sensing and communication, we are dependent on the display of feedback (either visual, audio, or haptic) to help us pursue our sometime constantly changing goals feedback, which may influence a user's actions as more information becomes available (Doherty & Massink, 1999; Faconti & Massink, 2001). So developing interaction for such devices is closely related to the engineering of mobile interfaces based on dynamics.

Control Theory and Fitts' Law

A branch of control theory that is used to analyse human and system behaviour when operating in a tightly coupled loop is called *manual control theory* (Jagacinski & Flach, 2003; Poulton, 1974). The theory is applicable to a wide range of tasks involving vigilance, tracking and stability, and so forth. The general approach followed in manual control theory is to express the dynamics of combined human and controlled element behaviour as a set of linear differential equations in the time domain (Poulton, 1974). Several models include human-related aspects of information processing explicitly in the model, such as delays for visual process, motor-nerve latency, and neuromotor dynamics (Jagacinski & Flach, 2003). Control theory can be linked to Fitts' Law (Fitts, 1954; MacKenzie & Ware, 1993; Mackinlay, Robertson, & Card, 1991) by viewing the pointing movements towards the target as a feedback control loop based on visual input, and the limb as a control element (Bootsma, Fernandez, & Mottet, 2004; Crossman & Goodeve, 1983; Hoffmann, 1991; Jagacinski & Flach, 2003; Langolf, Chaffin, & Foulke, 1976).

This research outlines the use of model-based sonification to shape human action when users interact with small devices based on auditory feedback. In this work, we investigate the usability of nonspeech sounds and haptic feedback to augment the display of a mobile device controlled by a gesture input. Nonspeech sound has advantages over speech in that it is faster as well as language independent. We look at control strategies of users in browsing the audio/haptic state space. We also suggest one possible way of improving performance based on models of human control behaviour in a few example applications.

BACKGROUND

The single audio output channel has been little used to improve interaction in mobile devices. Speech sounds are, of course, used in mobile phones when calls are being made, but are not used by the telephone to aid the interaction with the device (Blattner et al., 1992; Brewster, 2002; Gaver, Smith, & O'Shea, 1991; Smith & Walker, 2005; Walker & Lindsay, 2006). Nonspeech sounds and vibrotactile devices are used for ringing tones or alarms but again, do not help the user interact with the system beyond this. Some signals provide feedback that some event has been successful, such as when buttons are pressed or devices are switched on. Selecting items with a stylus in PDAs without tactile feedback is often confusing for users because it is hard to know whether they have hit the target or not, especially if used in a mobile setting (Brewster, 2002). In this case, vibrators in mobile phones could be a good haptic feedback. It assures the user that s/he is in the target, and if the user wants to select a target, s/he can then press a key in the vibration area to select it.

If using continuous sounds as opposed to the more common brief signals, auditory interfaces do not need to be more transitory than visual interfaces. However, such sounds probably benefit from being quite discreet. While, most existing sound feedback today occurs in the foreground of the interface, subtle background sounds can be a useful complement in advanced auditory interfaces. Films and computer games generally make use of music and sound effects. Film sound theorist Michel Chion (Chion, 1994) has made the following statement concerning sound in film: *there is no soundtrack*. An extreme statement coming from a researcher of sound, Chion means that there is no way to separate the auditory and visual channels of a film. We experience them only through a unified sense, which he terms "audio-vision."

In a similar way, an interface that uses sound cleverly can enhance the user's immersion and improve interaction. Gaver (1997) found that during an experimental process control task, the participants' engagement increased when provided with relevant sound feedback. There is now evidence that sound can improve interaction and may be very powerful in small screen devices (Brewster, 2002). If the possibility of conveying information sonically were used to its full potential, it would be a powerful complement to visual interfaces (Brewster & Murray, 1998). A strong argument against the use of sound in interfaces is that it easily can become annoying, both for the user and other people around them, since it is more intrusive than visual impressions. It is not useful in noisy environments, for instance, train stations, undergrounds, and so forth. However, by skilfully designing auditory interfaces or using haptic feedback, this can be avoided, and interaction with machines can become easier and hopefully more pleasant.

The most advanced auditory/haptic feedback seems to exist in computer games and multimedia products. Gaver (1997) claims that memory limitations in the technical product are one reason why sound feedback has not been used on a larger scale. Until quite recently, it has been too expensive computationally to use sound of good quality in computers and handheld devices. Today, only lightweight electronic devices, such as mobile phones or handheld computers, have limited memory capacities, although this is rapidly changing with the development of memory cards and effective compression algorithms for sound. However, nowadays these devices give various choices of discrete audio/haptic ring tones and alarms and to their users. The potential to use sound and haptic in small electronics is growing fast.

MODEL-BASED SONIFICATION

As there are many ways in which sound can be employed in interfaces, it is important to define the purposes of every sound at an early stage in the design process. A sound that conveys crucial information should have different attributes to one that serves as a complement to visual information. It is important to distinguish between two very different approaches (Chion, 1994): the practical and the naturalistic approach. The "practical" approach to auditory interfaces deals with sound as the main feedback. This can be the case when designing interfaces for visually impaired people, who must rely on sound feedback to provide sufficient assistance in performing a task. Furthermore, sound is often the only means of communication when using a portable hands-free device with a mobile phone. Auditory interfaces based on a practical approach should be comprehensive and simple (Brewster, 2002; Brewster & Murray, 1998; Smith & Walker, 2005; Walker & Lindsay, 2006). The drawback of this approach is sound might be noisy and tiresome over time. The "naturalistic" view regards sound mainly as a complement to a visual interface. A naturalistic interface combines sound and vision in a way as similar as possible to corresponding phenomena in the natural world. Such auditory interfaces are supposed to enhance interaction between the user and a machine, especially in situations where the visual interface is ineffective on its own. Sounds that complement a visual interface can generally be subtle background events that do not disturb. In a way, such sounds correspond to the background music of films, since they convey information to the audience without interfering with the main events. Sound feedback based on the naturalistic strategy is thus very subtle, and might only be recognised subconsciously. The focus of this work is on the "practical" approach.

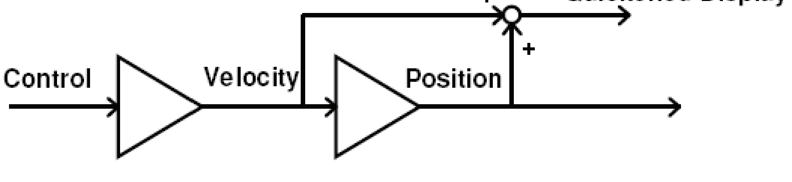
Sonification is a method suggested in "practical" domain, which is defined as the use of nonspeech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation (Gaver, 1989). Many of the major current research areas in sonification are similar in that they focus on the identification of applications for which audition provides advantages over other modalities, especially for situations where temporal features are important or the visual modality is overtaxed. The main issues that will move sonification research forward include (1) mapping data onto appropriate sound features like volume, pitch, timbre, (2) understanding dynamic sound perception, (3) investigating auditory streaming, (4) defining and categorising salience in general auditory contexts, and understanding where highly salient sonic events or patterns can surpass visual representations in data mining, and (5) developing multimodal applications of sonification (Kramer et al., 1999); sonification is a way to help in the exploration of complex data. Various kinds of information can be presented using sonification, simply by using different acoustic elements. This information has been organised in Hermann, Hansen, and Ritter (2000).

Studies such as Cook et al. (2002) and Cook and Lakatos (2003) have investigated the human ability to perceive various physical attributes of sound sources, and have proved that feature-based synthesis is of use in studying the low-level acoustical properties that human listeners use to deduce the more complex physical attributes of a sound's source. The generated sounds from a set of features are correlated with the listener's perception of, for example, size, speed, or shape of the source. Two methods of sonification have been used in this chapter, the Doppler effect and derivative volume adaptation. Both of these methods create a continuous sound for each data point. Thus, the relative position to the targets is perceived by a change of volume when passing the data point and pitch shift for Doppler effect as well. From the data points obtained in this way, we may be able to discover consistent relationships between acoustical and human-generated features that can be used to predict how a sound manifesting certain acoustic feature values will be perceived.

Quickening

"Quickening" is a method for reducing the difficulty of controlling second-order or higher-order systems, by changing the display to include predictions of future states, that was proposed by Birmingham and Taylor (1954), and is reviewed in Jagacinski and Flach (2003). A quickened display for an acceleration control system like the system described in this chapter shows the user a weighted combination of position and velocity (see Figure 1). This weighted summation effectively anticipates the future position of the system. It can greatly improve human performance in controlling these

Figure 1. A block diagram for a second-order system with a quickened display. The output to the quickened display is the sum of position and velocity. Effectively, the quickened display projects the output into the future based on the current velocity.



systems. Quickening in general is a prediction of the future state of the system based on the current state vector (for example position, velocity, acceleration) and a model of system behaviour and expected user action.

An example of this is based on the Doppler effect, which highlights the user's approach to a target, or a target's movement from the current state. Another example could be derivative of volume of sound source. When the user is further from the audio source, the sound is quieter than when the user is close to it. Another predictive method that has been investigated in Williamson, Strachan, and Murray-Smith (2006) is *Monte Carlo* simulation in a tilt-controlled navigation system.

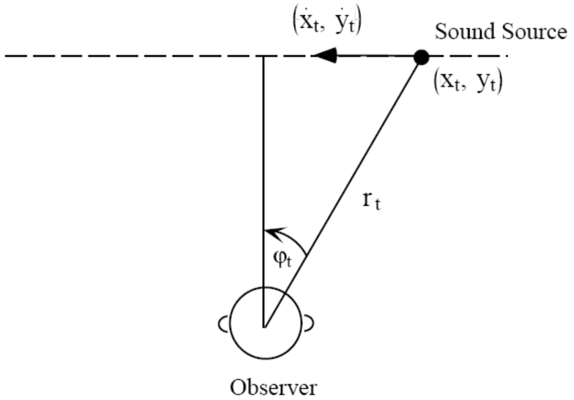
Doppler Effect

The auditory system is responsible for constructing a map of the auditory scene around us, using information from audio input, that is, sound localisation (Bregman, 1990; Smith, 2004). There are various types of cues that humans can use to localise the position of a sound source. These cues can be divided into monaural and binaural cues. The two different types of monaural cue are loudness and Doppler shift. The loudness cue relies on the fact that when a sound source is far away, it is quieter than when it is close by. The Doppler shift corresponds to a frequency shift associated with a sound source moving through a homogeneous medium (Smith, 2004). Pressure wave crests emerge from the sound source at intervals corresponding to the acoustic wavelength. Each crest spreads spherically out from the point of origin at the speed of sound *c* (Figure 2). The successively generated spheres of wave crests are closer together ahead of the sound source but farther apart behind the source. For a stationary observer, the measured frequency corresponds to the number of crests per unit time, so the composite frequencies will be higher when the observer is in front of the moving sound source, and less when behind the moving sound source (Hermann et al., 2000; Hermann & Ritter, 1999). A familiar example is the shift in frequency of an ambulance siren as the vehicle approaches, passes, and then recedes. The well-known lawful dependence of the Doppler shifted frequency, here denoted Ψ_t, on velocity of the sound source relative to an observer is:

$$\Psi_t = f(1+\frac{v}{c}\cos\Phi_t) \tag{1}$$

where *f* is the intrinsic frequency of the sound source, *v* is the velocity magnitude (speed), and *c* is the speed of sound. The shifted frequency Ψ_t depends only on the velocity component directed toward the observer with angle Φ_t (see Figure 2). The shifted frequency has the maximum value when Φ_t is zero. As this angle reaches 90°, all motion is across the line of hearing and the Doppler shift is zero. This result holds true regardless of

Figure 2. The geometry for the Doppler shift of a moving sound source relative to an observer.



the time history of the trajectory (Jenison, 1997). These aural cues can be used to navigate through the virtual environment on a Pocket PC.

In the next sections we present pros and cons of different quickened methods and control strategies in browsing the state-space on a mobile device using tilt-input.

EXPERIMENT

Goals

There is a concept of accuracy explored in this work. The type of accuracy that is under primary consideration in this study is the capability of subjects to accurately identify audio sources in large audio data sets with a PDA and tilt-sensor using sound only. In navigating a computer display of data visually, accuracy is seldom a concern. Using a scrollbar or clicking a 10 × 10 pixel icon using one's vision is trivial from the perspective of the accuracy needed to accomplish this task (Holmes, 2005). Designers of auditory displays, on the other hand, are in need of research into the accuracy that is possible in this environment. Establishing the accuracy with which humans can navigate using sound alone is an early step in integrating sound into a multi-modal information system.

The other questions we answer in this work are: How do people begin searching in unfamiliar spaces? What patterns or techniques are employed to accomplish the experimental task? How will predicting the future state in the audio space change subjects' accuracy in targeting?

Apparatus

The experiment was conducted on a Pocket PC (hp5450), running windows CE, with a 240 × 320 resolution, colour display, an accelerometer Xsens P3C, 3 degree-of-freedom, attached to the serial port, which allows the users to navigate through the environment by tilting the device, and a stereo headset (Figure 3). The built-in vibrator unit in the Pocket PCs provides the haptic feedback in the experiment.

The experiment was written using the *FMOD* API (version 3.70CE)(FMOD, 2004), a visual programming environment with an object-oriented language (Embedded Visual C++) used primarily to manipulate and control sound production and *GapiDraw* (version 2.04) (GAPIDraw, 2004), a runtime add-in to *FMOD* used to generate real-time Pocket PC graphics. *FMOD,* and *GapiDraw* are available for free under the condition of the GNU public license (GPL).

Using *FMOD* and *GAPI*, an interface was developed with the following parameterisations: speed of sound, 340ms^{-1}, Doppler factor, 1.0, distance scale 100.0, minimum audible distance 80m, full volume (255)(minimum volume is 0

Figure 3. Left- Pocket PC, Accelerometer and experiment I running on the system (target sound sources displayed, for illustrative purposes). Right- A user interacting with the system.

and max volume is 255 in *FMOD*), and maximum audible distance 8000m. Each pixel on the display represents 100 metres. An empty window (240 × 320 pixels) was centred on the screen. Audio sources represented by small (10 × 10) speaker icons are shown on the screen only for training (Figure 3). In the main experiment, sound sources are hidden and an empty window is shown on the screen. Only the cursor, represented by a small (10 × 10) ear icon, is visible in both training and main experiment.

Experiment I

We first conducted a pilot study with 12 subjects, 3 women and 9 men, all sighted, with a mean age of 29 years. Four participants were research fellows, and the rest were postgraduate students at the NUIM campus. All but one of the participants had neither experience of using Pocket PCs nor with accelerometer-based interfaces. Two of them were left-handed (Eslambolchilar, Crossan, & Murray-Smith, 2004).

Task and Stimuli

The task in this study was to select the centre of individual targets that appear (in audio but not visually) in different locations on the screen as accurate as possible. The individual targets are audible when the cursor is in their locality, and they have full volume only in the centre of the target (imagine a Gaussian distribution of the volume centred on the target). For each target, a vibration feedback has been assigned and whenever the user is in very close distance to the target, 10 pixels, s/he feels the vibration continuously. Our aim in using the vibration in this task is the vibration assures the user that s/he is very close to the centre of the target.

First, participants were asked to sit on a chair in a quiet office and were equipped with a headset and a Pocket PC in their palm. Then they were informed about the functioning of the accelerometer, Doppler effect, and the procedures of the experiment, in order to reduce the chance of any terminological misunderstanding. Subjects were asked to move the cursor to audio targets by tilting the PDA, and to select them by pressing a key on a small keyboard of the PDA. They were told to emphasise accuracy over speed.

Design

There were four experimental conditions: (1) No Doppler effect, no vibration feedback (2) No Doppler effect, vibration feedback, (3) Doppler effect, no vibration feedback, and (4) Doppler effect, vibration feedback. The participants performed the conditions in a counterbalanced order. This resulted in 12 different orders of experiments for participants. In each experiment, seven audio sources were used (a selection of different music) summarised in Table 1.

Visualisation

Matlab was used for visualising the logged experimental data. We use a number of techniques for investigating the users' behaviour in these experiments.

Audio and exploration Density Plots

These plots show the audio density (in pixels) at different points in the 2-D space (Figure 4 (Left)). The contour indicates the density of the sum of the amplitude of the mixture components associated with the different audio tracks. The exploration

Table 1. Audio sources in the first experiment in all conditions

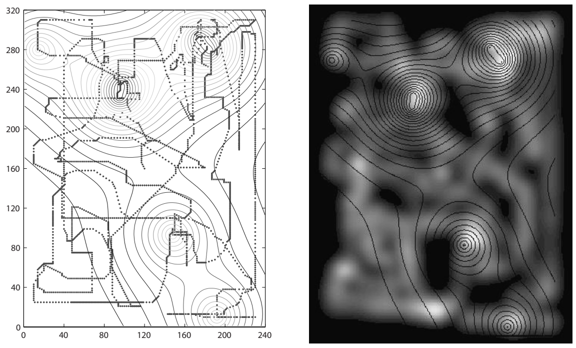
Target Index	Music Type
1	Hip-Hop
2	Celtic
3	Arabic
4	Country
5	Jazz
6	Farsi
7	Opera

density plot for visualisation of cursor trajectories, used previously in Williamson and Murray-Smith (2004), has been used here, which plots a density around the trajectory, which is a function of the position and the length of time spent in that position. These plots give some indication of how users navigated when completing the task. An example is given in Figure 4(Right). This plot is created by placing a Gaussian distribution centred on the (x,y) position of the cursor for each point in the log file, with standard deviation proportional to that used in the audio sources. The Gaussians are summed for each pixel, and the resulting image gives an impression of the areas of the input space that were explored, and how long the user spent in them. The image can be summarised numerically by counting the percentage of pixels greater than a selected threshold *e*. In this experiment *e*=5.0. The image's resolution is 240 by 320 pixels (Eslambolchilar et al., 2004).

Distance to the Target

Whenever the user feels s/he is at the target, s/he presses a key indicating the selection of the target. For each selection made by the user, the distance to the nearest target is calculated as below, and recorded.

Figure 4. Left - The cursor trace of the 4th participant in the "no Doppler-no vibration" condition, is plotted over the density of the local audio amplitude of the different tracks. Right - the density contour plot and cursor trajectory density indicating the exploration of the space by the same participant in the same condition.



$$Dist = \sqrt{(x_{source} - x_{selected})^2 + (y_{source} - y_{selected})^2} \qquad (2)$$

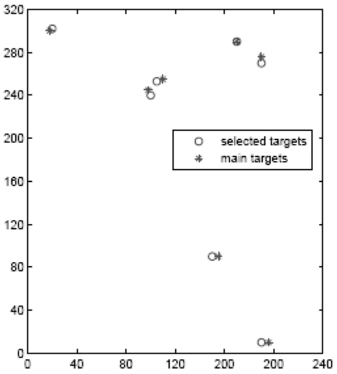
An example plot is shown in Figure 5. The distance to the location of the target (in pixels) gives some insight into the acuity with which the location can be perceived with the given display.

Results

Search Patterns Observed

In looking at the audio and exploration density plots, we are not attempting to establish a link between the search pattern used and the resulting measurement of accuracy. We simply make a subjective classification and qualitative assessment of the types of search patterns employed to accomplish the task. A subject may employ one of the search techniques and still not be very accurate, or they may be very accurate in spite of using no detectable systematic pattern. However, this factor gives an indication about the ease with which the audio environment could be clearly perceived by participants. In a clear and easy to navigate environment, with appropriate feedback, this should be similar to the density of targets, and linked to the smoothing used.

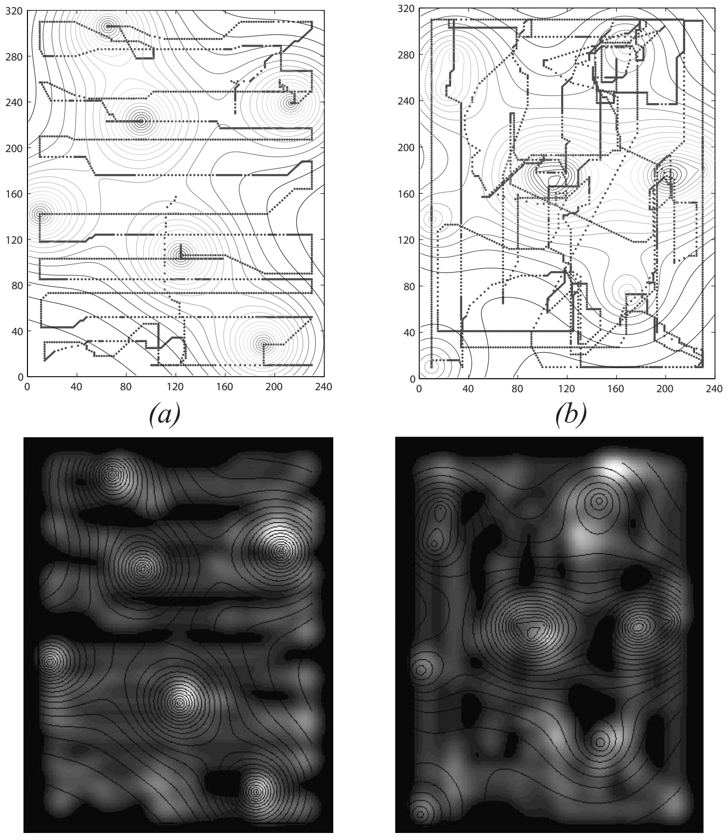
Figure 5. Hidden target positions (circles), and points selected by user 4 in the "no Doppler-no vibration" condition, as the best guess (crosses)



Some of the terms and their basic definitions used here are taken from search theory, a subfield within operations research (Patrol, 1999). The patterns developed by search theory are visual search patterns of physical space, but there is some crossover in the types of patterns used in the auditory interface used in the experiment to search in a virtual space.

1. **Parallel sweep:** The parallel sweep is used when uniform coverage of an area is desired and the area is unfamiliar. It is an efficient method of searching a large area in a minimum amount of time. Several subjects used the horizontal parallel sweep, "raster scan," similar to the one seen in Figures 6(a) or 7. This pattern can be related to the text-reading pattern we learn in the childhood.
2. **Quadrant search:** The quadrant search pattern is one in which the searcher mentally breaks down the screen into quadrants to divide the area into a more manageable size. Within the quadrants, the searcher may use another pattern to search each quadrant, such as a parallel sweep (Figure 8(a)).
3. **Sector search:** A sector search pattern begins once the approximate location of the target is located. In this pattern, the searcher explores out from the approximate location of the target and returns again, then conducts another exploration in another area, and returns again. This is repeated until they

Figure 6. (Left) The traces of the cursor for participant 12 in "no Doppler with vibration" experiment (a) and its exploration density plot (c), (Right) The traces of the cursor for participant 6 in "Doppler with vibration" experiment (b) and exploration density plot of this experiment (d)



(a) *(b)*

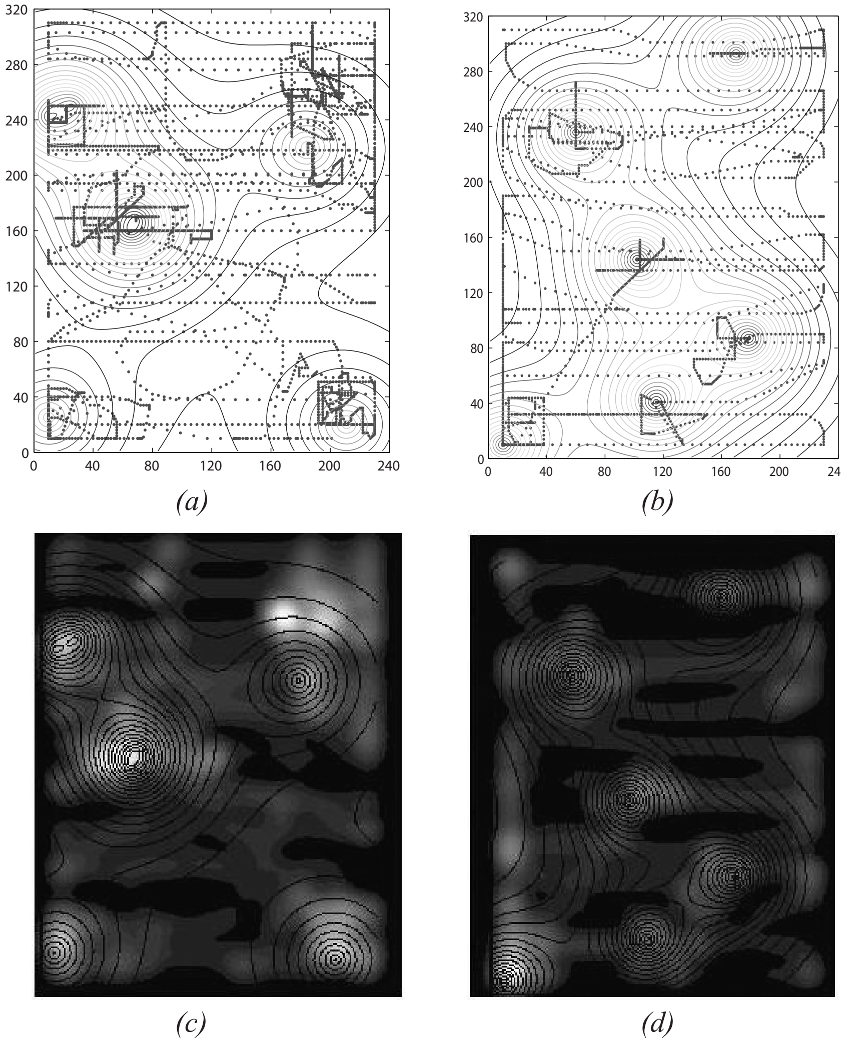
(c) *(d)*

are confident that the space is adequately explored (Figure 8(b)).

4. **Perimeter search:** The perimeter search is one in which the boundaries of the space are explored, but little or none of the middle is traversed. The pattern of search can be a circle to circumscribe the border or a square-shaped pattern turning at a 90° angle. This type of search pattern would typically lead to inaccuracy given that none of the targets are located at the perimeter. This search pattern was not observed in this research.
5. **No formulaic search:** For some searchers, no discernable systematic technique was employed in exploring the space to accomplish the task. For these search patterns, there is no attempt to thoroughly explore the information space. Figure 8(c) illustrates the path used in the only trial to actually select the target exactly.

The search patterns of each subject were analysed to see if there were any tendencies based on demographic characteristics; 48 total patterns were analysed. The most common technique employed

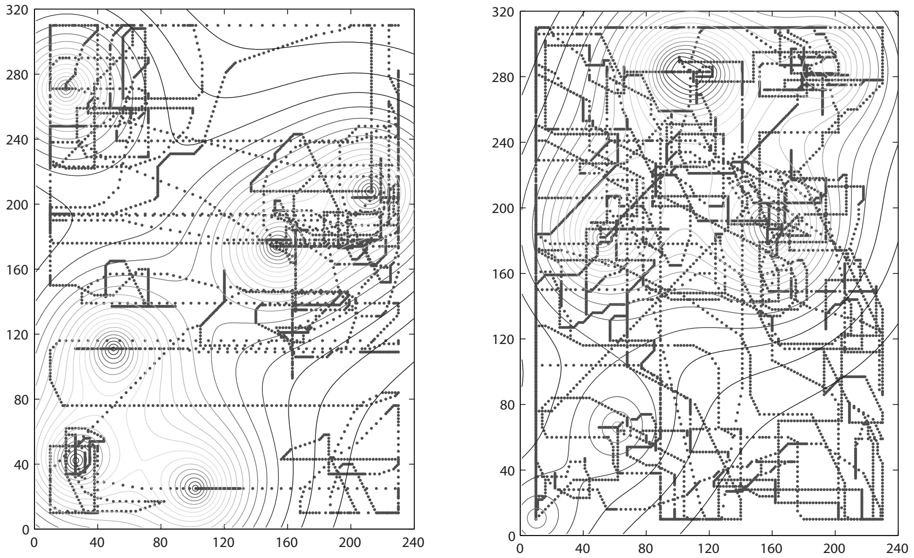
Figure 7. (Left) The traces of the cursor for participant 5 in "Doppler with vibration" experiment, (Right) The traces of the cursor for participant 9 in "no Doppler with vibration" experiment.



(a) *(b)*

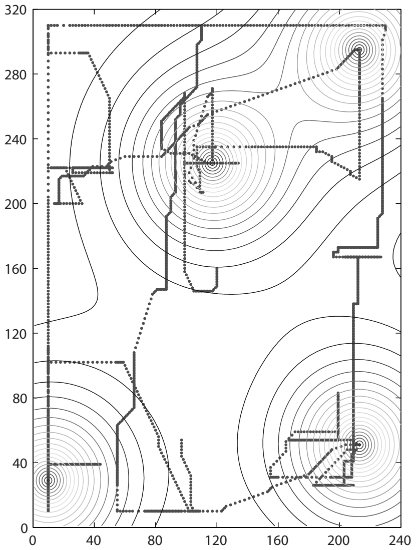
(c) *(d)*

Figure 8. Examples of few search patterns in different conditions



a) Example of quadrant search: Subject 9, "Doppler-no vibration"

b) Example of sector search: Subject 1, "no Doppler-no vibration"



c) Example of no formulaic search: Subject 7, "Doppler-no vibration"

was the sweep search (76%). The next most common was the no distinguishable pattern (14%), followed by quadrant (6%), and sector (4%). Participants' audio and exploration density plots show "Doppler-no vibration" has the least covered space with 34.5%, and the rest have similar percentage of coverage, 37.6%.

Chosen Songs

The accuracy relative to the number of song chosen is another factor in improving audio interfaces. Because the type of songs may affect the perception of distortion due to the Doppler effect, and affect the users' ability to recognise and locate them. We measured the number of audio sources

Figure 9. Mean distance in pixels from target in different tasks

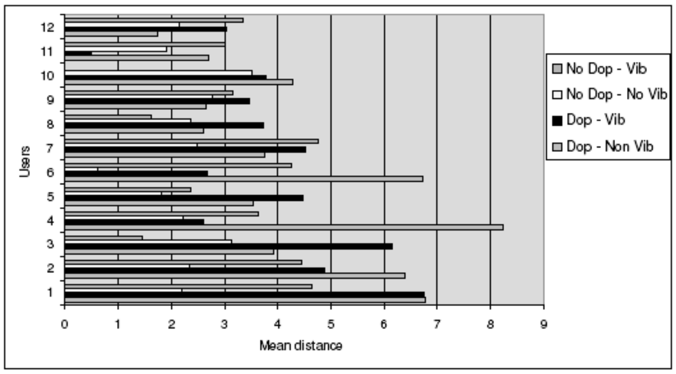


Figure 10. Count of most accurately chosen songs in different conditions for all users

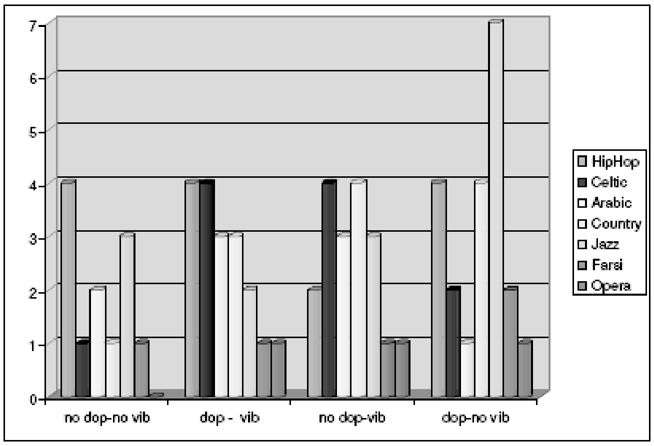


Figure 11. Mean distance (pixels) of selected songs in all conditions for all users.

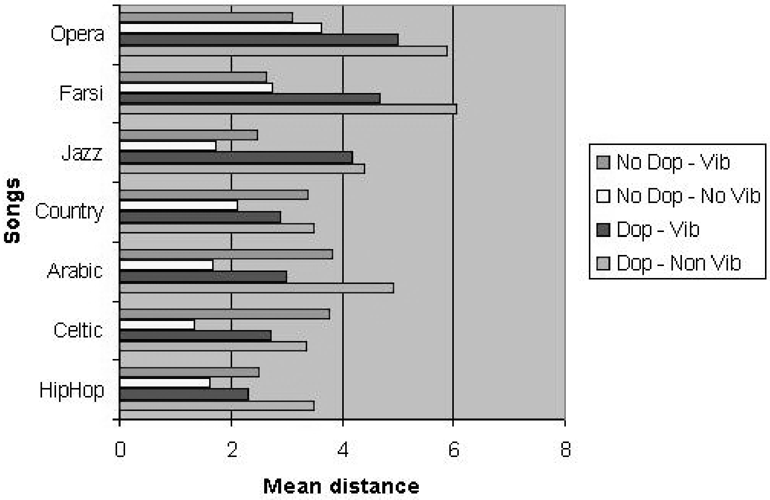


Table 2. Accuracy scores for audio sources in the first experiment

-	no dop- no vib	Dop-vib	no dop-vib	dop-no vib
Hip-Hop	4	4	2	4
Celtic	1	4	4	2
Arabic	2	3	3	1
Country	1	3	4	4
Jazz	3	2	2	7
Farsi	1	1	1	2
Opera	0	1	1	1

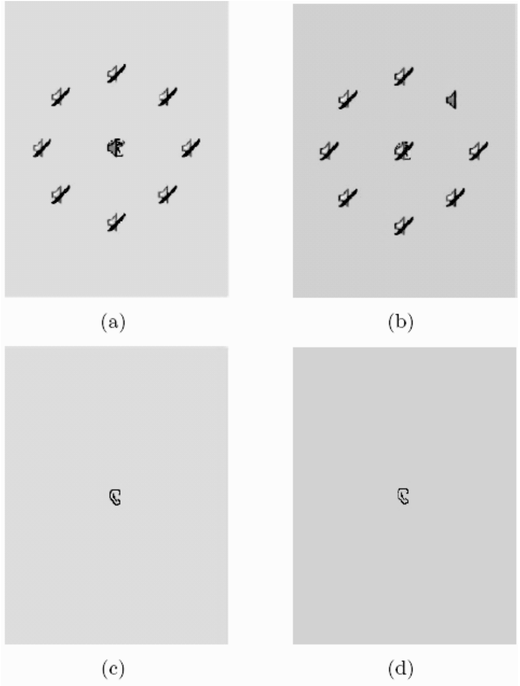
participants have selected. The mean accuracy for each of these sources has been summarised in Table 2. Figure 10 shows the mean accuracy count of songs in all conditions for all participants. This result is based on the number of times each source was selected with the smallest distance to the target in each condition. There is a large amount of variability in the results. Jazz music was selected more than others, on average, but Hip-hop music was chosen more accurately in the "no Doppler-no vibration" condition. Figure 11 shows mean error for songs in all conditions. In general "no Doppler-no vibration" has the lowest error among others and "Doppler-no vibration" has the highest error. Farsi and Arabic sources had high mean and maximum errors in the Doppler case.

Discussions

Post hoc examinations of the cursor's trace in this experiment showed that the subjects tended to use the same technique regardless of the sounds they heard and the audio condition. Six subjects (50%) used one search technique exclusively. Of these six subjects, five used the sweep technique, one used no distinguishable pattern exclusively. Another six subjects (50%) used the same technique in two out of the four conditions. This consistency in the application of a searching technique has several notable points. First, the same technique was employed regardless of the sound treatment. This would indicate that the subjects brought with them a technique that was not altered by the change in the treatments used in the auditory interface. The subjects were given no experimental feedback that might prompt them to change their search pattern to one that might be more effective. Left to their own means, the subjects tended to continue with the application of the search pattern with which they felt most comfortable. Second, the most common type of search pattern (sweep search) was also the least effective, given that the target in all four conditions was located towards the interior of the information space. In these cases, the subject was less likely to notice a change in the sounds they were hearing because of the low intensity of the sounds generated at the borders of the information space. Because they typically did not explore the interior, they would not hear the more intense sounds that might lead them to the target. In conditions with vibration feedback sweep search is combined with circular movements around the vibration source (Figures 6(a), 7(a) and 7(b)) and has led the users to the target. This suggests that the vibration was more important for the users in locating a target, and whenever they felt they were close to the song, they looked for the vibration source before clicking; so, feeling a vibration source meant they were at the centre of the audio source. This might also explain the fact that errors were not smaller, as the user may often have selected the location as soon as vibration was perceived, at the edge of the circle, rather than at the centre of the target itself.

The "no formulaic search" is the least thorough of the systematic techniques. Even though there was essentially no effort involved in exploring more thoroughly by applying a different search pattern, the subjects tended to use the no distinguishable search pattern. This could be accounted

Figure 12. The state space in experiment II and corresponding angles. Top pictures show the training application and bottom ones show the main application. Left pictures show the screen before pressing the button for Jazz music and right ones show the screen after pressing the button (colour changes and covered speakers are indicated in training application).



for by assumptions the subjects made about the nature of the information space. It would seem that some subjects took the experimental task seriously by systematically exploring the information space. Other subjects did not seem to be interested in exploration, but instead made a quick "stab" in the general direction of the high point. The case could be made that those subjects who explored liked the sonic interface, and those who did not explore did not like the interface. It may well be the case that auditory display is not for everyone. Some will like it and make use of it, others will not.

The results show that the mean distance from the selected position to the target in "no Doppler-no vibration" is less than other experiments (Figure 9). The extra clicks and navigation activities in the cursor trajectories for Doppler might be an effect of the extra sensitivity of the feedback to movement, which makes the users explore by varying their velocity vector. Variability in localisation accuracy is greater with the Doppler effect for the Farsi and Arabic sources, suggesting that for the mainly western European participants, their poorer familiarity with these sources made the distortions introduced by the Doppler effect more difficult to perceive. Opera also had larger errors, again suggesting that less familiarity with the target sources can affect the usefulness of this approach. The large number of falsely placed points for the Doppler method might be because of the amplification involved in moving towards something and potentially, frequency and speed of sound, which makes people feel they are getting a stronger response, and they over-interpret the quickened signal, believing they are already at the point - a common cited risk associated with quickened displays (Poulton, 1974).

Experiment II

Twenty-four paid participants (8 male, 16 female), all sighted with normal or corrected vision and normal hearing with mean age 24, were recruited through sign-up sheets in Glasgow University Psychology department and through e-mail. Two of the participants were left handed. Three participants had experience using a Pocket PC, and of these, one had frequent use. None had experience using an accelerometer as an input device.

Design

Given that the results of the pilot study showed we had possible confounding factors, that is, four different audio conditions, haptic feedback, few subjects, unfamiliar songs, and random located targets, the next experiment reduced the number of independent variables. This resulted in three different types of audio feedback without haptic feedback:

1. Doppler feedback
2. Derivative volume adaptation
3. No quickening

There were eight possible audio sources (targets) arranged in a circle (radius = 100 pixels) around the centre at 45° intervals (Figure 12). The audio feedback at the centre was jazz music, which played continuously for all conditions. When an outside target was to be located, the audio feedback was "Hotel California" played in a loop. The audio source to be located alternated between the centre (jazz music) and one of the outside targets ("Hotel California"), and always began with the centre target. The audio sources around the outside were presented in a random order, twice for each target for the training session (16 trials in all), and five times for each target for the experimental trial (40 trials altogether). Once one target had been located, a button was pressed. For each key pressing, there was a screen colour change and a short "beep" sound. Audio was noticeable within a radius of 90 pixels from sources (3s) in conditions 1 and 2, but audio was noticeable just within a radius of 15 pixels from the sources in condition 3, and there was no feedback at any other locations in this condition. Each participant was tested individually, and participants were told to commence the training session when they felt ready. After the training session, it was ensured that they understood the procedure fully and that they felt comfortable using the equipment. They were given a break between the training and experimental sessions if they wanted one.

Visualisation

In addition to exploration and audio density plots and distance to the target, we used another visualisation method that measures the orientation of each target with respect to the centre point, showing which angles in the state space have had the most accurate data in selecting targets. This measurement is important in this experiment to see whether the orientation of audio sources has any effect on the targeting task. Results in this experiment were analysed using a *GLM ANOVA* test.

Results

Proportion of Distance to the Target

Figure 13 shows the box plot of medians, means, and measures of spread of the distance between the audio sources and the position selected for each of the three audio feedbacks. The red triangle indicates the mean and the blue square indicates the median (50% of the observations lie below this line). The top of each box indicates the upper quartile (75% of the observations lie below this point) and the bottom indicates the lower quartile (25% of the observations lie below this point). The tops of the lines above the boxes indicate the highest observation, and the bottom of the lower line indicates the lowest observation. The blue stars indicate outliers: those observations that differ significantly from the mean. Figures 13 and 14 show a difference in the average distances between the actual audio source and the target selected (accuracy) for the three audio conditions. The most accurate target selection

Figure 13. Boxplot of distance versus audio conditions and angles

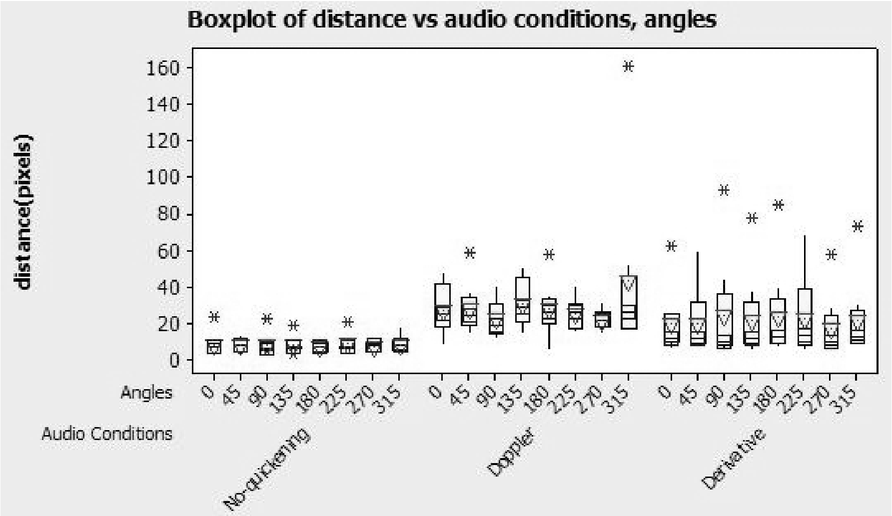
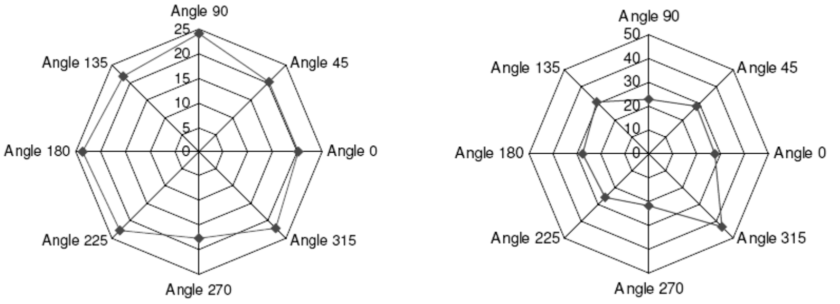
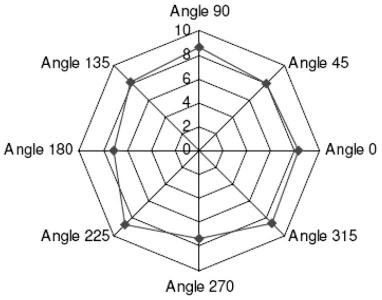


Figure 14. Average of position error in pixels for all participants in 3 audio conditions



(a) Average of position error in the Derivative condition for all users

(b) Average of position error in the Doppler condition for all users



(c) Average of position error in the No-Quickening condition for all users occurred in the no-quickening condition, and there was little difference between the levels of accuracy for each orientation. The spread for no quickening was very small, and the five outliers are not very far away from the median, suggesting that, overall, most people in this condition took approximately the same length of time to select a target. The derivative condition takes longer overall than the no-quickening condition. The spread is larger (largest of all three conditions) and the outliers further away from the median than in the no-quickening condition. The participants in the Doppler condition took longer on average to select the targets, compared with those in the other two conditions. The spread in the Doppler condition is smaller than in the derivative condition, but larger than in the no-quickening condition. This shows that angles 90° and 270° have higher accuracy. A *GLM ANOVA* analysis found that there was a significant effect of audio type on the distance ($F(2,21)=4.345$; $p<0.05$). There were, however, no significant differences between the eight audio orientations. Post-hoc analysis using a *Tukey* test showed that there was a significant difference between the estimated mean distances of the selected targets away from the audio sources in two of the three audio conditions. It was found that the mean distances were not significantly different between two quickening feedbacks and nonquickening condition ($p<0.087$). The *Tukey* test showed that there was no significant difference between the estimated mean distances of the selected targets away from the audio sources in the derivative condition. Running this test on Doppler results revealed that a significant difference ($p<0.025$) between the estimated mean distances of the two selected targets away from the eight audio orientations (90° and 270°). In the no-quickening condition, the *Tukey* test showed that there is no significant difference among the estimated mean distances of the selected targets away from the audio sources.

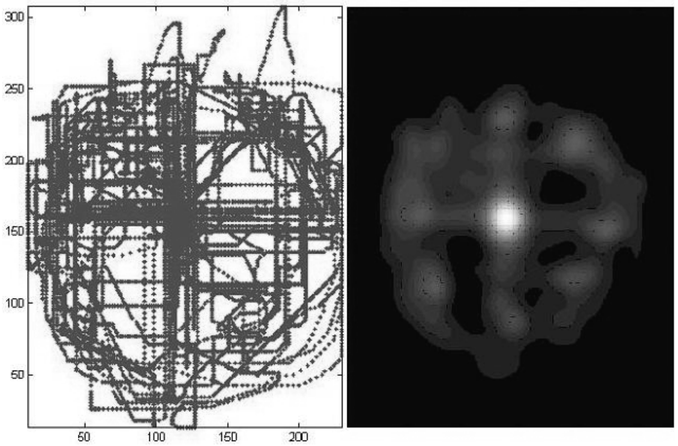
Search Patterns and Covered Space

In this experiment, audio targets had fixed positions in the state space so the observed search patterns were different than those we found in the pilot study. Twenty-four total patterns were analysed in this experiment. From the training sessions, subjects knew the approximate location of the audio sources. In 100% of search patterns in the no-quickening condition, the subjects moved to the edge of a circle in the size of the actual radius of the points and started circling to find the active target. In the Doppler condition, 87% of subjects could guess in which direction the target was located and after doing a few back and forth movements, they landed on the target. Subjects, therefore, followed a sector search pattern. Search patterns of subjects who worked with derivative volume adaptation were mixtures of the patterns of the no-quickening and Doppler conditions. Figure 15 shows some of the subjects' trajectories and density plots in different conditions. Figure 16 shows the percentage of the screen covered by participants' movement in three conditions. In the derivative condition, the top-left sections (90°-180°) were explored more than other parts. In the Doppler condition, the top-right (0°-90°) sections were popular to explore, and in "no quickening" there was no significant difference in the sections covered by participants' movement and all of the partitions were explored equally. These plots show the Doppler condition had the most covered space with 44.5% and the rest had a similar percentage of coverage, 39%.

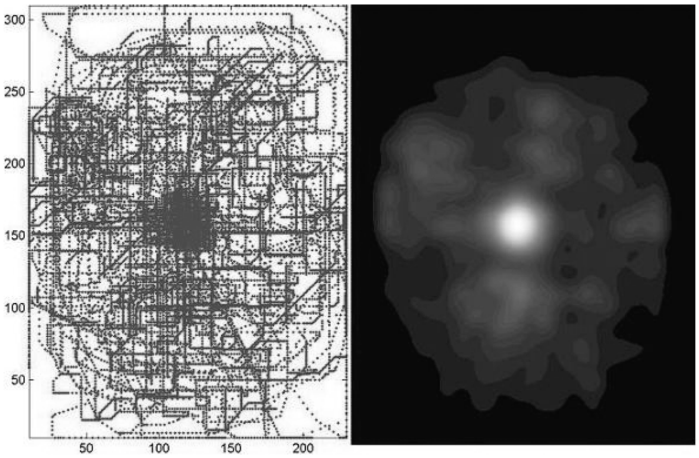
Discussions

In this experiment, it was found that there was an effect of audio condition on the level of accuracy. When the feedback was no quickening, participants were more accurate than when the feedback was Doppler or derivative. This is due to the fact that with the no-quickening condition, the only time that audio feedback is heard is when the cursor is directly over the target, and the difference between hearing and not hearing audio feedback is larger than the difference between hearing different levels of audio feedback. There was also found to be no effect of angle. The level of accuracy was the same, irrespective of the orientation of the target.

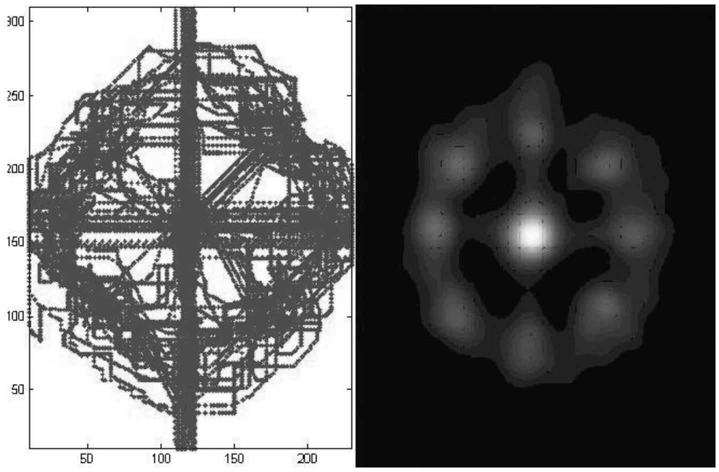
Figure 15. Trajectories of different subjects in 3 audio conditions



a) Trajectory and density plot of subject 4 in Derivative

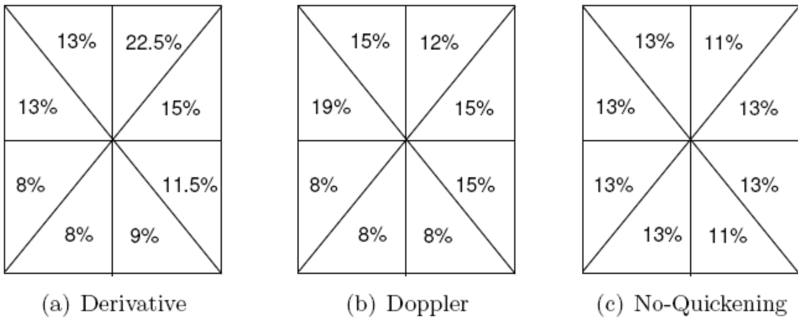


b) Trajectory and density plot of subject 4 in Dopler



c) Trajectory and density plot of subject 4 in no quickening

Figure 16. Percentage of the screen covered by users' movement in different conditions. The variance of the coverage in the derivative condition is 0.0023, in the Doppler condition is 0.0017 and in the no-quickening condition is 8.5714e-005.



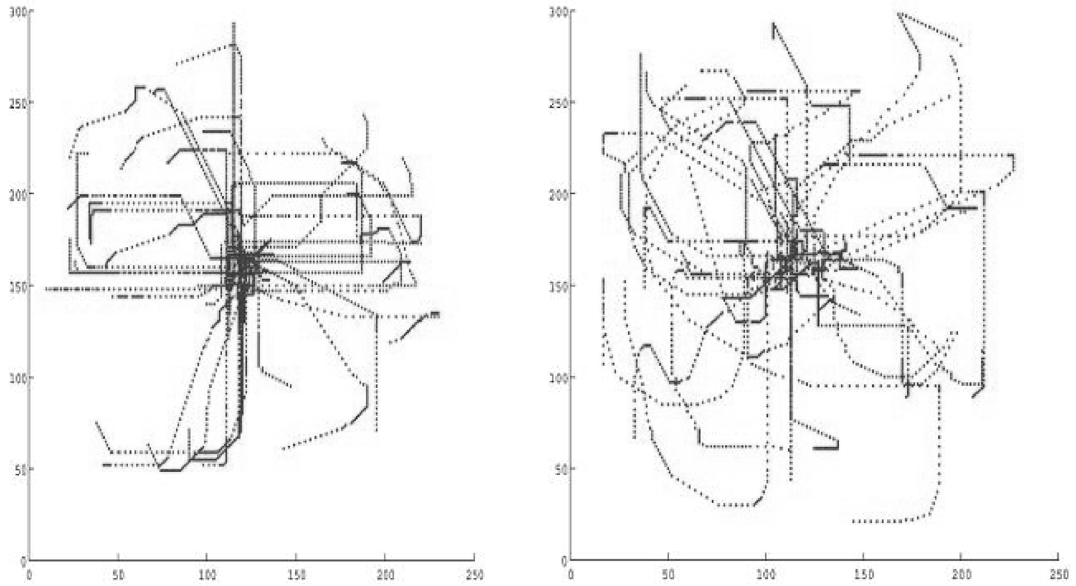
(a) Derivative (b) Doppler (c) No-Quickening

Many participants reported that sometimes they would just "land" on the audio source by chance, and at other times they would search for a long time and still not feel they had located the point accurately. This was an especially common complaint by participants in the Doppler condition. Disorganised search patterns observed in the Doppler condition, for instance in Figure 15(b), may correspond with the claim made by some participants that by the time they had established where the target was through the audio feedback, they had already passed the audio source, and had to go back to it. We plotted users' trajectories when they had moved from the centre of the screen to any active audio source around the centre. Figure 17(b) and other users' trajectories in the Doppler condition highlight that in the first moments after audio source activation, the subjects could guess the direction (left or right side of the space) and approximate position of the target and consequentially, moved towards the target correctly, but it was difficult to establish a correct target acquisition, and they made some back and forth or up-down movements to land on the target, which is compatible with the observed sector search pattern. This is shown more clearly in Figure 18(a-left) and the user's trajectories as time series in Figure 18 (b-left) in an individual target acquisition task when the user has moved from the centre to the target in angle 45°, which has been activated.

Figures 15(a) and 17(a) show a trajectory that is fairly typical for most participants in the derivative condition. It can be seen that the trajectory is far more ordered, with participants moving in the horizontal and vertical directions (in the directions of 0°, 90°, 180°, and 270°), more so than in the Doppler condition. It becomes obvious that participants moved in a circular motion that they learnt during the training session, far more so than those in the Doppler condition. From this, they have established that the audio sources in the experimental session were also arranged in a circle. This suggests that they were not necessarily using only audio feedback, but also prior knowledge about the probable locations of the audio sources. Since this circle is not as clearly defined in the Doppler trajectories, it suggests that participants in the Doppler condition were using predictive information, but were also less able to control their movements efficiently. There is a risk that any significant effects were masked by prior knowledge of the way the audio sources were arranged, and visual feedback has affected the users' behaviour in exploring the audio space.

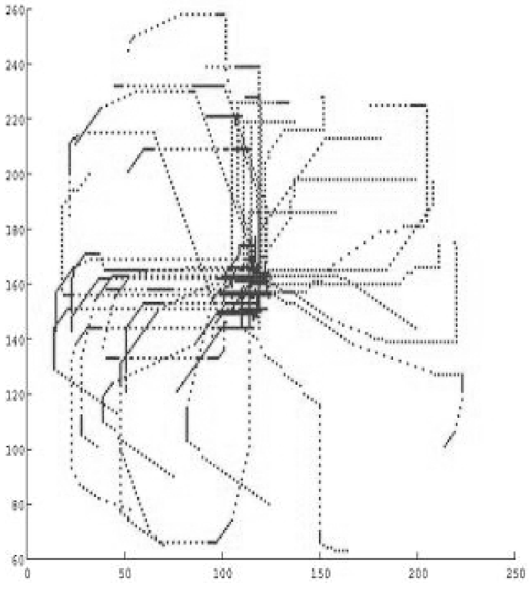
Figures 17(a) and 19(a-left) provide a clearer picture of the users' browsing behaviour in this condition. As a result of the first impressions that the users have received from the volume of the audio source, they have chosen vertical or horizontal directions. Whenever they have not found the target in these directions, for instance in an individual target acquisition in angle 45° presented in Figure

Figure 17. Trajectories of different subjects in 3 audio conditions when they have moved from the centre to outlying active audio sources



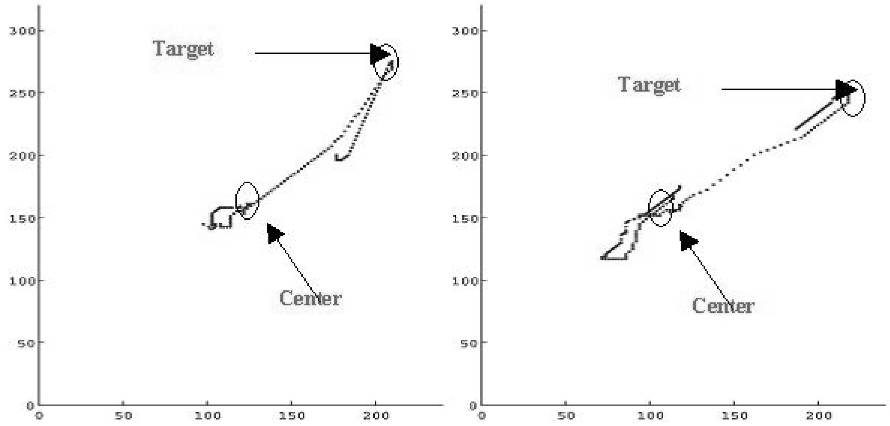
a) Trajectory and density plot of subject 1 in Derivative case

b) Trajectory of subject 4 in Doppler case

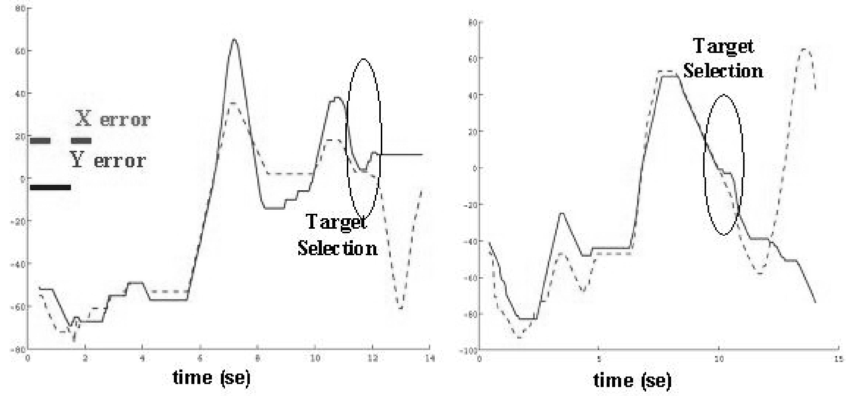


c) Trajectory of subject 4 in no quickening case

Figure 18. (a) One of the participants' trajectories in the Doppler condition with and without predictive feedback in an individual target acquisition task when the user has moved from the centre to an activated target in angle 45°. (b) The time series of the participant's X and Y position error in the same task.



(a) An individual target acquisition task in the Doppler condition. (left) standard feedback, (right) predictive feedback



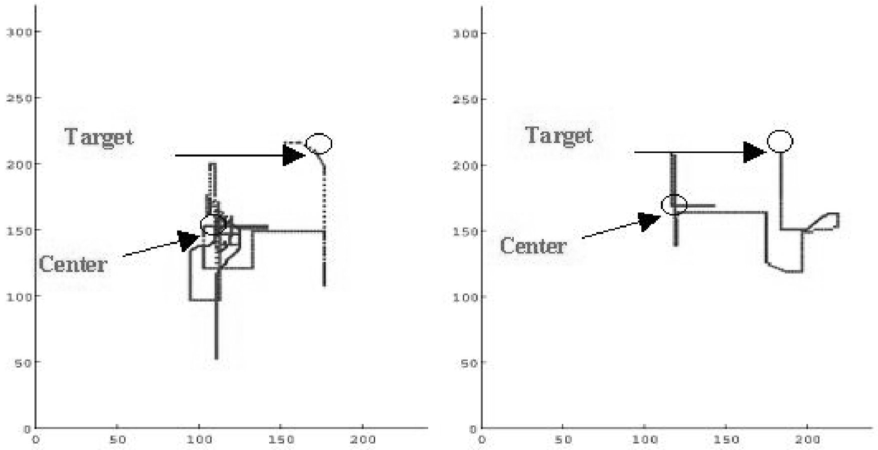
(b) The time series of one of the participants' position error in X and Y axis in the target acquisition task shown above, in the Doppler condition. (left) standard feedback, (right) predictive feedback

19, they have moved around the circle using prior knowledge of the landscape. Figure 15(c) shows a more pronounced circle and cross-shape for the no-quickening condition. Since the participants in the no-quickening condition were presented with no aural feedback except when directly over the target, it is, most likely that they were relying on the circular target distribution previously seen in the training. This led to a systematic search strategy, less "browsing around" because of the lack of predictive ability without quickening. All users who participated in this condition claimed that this was not an exciting method of exploring the auditory space.

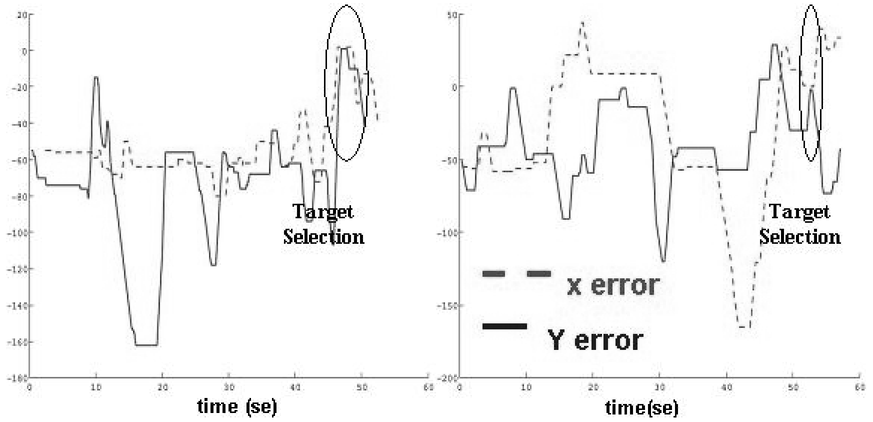
Human Operator Modeling

In continuous control tasks, for instance browsing and finding audio targets in the audio space, the human operator can be modeled using the tools of

Figure 19. (a) One of the participants' trajectories in the derivative condition with and without predictive feedback, in an individual target acquisition task when the user has moved from the centre to an activated target in angle 45°. (b) The time series of the participant's X and Y position error in the same task.



(a) An individual target acquisition task in the Derivative condition (left) standard feedback, (right) predictive feedback



(b) The time series of one of the participants' position error in X and Y axis in the target acquisition shown above, in the Derivative condition (left) standard feedback, (right) predictive feedback

manual control theory. Quantitative models of the human operator may provide predictions and insights into basic properties of human performance in human-machine interaction, and the ability to derive transfer function for human operators means they can be directly used in human-machine systems interaction design (Jagacinski & Flach, 2003).

Human interaction with the computing systems is two-way. The user and machine form a closed-loop system, where s/he issues commands through the system's input channels and receives results fed back on output channels. Subsequent input depends on the latest output. The performance is adversely affected when the feedback is subject to delay or lag (Jagacinski & Flach, 2003; MacKenzie & Ware, 1993).

In some genres of interactive systems, which rely heavily on the tracking of hand, head, and/or body motion in a simulated environment, the pretence of reality requires a tight coupling between the user's view and hearing of the environment and the actions, usually hand, head, and body motions, that set the view and hearing. When changes in the environment lag behind input motions, the loss of fidelity is dramatic (MacKenzie & Ware, 1993). In an extension to the previous experiments, we carried out a primary investigation to measure and model the speed, accuracy, and bandwidth of human motor-sensory performance in interactive tasks subject to lag.

Manual control theory suggests that in a simple tracking task the human operator can be modeled via a transfer function that consists of a gain, a lag (an integrator at higher frequencies) and a time delay (Jagacinski & Flach, 2003; Poulton, 1974).

$$Y_h(j\omega) = \frac{Ke^{-j\omega\tau}}{j\omega} \tag{3}$$

The gain ***K*** is a scaling factor that influences the bandwidth of the control system. The time delay t reflects human reaction time. In simple tracking tasks, the range of the time delay is between 20 to 150 ms, which overlaps with measures of reaction time in response to continuous stimuli. If ***K*** is low, the system will respond very sluggishly, moving only slowly towards the target signal. Conversely, if ***K*** is high then the system is likely to overshoot, requiring adjustment in the opposite direction, which itself may overshoot, leading to oscillation. However, humans adjust their gain to compensate increases or decreases in plant gain. For example, pilots change their behaviour when they switch from Boeing 747 (heavy) to an aerobatic airplane so the total open-loop gain remains constant. So if Y_p represents the plant transfer function and Y_h represents the human transfer function, then:

$$Y_h(j\omega)Y_p(j\omega) = const \tag{4}$$

The delay τ can also contribute to this behaviour; a high delay makes oscillatory behaviour much more likely (refer to time and delay section discussed earlier). The lag suggests that the human tracker has a low pass characteristic, that is, the human responds to low-frequency components of errors and ignores (or filters out) the high-frequency components of error (MacKenzie & Ware, 1993).

Using the platform in the previous experiment, we did a preliminary investigation to measure and model the accuracy, and bandwidth of human motor-sensory performance in interactive tasks subject to lag. We kept the same format of the second experiment but instead of providing audio feedback to the user's current position, we provided feedback to the user's predicted position, calculated according to equation (3):

$$X_{t+\tau} = X_t + V\tau \tag{5}$$

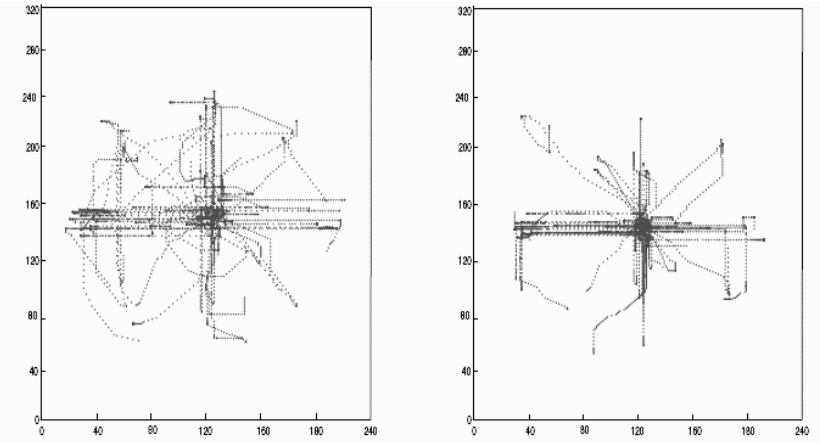
The volume of the audio source, which provides feedback about the target's position, is a function of the user's current velocity and position. Here, X_{t+1} and X_t are the user's position at time $t + \tau$ and t or current position and next possible position respectively. τ is the human's time-delay or reflection time, which becomes our "prediction horizon" for the predictive model, and V is the user's speed of motion in the audio space. This has the effect that, as the user moves toward the target, s/he feels him/herself in the position predicted to be reached at time $t + \tau$.

Design

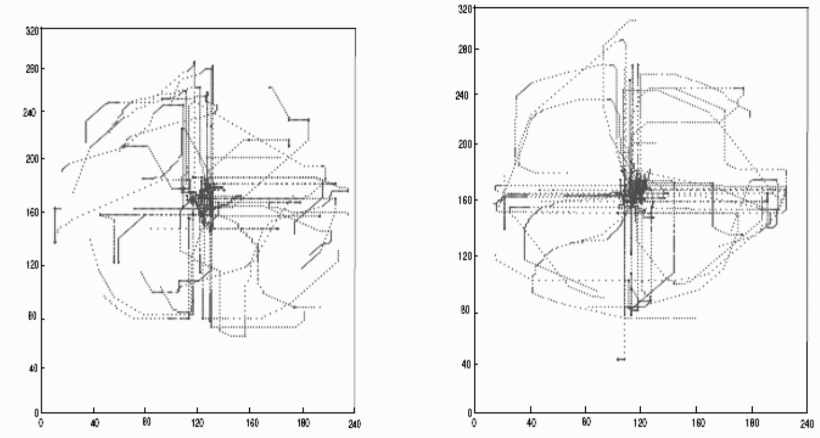
In the second experimental setup, we added a smooth drop-off in the time horizon of the prediction as the target was approached. The falloff began at radius 15 pixels, and once the user was within 5 pixels of the source, the feedback reverted to standard feedback with no predictive element.

In a pilot study, we ran the application for three participants familiar with the Pocket PCs and accelerometer. Neither felt any difference in the two derivative conditions, with and without prediction. In Figures 19(a) and 20(b), we see one of the participant's trajectories when he has moved from the centre to any active target. Providing feedback to the user's future position in the derivative condition has not much changed the user's exploratory behaviour. But users reported a great difference between Doppler with the pre-

Figure 20. The trajectories of 2 subjects in the Doppler and derivative conditions, without and with predictive feedback



(a) Doppler-normal feedback (b) Doppler-Predictive feedback



(c) Derivative-normal feedback (d) Derivative-Predictive feedback

dictive model-based feedback and Doppler with no predictive feedback. They said it felt they were able to acquire the direction of the audio source more quickly in the predictive case, but that it was more difficult to land on the source. In the standard model with no predictive element they felt it was slower to find the direction, but easier to land. Despite their perceptions, the trajectories in Figures 18(a-right) and 20(a) suggest the opposite case, that they performed fewer oscillatory movements around the target in the predictive model, compared to the standard case.

Prediction in the Doppler case allowed the users to converge more rapidly and directly to the target, but it seemed less helpful very close to the target. In the derivative condition, predictive feedback seems to have smoothed the behaviour, but has not improved the initial target localisation. These preliminary explorations suggest that a more detailed investigation of incorporating the predictive element in the feedback system would be of interest.

CONCLUSION AND SUMMARY

This work presents initial experimental results exploring the use of quickened audio displays for localisation and selection based on tilt control of mobile devices. The experiments provided useful exploratory information about how users navigate in such environments, and highlighted some benefits and disadvantages of each of the display options investigated. Users used a search method, which they felt more comfortable with for browsing the space regardless of the sound treatment. Vibration was clearly perceived by users, but led them to spend more time circling around targets.

Average results in the pilot study on the metrics used suggest that participants were more accurate in target selection in the "no Doppler-no vibration" than other conditions. The results do not suggest that the use of Doppler or vibration brought consistently improved accuracy, but some people did very well with Doppler, and most stated that they found the vibration feedback useful. Longer studies might show different use in real-life tasks once users had familiarised themselves with the system.

The main study represented a more focused investigation, with fewer confounding factors. We increased the number of participants, placed all the targets at equal distances from the starting point (centre of the screen), did not include haptic feedback, chose western pop music that was familiar to all users, and allocated more time for allowing users to learn how to use the specific interface, which was new to all. We investigated whether quickening was more useful to users searching for targets in state-space than no quickening audio feedback. We also investigated if there was any advantage of using Doppler feedback over derivative, and if there was an effect of orientation in either Doppler, derivative, or no quickening; therefore, to find out if the results of the first experiment could have been masked by an interaction with the orientation of the targets. It was also found that there was no effect of the angle at which the audio source was located.

In a preliminary investigation to better understand the results and to guide future work, we performed an exploratory experiment with predictive model-based feedback. The model is based on human operator modeling in continuous tracking tasks, and it could take human response delays and lags into account (not considered in this work). Using this predictive model, we could improve the users' performance in the Doppler condition, and reduce their overshoots during landing on the target just by providing audio feedback about the user's predicted position instead of their current position. This suggests further research to investigate the benefits of explicitly incorporating models of human behaviour in the design of feedback methods.

Outlook for Mobile Interface Designers

These results are a useful starting point for further investigation into the types of feedback that are most useful and informative in assisting users of a tilt-controlled mobile device with multimodal feedback. Some of the visualisation tools used will be useful for other designers, but the work also gives an indication of the difficulty of designing experiments that test aspects of low-level perception of multimodal displays, without confounding factors from prior knowledge influencing the results. The experiments also show the need for longitudinal studies. As in early exploration of novel interfaces, much observed behaviour is related to the user exploring the novel interface, and might not be a reliable indicator of typical practiced behaviour. Supporting the design of interaction in mobile devices with multimodal interfaces is a key challenge in mobile HCI. We believe that further development of the model-based prediction techniques we have begun to explore in this chapter will not only give us a better understanding of typical user behaviour, but will provide a promising, scientific basis to support designers in creating more useable systems in a wide range of novel settings, with a range of sensors and displays.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of IRCSET BRG SC/2003/271 project *Continuous Gestural Interaction with Mobile devices*, HEA project *Body Space*, and SFI grant 00/PI.1/C067, the IST Programme of the European Commission, under PASCAL Network of Excellence, IST 2002-506778. This publication only reflects the views of the authors.

REFERENCES

Birmingham, H. P., & Taylor, F. V. (1954). A design philosophy for man-machine control systems. In *Proceedings of the Institute of Radio Engineers* (pp. 1748-1758).

Blattner, M., Papp, A., & Glinert, E. (1992). Sonic enhancements of two dimensional graphic displays. In *Proceedings of International Conference on Auditory Display, ICAD'92* (pp. 447-470). Santa Fe, NM.

Bootsma, R. J., Fernandez, L., & Mottet, D. (2004). Behind Fitts' law: Kinematic patterns in goal-directed movements. *International Journal of Human-Computer studies, IJHCS, 61*(1), 811-821.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound.* Cambridge, MA: MIT Press.

Brewster, S. A. (1997). Using non-speech sound to overcome information overload. *Special Issue on Multimedia Displays* (pp. 179-189).

Brewster, S. A. (2002). Overcoming the lack of screen space on mobile computers. In *Proceedings of Personal and Ubiquitous Computing* (pp. 188-205).

Brewster, S. A., & Murray, R. (1998). Presenting dynamic information on mobile computers. *Personal Technologies* (pp. 209-212).

Chion, M. (1994). *Audio-vision*. New York: Columbia University Press.

Cook, P. R. (2002). *Real sound synthesis for interactive applications.* Peters, A K, Limited.

Cook, P. R., & Lakatos, S. (2003). Using Dsp-based parametric physical synthesis models to study human sound perception. *IEEE Workshop, Applications of Signal Processing to Audio and Acoustic*s (pp. 75-78).

Crossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' law. *Quarterly Journal of Experimental Psychology, (Original work presented at the meeting of the Experimental Psychology Society, Oxford, England, July 1963), 35A*, 251-278.

Doherty, G., & Massink, M. (1999). Continuous Interaction and Human Control. In *Proceedings of the XVIII European Annual Conference on Human Decision Making and Manual Control* (pp. 80-96). Loughborough: Group-D Publications.

Dong, L., Watters, C., & Duffy, J. (2005). Comparing two one-handed access methods on a PDA. In *Proceedings of the MobileHCI'05* (pp. 291-295). Salzburg, Austria.

Eslambolchilar, P., Crossan, A., & Murray-Smith, R. (2004). *Model based target sonification on mobile devices.* Paper presented at the Interactive Sonification Workshop, ISON, Bielefeld University, Germany.

Faconti, G., & Massink, M. (2001). Continuous interaction with computers: Issues and requirements. In *Proceedings of Universal Access in HCI, Universal Access in HCI - HCI International 2001* (pp. 301-304), New Orleans.

Fallman, D. (2002a). An interface with weight: Taking interaction by tilt beyond disembodied metaphors. *4th International Symposium on Mobile Human-Computer Interaction* (pp. 291-295).

Fallman, D. (2002b). Wear, point and tilt. In *Proceedings of the Conference on Designing Interactive Systems: Processes, practices, methods, and techniques* (pp. 293-302).

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*(1), 381-391.

FMOD. (2004). *Music and sound effect system*.

GAPIDraw. (2004). *Graphics API Draw.*

Gaver, W., Smith, R., & O'Shea, T. (1991). Effective sounds in complex systems:The ARKola simulation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'91* (pp. 471-498).

Gaver, W. W. (1989). *The sonic finder* (Vol. 4). Elsevier Science Publishing.

Gaver, W. W. (1997). *Auditory interfaces* (2nd ed.): *Handbook of human-computer interaction*.

Harrison, B., & Fishkin, K. P. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'98* (pp. 17-24), Los Angeles.

Hermann, T., Hansen, M. H., & Ritter, H. (2000). Principal curve sonification. In *Proceedings of International Conference on Auditory Displays, ICAD'00* (pp. 81-86).

Hermann, T., & Ritter, H. (1999). Listen to your data: Model-based sonification for data analysis. *Intelligent Computing and Multimedia Systems* (pp. 189-194). Baden-Baden, Germany.

Hinckley, K., Pierce, J., Horvitz, E., & Sinclair, M. (2005). Foreground and background interaction with sensor-enhanced mobile devices. *ACM Transactions on Computer-Human Interaction (TOCHI)* (pp. 31-52).

Hoffmann, E. R. (1991). Capture of moving targets: A modification of Fitts' law. *Ergonomics, 34*(2), 211-220.

Holmes, J. (2005). Interacting with an information space using sound: Accuracy and patterns. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display* (pp. 69-76). Limerick, Ireland.

Jagacinski, R. J., & Flach, J. M. (2003). *Control theory for humans: Quantitative approaches to modeling performance.* Lawrence Erlbaum Associates, Inc.

Jenison, R. L. (1997). On acoustic information for motion. *Ecological Psychology, 9*(2), 131-151.

Johnson, C., Brewster, S. A., Leplatre, G., & Crease, M. G. (1998). Using non-speech sounds in mobile computing devices. *First Workshop on Human Computer Interaction with Mobile Devices*, Department of Computing Science, University of Glasgow, Glasgow, UK (pp. 26-29).

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). *Sonification report: Status of the field and research agenda*. Report prepared for the National Science Foundation by members of the International Community for Auditory Display.

Langolf, G. D., Chaffin, D. B., & Foulke, J. A. (1976). An investigation of Fitts' Law using a wide range of movement amplitudes. *Journal of Motor Behaviour, 8*, 113-128.

MacKenzie, I. S., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems - INTERCHI '93* (pp. 488-493). Amsterdam.

Mackinlay, J. D., Robertson, G. G., & Card, C. K. (1991). The perspective wall: Detail and context smoothly integrated. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'91* (pp. 173-179). New Orleans, LA.

Oakley, I., Ängeslevä, J., Hughes, S., & O'Modhrain, S. (2004). Tilt and feel: Scrolling with vibrotactile display. In *Proceedings of Eurohaptics 2004*, Munich, Germany.

Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G., & Want, R. (2002). TiltType: Accelerometer-supported text entry for very small devices. *UIST'02: Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (pp. 201-204). Paris.

Patrol, C. A. P.-R. A. (1999). *Mission observer course manual.* Civil Air Patrol.

Poulton, E. C. (1974). *Tracking skill and manual contro.* Academic press.

Rekimoto, J. (1996). Tilting operations for small screen interfaces. *UIST'96: Proceedings of the*

9th Annual ACM Symposium on User Interface Software and Technology (pp. 167-168). Seattle, WA.

Rinott, M. (2004). Sonified interactions with mobile devices. In *Proceedings of the International Workshop on Interactive Sonification*, Bielefeld, Germany.

Sazawal, V., Want, R., & Borriello, G. (2002). The unigesture approach, One-handed text entry for small devices. In *Proceedings of Mobile Human-Computer Interaction - Mobile HCI'02* (pp. 256-270).

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology, 19*(8), 1065-1087.

Smith, N., & Schmuckler, M. A. (2004). The perception of tonal structure by the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 268-286.

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors, 48*(2), 265-278.

Wigdor, D., & Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. *UIST'03: Proceedings of the 16th annual ACM symposium on User interface software and technology* (pp. 81-90). Vancouver, Canada.

Williamson, J., & Murray-Smith, R. (2004). Pointing without a pointer. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'04* (pp. 1407-1410). Vienna, Austria.

Williamson, J., Strachan, S., & Murray-Smith, R. (2006). It's a long way to Monte Carlo: Probabilistic display in GPS navigation. *In Proceedings of Mobile Human-Computer Interaction, Mobile HCI'06* (pp. 89-96). Helsinky, Finland.

KEY TERMS

Continuous Control: A continuous control system measures and adjusts the controlled quantity in continuous-time.

Gestural Interfaces: Interfaces where computers use gestures of the human body, typically hand movements, but in some cases other limbs can be used, for example, head gestures.

Haptic Interfaces: Convey a sense of touch via tactile or force-feedback devices.

Manual Control: A branch of control theory that is used to analyse human and system behaviour when operating in a tightly coupled loop.

Nonspeech Sound: Audio feedback, that does not use human speech. The use of nonspeech sound in interaction has benefits such as the increase of information communicated to the user, the reduction of information received through the visual channel, the performance improvement by sharing information across different sensory modalities.

Prediction Horizon: How far ahead the model predicts the future. When the prediction horizon is well matched to the lag between input and output, the user learns how to control the system more rapidly, and achieves better performance.

Quickened Displays: Displays that show the predicted future system state, rather than the current measured, or estimated state.

Sonically Enhanced Interfaces: Interfaces where sound represent actions or content.

Sonification: The use of nonspeech audio to convey information or perceptualize data.

Sound Localisation: The act of using aural cues to identify the location of specific sound sources.