Crossmodal Audio and Tactile Interaction with Mobile Touchscreens

Eve Elizabeth Hoggan

Submitted for the degree of Doctor of Philosophy

Department of Computing Science, University of Glasgow

February 2010

Abstract

Touchscreen mobile devices often use cut-down versions of desktop user interfaces placing high demands on the visual sense that may prove awkward in mobile settings. The research in this thesis addresses the problems encountered by situationally impaired mobile users by using crossmodal interaction to exploit the abundant similarities between the audio and tactile modalities. By making information available to both senses, users can receive the information in the most suitable way, without having to abandon their primary task to look at the device.

This thesis begins with a literature review of related work followed by a definition of crossmodal icons. Two icons may be considered to be crossmodal if and only if they provide a common representation of data, which is accessible interchangeably via different modalities. Two experiments investigated possible parameters for use in crossmodal icons with results showing that rhythm, texture and spatial location are effective.

A third experiment focused on learning multi-dimensional crossmodal icons and the extent to which this learning transfers between modalities. The results showed identification rates of 92% for three-dimensional audio crossmodal icons when trained in the tactile equivalents, and identification rates of 89% for tactile crossmodal icons when trained in the audio equivalent.

Crossmodal icons were then incorporated into a mobile touchscreen QWERTY keyboard. Experiments showed that keyboards with audio or tactile feedback produce fewer errors and greater speeds of text entry compared to standard touchscreen keyboards. The next study examined how environmental variables affect user performance with the same keyboard. The data showed that each modality performs differently with varying levels of background noise or vibration and the exact levels at which these performance decreases occur were established.

The final study involved a longitudinal evaluation of a touchscreen application, CrossTrainer, focusing on longitudinal effects on performance with audio and tactile feedback, the impact of context on performance and personal modality preference. The results show that crossmodal audio and tactile icons are a valid method of presenting information to situationally impaired mobile touchscreen users with recognitions rates of 100% over time. This thesis concludes with a set of guidelines on the design and application of crossmodal audio and tactile feedback to enable application and interface designers to employ such feedback in all systems.

Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my supervisor Stephen Brewster for giving me the opportunity to conduct this research and for continually inspiring me throughout my PhD. Thanks also to my second supervisor, Matthew Chalmers, for all of his input and bringing a different point of view to the table.

Special thanks go to Topi Kaaresoja for acting as a third supervisor, supporting my research and for giving me the brilliant opportunity to experience research in an industrial environment.

Having been lucky to be part of the Multimodal Interaction Group at Glasgow, I have been fortunate to work with a large number of excellent researchers. Huge thanks go to everyone in the Computing Science Department who helped to proofread this thesis.

I would like to thank Helen Purchase for inviting me to do a summer project that inspired me to enter academic research.

Last but not least, I would like to offer many thanks to my family for helping me throughout my PhD with lots of encouragement and support.

This research was funded, in part, by Nokia Research Center, Helsinki.

Declaration

The contents of this thesis are entirely the author's own personal work. This thesis only makes use of parts of these papers that are directly attributable to the author. All other material has been referenced and given full acknowledgement in the text.

Experiment 1a in Chapter 4 has been published as a Work In Progress in CHI 2006 [60], co-authored by Stephen A. Brewster.

Experiment 1b in Chapter 4 has been published as a Work In Progress in CHI 2007 [64], co-authored by Stephen A. Brewster.

Experiment 2a and 2b in Chapter 4 have been published in NordiCHI 2006 [61], co-authored by Stephen A. Brewster.

Experiment 3 in Chapter 5 has been published in ICMI 2007 [63], co-authored by Stephen A. Brewster.

Experiment 4 in Chapter 6 has been published in CHI 2008 [65], co-authored by Stephen A. Brewster and Jody Johnston.

Experiment 5 in Chapter 6 has been published in CHI 2009 [66], co-authored by Andrew Crossan, Stephen A. Brewster and Topi Kaaresoja.

Experiment 6 in Chapter 7 will appear in CHI 2010 [62], co-authored by Stephen A. Brewster.

Chapter 1 Introduction

This thesis presents a study of crossmodal audio and tactile interaction with mobile touchscreen displays. We spend increasing amounts of our daily lives using mobile devices in many different contexts, thus it is necessary to research how to present information to users in ways suitable for these varying situations. Mobile devices have transformed significantly over the years, from initially only having basic phone call or messaging features, into powerful, Internet connected, music and video playing devices whilst managing all kinds of personal information like email, calendars and address books. Furthermore, a single device now integrates a camera, GPS, music player and voice recorder in one, freeing users from carrying multiple devices. There is an increasing number of mobile applications available for these devices and there is a challenge in designing these applications given that there is a large variety of information that needs to be displayed visually on extremely small screens. This places a high demand on the visual sense and explains the extent to which users can often spend more time focused on the screen than on the environment or task in hand.

Many commercial devices employ the use of audio and tactile feedback to provide simple alerts, such as incoming call notifications, through the use of ringtones and vibrations. The possibilities of communicating information and enhancing interaction through senses other than vision, for example, sound and touch, has generated a rich body of research [11, 72] [83] [111]. The existing research has demonstrated that audio and tactile feedback can be beneficial to users, increasing typing speeds and reducing errors with some training.

Mobile phones are personal devices, always on and always with us, which means that whether it is in our bag or pocket, or we are in a meeting, at a party, or listening to music, we still want to be able to interact with our device. In these situations, visual feedback is not always appropriate. Although a user's eyes may be busy focusing on the primary task, many activities do not otherwise restrict users from attending to information using their remaining available senses. This is when multimodal interaction is of benefit so that, for instance, messages can be presented through the audio modality and alerts can be presented through the tactile modality. Unfortunately, when the device is in a bag or pocket, tactile feedback can go unnoticed. When a user is in a noisy environment like public transport or listening to music, audio feedback can be ineffective. For example consider this typical usage scenario:

Sam is on her way to a business meeting walking along a busy street with her mobile phone in her bag when she receives an important calendar reminder. As her phone is not in contact with her body, a tactile alert would probably go unnoticed so the reminder would be best presented in audio. Next, Sam boards a train to continue her journey and as the train leaves the station, she starts downloading some music for her phone. Given that the train is noisy and she has placed her phone back in her pocket so she can read the newspaper, audio alerts alone would be insufficient to inform her of her completed download. At the same time, tactile alerts would be slightly masked, as the phone is not in direct contact with her skin. At this time, a combination of audio and tactile feedback could let her know when her song has been downloaded. Finally, Sam arrives at her business meeting. As the boss makes a presentation, Sam receives an urgent email from her husband. Everyone in the meeting room is listening to the presentation and it would be rude for Sam to disrupt the meeting with audio feedback informing her of the incoming email. In this case, a tactile cue would be much more subtle and more socially acceptable. This scenario is an example of the need for mobile devices to provide alternative presentation modalities through which information may be presented if the context requires. As the context changes, so should the feedback modality.

As mentioned, multimodal feedback is often used to reduce the visual load on mobile device users. There has been a large body of research into mobile multimodal interaction with each individual modality [46] [11] [44] [34] [55] [98]. However, as this scenario has demonstrated, users need to be able to switch effortlessly between different modalities depending on the situation. Users also need the option of several different modalities. Much of the research so far does not give the user a choice of modalities but simply provides one output modality, resulting in unimodal interaction.

The approach used in this research to combat the problems mentioned above involves crossmodal audio and tactile feedback. Unlike multimodal interaction, crossmodal interaction uses the different senses to provide the same information [52] (a more in-depth definition can be found in Chapter 3). This is much like sensory substitution where one sensory modality is used to supply information normally gathered by another [87]. Sensory substitution systems have proven to be an effective means of communicating information to people with sensory impairments so could provide an alternative method through which information can be presented to mobile device users. By employing concepts from sensory substitution, mobile devices could translate data into an auditory or tactile form so that it can be presented in the most appropriate modality to suit the context. For example, alerts providing information to the user about incoming messages (for example, SMS, MMS, or phone call) could be encoded using crossmodal methods in both the audio and tactile modalities. By making these alerts available to both the auditory and tactile senses, users can receive the information in the most suitable way, without having to abandon their primary task to look at the device.

The research presented here investigates the design of crossmodal auditory and tactile messages, called crossmodal icons. Two icons may be considered to be crossmodal icons if and only if they provide a common representation of data, which is accessible interchangeably via different modalities. These can be used in interfaces as a means of non-visual output and allow the investigation of user performance in different situations (in users' everyday lives) to establish whether one modality is more suited than the other and whether crossmodal audio and tactile feedback could be effective in real world applications in different contexts and under different degrees of workload. This thesis presents the very first formal investigations into crossmodal icons and the design of crossmodal audio/tactile feedback for mobile touchscreens.

1.1 Thesis Aims

This thesis asserts that using crossmodal auditory and tactile interaction can aid mobile touchscreen users in accessing data non-visually and, by providing a choice of modalities, can help to overcome problems that occur in different mobile situations where one modality may be less suitable than another. By encoding data using the crossmodal parameters of audio and vibration, users can learn mappings and translate information between both modalities. Therefore, data may be presented to the most appropriate modality given the situation and surrounding environment.

1.2 Research Questions

This thesis aims to answer the following questions:

> **RQ1**: What are the parameters of vibration and non-speech audio that can be manipulated to encode data in crossmodal icons?
>
> **RQ2**: What levels of performance can be achieved when these parameters are used to create multi-dimensional crossmodal icons?
>
> **RQ3**: Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?
>
> **RQ4**: Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

1.3 Thesis Walkthrough

Chapter 2, *Literature Review*, reviews related work on perception and the presentation of information through vibrations and audio along with current research on the use of mobile touchscreen devices. This chapter places the work of this thesis in context by summarising related work and identifying an area which has received little attention: crossmodal interaction. In addition, the findings from this related work are considered in terms of how they could be used to inform the design of crossmodal audio and tactile icons for mobile touchscreens.

Chapter 3, *Crossmodal Interaction*, defines crossmodal interaction with a focus on initial perceptual studies in the field of psychology. Then, the audio and tactile modalities are analysed in more depth with respect to their potential for use in crossmodal interaction.

Lastly this chapter contains a discussion of audio/tactile crossmodal icons and outlines the design approach used in this thesis.

Chapter 4, *Individual Design Parameters*, reports two experiments investigating the different possible parameters and mappings that can be used to facilitate crossmodal auditory/tactile feedback. The implications of the experimental findings are discussed and guidelines are drawn out from the results of these studies to help designers who wish to use these crossmodal parameters.

Chapter 5, *Multidimensional Crossmodal Icons*, discusses the development of a three-dimensional set of crossmodal icons, and then reports an experiment investigating the learning of such icons and the extent to which this learning transfers between the two modalities.

Chapter 6, *Applying Crossmodal Icons: Audio/Tactile Touchscreen Text Entry*, focuses on examining the incorporation of crossmodal icons into a mobile touchscreen application with an aim to find out if situationally impaired users can benefit from such crossmodal feedback. The design, implementation and evaluation of this crossmodal mobile touchscreen application explore the combination of many of the key features discussed in the preceding chapters.

Chapter 7, *CrossTrainer: Testing the Long-Term Use of Crossmodal Interfaces,* involves a longitudinal summative evaluation of a touchscreen application with crossmodal feedback for a range of different interface widgets with the aims to investigate the everyday use of crossmodal audio and tactile feedback and to study user performance and preference over time.

Chapter 8, *Discussion and Conclusions*, reviews the work presented in the thesis and its novel contributions in terms of the research questions outlined in the introduction. A set of guidelines is included, which can be used to inform the design of crossmodal interfaces. Lastly, the limitations of this work are outlined and possible future research directions are proposed.

Chapter 2 Literature Review

The aim of this research is to investigate crossmodal interaction with audio and tactile mobile touchscreen displays. Therefore, the purpose of this chapter is to provide an overview of the existing research in related fields such as audio and tactile feedback, multimodal interaction and current mobile touchscreen applications or solutions to place the contributions of this thesis in context.

The chapter begins by discussing the basic concepts in audio and tactile displays including human perception capabilities and methods of encoding information in these modalities. The remainder of the chapter reviews related research in the field of mobile touchscreen interaction using different feedback modalities, and is structured in terms of its main applications and evaluation environments. The chapter concludes with a summary of the main findings of the chapter and positions the contributions of this thesis within these related areas of research.

Research Question 1 asks:

RQ1: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

The literature review will detail aspects of both the auditory and tactile modalities with an aim of establishing the most successful parameters in each modality. These parameters, in turn, can then be investigated as potential parameters for crossmodal interaction.

2.1 The Auditory Modality

Given that this research focuses on the similarities between the auditory and tactile modalities, it is necessary first to examine each modality on its own. The audio modality, in terms of non-speech audio, is a widely researched field and provides a large body of literature from which this research draws. Sound, as defined by Moore [100], *"originates from the motion or vibration of an object. This motion is impressed upon the surrounding medium (usually air) as a pattern of changes in pressure"*. In terms of this research, the audio modality is used as output from mobile devices and involves the use of one of the many types of non-speech audio feedback: earcons [8] as described later in this section. By understanding both audio and tactile in depth it seems likely that the work from both these domains could be used to inform the design of crossmodal interaction.

Section 2.1 contains a brief introduction to audio perception, a discussion of the advantages and disadvantages of using non-speech audio feedback, ways in which to encode information in the audio modality through the use of earcons and Auditory Icons, followed by a review of different applications which employ audio feedback.

2.1.1 Perception and Parameters

Before using the audio modality to transmit information it is necessary to gain an understanding of the capabilities of humans to process audio stimuli. This section begins by providing an overview of the sense of hearing and then goes on to present results from the literature regarding perception of the different parameters of audio.

In simple terms, sound is made up of two measurable parameters: frequency and amplitude. Frequency is the number of times a waveform is repeated in a given amount of time. This is measured in Hertz (Hz) where 1Hz is equal to 1 complete cycle of the waveform per second. Humans can hear sound with frequencies in the range of 20Hz to

20kHz [101]. However with the onset of age, the upper limit on hearing tends to reduce to about 15kHz [108].

Amplitude is the difference between the mean pressure and the size of the pressure increase or decrease. The highest amplitude sounds that can be heard by humans, without damage to our sense of hearing, is approximately 120dB above the quietest sound we can perceive. In order to hear the above-mentioned frequencies, the amplitude must be altered. For example, low frequency sounds should be presented with high amplitudes whilst high frequencies need to be accompanied by lower amplitudes.

In addition to the above-mentioned primary components, there are many other dimensions of the audio modality as detailed below:

2.1.1.1 Duration

Duration is the attribute of audio that determines the length of the stimulus. The smallest detectable increase in duration is 4ms for 10ms stimuli, 15ms for 100ms stimuli and 60ms for 1000ms stimuli [100]. Additional factors must also be taken into account given that these values are based on the mean performance of participants. Performance could perhaps increase if users are musically trained.

2.1.1.2 Pitch

Pitch is the audio attribute that dictates the way in which sounds are ordered in a musical scale [4]. In general terms, it is related to the repetition rate of the waveform of a sound i.e. the frequency [100]. Unlike duration, pitch is a subjective attribute and assigning a pitch value simply means specifying the audio frequency. The problem with pitch is that it is difficult to distinguish. It is fairly easy to determine whether a sound is high or low but it is much harder to absolutely identify the pitch of a sound without having a reference note for comparison. Only 0.01% of the population have ‘perfect pitch’, which is the ability to absolutely identify the pitch of a sound [119]. Therefore, unless large differences between pitch values are used, pitch is not a particularly useful attribute for encoding data.

2.1.1.3 Localisation

The term localisation refers to the direction and distance of a sound source [100]. Our ability to localise sound depends on several factors, the most important of which are the time differences between sounds reaching our left and right ears. To be able to distinguish the direction and elevation of a sound source, binaural cues [100] consist of variations in the timing and range of sound between the ears.

There are now numerous software packages that allow the creation of spatial sound cues [6] such as AM:3D[1] (as used in this thesis in Chapter 4). The ability to create these synthetic spatial environments is made possible using head related transfer functions (HRTF). A HRTF modifies a sound source so that the listener perceives the sound to be coming from some position in space [154]. HRTFs allow designers to create virtual three-dimensional audio environments that can be easily rearranged without having to physically move sound sources such as speakers. Furthermore, these spatial auditory environments can be presented effectively via headphones or stereo speakers. This means that spatial audio cues can be used in a mobile environment [124].

2.1.1.4 Timbre

Plomp [110] defines timbre as the attribute of sensation in terms of which a listener can judge that two steady complex tones having the same loudness, pitch and duration are dissimilar. Timbre depends upon more than just the frequency spectrum of the sound; fluctuations over time can also play an important role [100].

Despite a vast amount of research, the components or dimensions of timbre are still not fully understood. However, it has been established that the number of harmonics and amplitude of each of these harmonics plays a large role in the perception of timbre [39]. Rigas [118] has carried out experimental studies in order to categorise MIDI (Musical Instrument Digital Interface) sounds in groups based on their subjective similarity. He presented listeners separately with tunes of 8 notes played on 23 different synthesised musical instruments and asked them to write down the name of the instrument that played the tune. He found that listeners most successfully identified pianos, organs, xylophones and drums. In a further study he presented listeners with a list of five named instruments (Piano, Guitar, Drums, Violin, Saxophone, Flute and Harp). Listeners were then played a sound of one of the instruments and had to select which one they heard. Rigas found high recognition rates with over 80% correct responses for each instrument except the harp, which had only 30% correct responses perhaps due to its esoteric nature.

---

[1] AM:3D Positional Audio, http://www.am3D.com/

2.1.1.5 Loudness

Loudness, defined as the *"attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud"* [152], is (as with pitch) subjective. Loudness is measured on the phon scale, where 1 phon is the loudness of a 1000Hz tone presented with the intensity of 1dB SPL [102]. Unfortunately, it can be difficult to use loudness as a parameter in audio feedback. Low levels of loudness can go easily unnoticed especially in mobile environments and high levels can be disturbing or painful.

2.1.2 Audio Encoding Strategies

Auditory display is an umbrella term referring to the use of any type of sound to present information to a listener. This may include, but is certainly not limited to warnings, alarms, status indicators, and data sonification [152].

The use of auditory icons to improve interaction with computer interfaces was first suggested by Gaver [46]. Two main types of audio encoding techniques exist: Auditory Icons [46] and earcons [8]. Auditory Icons are natural, everyday sounds used to represent events or items within a computer interface. The audio feedback used in Auditory Icons is semantically linked to the data it represents. For example, bumps, scrapes, or even files hitting mailboxes. This means that users can easily and quickly learn to interpret the icons. In contrast, earcons are structured, abstract non-speech audio messages. Earcons use musical, rather than naturally occurring, sounds and use an abstract mapping. There is no semantic link between the audio feedback and the data it represents. This means that users must be trained to understand the icons.

2.1.2.1 Earcons

Earcons are the auditory equivalent of visual icons, which have been defined as an image, picture, symbol, or sound representing a specific event, object or concept [130]. Earcons [8] are constructed from simple building blocks called motifs. These are short, rhythmic sequences that can be combined in different ways. Blattner *et al.* proposed the design of earcons but did not develop or test them. Using psychoacoustical methods, Brewster [17] [18] has conducted detailed investigations of earcons, which have shown that they are an effective means of communicating information in sound.

Blattner *et al.* [8] and Brewster [17] suggest their most important features are:

Rhythm: Using different rhythms can create a large number of distinguishable motifs. Blattner *et al.* describe rhythm as the most prominent characteristic of a motif. Brewster [17] stated that in order to ensure that rhythms can be absolutely identified they should be designed to be as different as possible from one another. The easiest way to achieve this is to use a different number of notes in each rhythm. Brewster also mentions that using musical rules can make earcons sound like a complete rhythmic unit. For example, the first note should be louder and the last note should have a longer duration to highlight the end of the unit. In terms of this research, a very important guideline is that earcon rhythms should be as short as possible so that they do not slow down any interaction with a system. Rhythm is a very effective parameter. McGookin [96] found that users were able to achieve absolute identification rates of over 90% for rhythm (melody) when used in three-dimensional Transformational earcons in combination with timbre and register. Transformational earcons map each attribute or dimension of data to a parameter as opposed to Inherited earcons which can be created using a tree structure, where every node in the tree is an earcon, and each earcon inherits from the levels above it.

Timbre: Motifs can be made to sound different through the use of different timbres, for example playing one motif with a trumpet and the other with a piano. Brewster [17] showed that using musical instrument timbres in earcons was more effective than using basic tones (e.g. sine waves). He also reported that it is important to select timbres that are subjectively easy to distinguish from one another. McGookin [96] found that users were able to achieve absolute identification rates of over 90% for timbre (piano/violin/trumpet) when it was used in three-dimensional Transformational earcons in combination with rhythm and pitch.

Register (Pitch): This is the position of the motif in the musical scale. A high register means a high-pitched note. There are 96 different pitches in the western musical system and these can be combined to produce a large number of different motifs [18]. The same motif in a different register can convey a different meaning. Brewster [17] suggests that register is a poor choice when absolute recognition is required and therefore it would be better to use it in combination with another parameter. Although, as mentioned above, pitch is subjective and can be difficult to identify absolutely. If register must be used then large differences between the different levels will be required. When participants were asked to complete absolute identification tasks on three-dimensional Transformational earcons (created with rhythm, register and timbre), McGookin [96] found that absolute identification rates of around 70% were achieved for register (low/med/high).

Dynamics: This is the change in volume of the motif. It can be made to increase as the motif plays (crescendo) or decrease (decrescendo).

Intensity (Loudness): Brewster [17] recommended that intensity should not be used as a parameter in earcons because users find loud sounds annoying and report annoyance when the volume level is out of their control.

Spatial Location: Spatial location has not been used a great deal in earcon design, except to help differentiate multiple earcons presented simultaneously [95] [94]. Brewster [18] suggested that different families of earcons could be presented from different locations but this has not been investigated.

Rate (Tempo): Changing the tempo, speeding up or slowing down the sounds, is another effective method for differentiating earcons [18].

Duration: earcons with up to six notes played in one second have been shown to be usable [8] [18].

This thesis examines audio and tactile feedback for use in crossmodal interaction. Earcons and the methods of encoding discussed above will be used as a basis for the audio portion of this work.

2.1.3 Key Audio Applications in Computing

The use of audio for non-visual information display has been widely investigated. One of the most extensive applications of auditory icons in research to date involves Gaver's Sonic Finder [47] that, when added to the Macintosh Finder, appropriates sounds for actions (for example, opening a file, dragging an object, or emptying the trash) using metaphorical mappings. No formal study was conducted but users commented that the sounds seem to be naturally integrated into the interface and appear intuitively accessible.

Feedback using the audio modality has also been applied to widgets in mobile devices. Brewster [12] developed the ideas of sonified buttons and applied them to buttons on the 3Com Palm series of pen-based handheld computers using simple earcons. One aim of the research was to see if the addition of non-speech audio feedback could reduce the size of the widgets so that screen space could be saved and another aim was to see the effects when users were walking. The overall results confirmed that the addition of sound allowed the participants to enter significantly more 5-digit strings compared to the silent condition, with smaller sonic buttons as effective as larger silent ones. In the walking condition there was a 20% drop in performance, with the audio interface still performing better than the standard one. The suggested reason for this was that users did not have to concentrate so much of their visual attention on the device, as much of the feedback was audio not visual, and so could focus on walking.

Leplatre and Brewster [85] added earcons to mobile phone menus to help users navigate the menu structure. An experiment was conducted to investigate menu navigation with and without non-speech audio feedback. The results showed that non-speech sound improves the performance of navigational tasks in terms of the number of errors made and the number of keypresses taken to complete the given tasks.

Sawhney and Schmandt [124] developed a wearable personal messaging audio system called *Nomadic Radio* to deliver information and messages to users on the move (Figure 2-1).

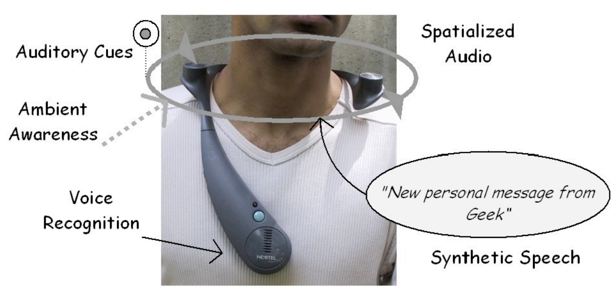

Figure 2-1: Example set-up of the Nomadic Radio system[2].

One of the aims of this system was to reduce interruptions caused by messages being delivered at an inappropriate time when users are situationally impaired (for example loud mobile telephone ringtones in a library). In the system, users wore a microphone and shoulder-mounted loudspeakers that provide a 3D soundscape through which the audio

[2] http://web.media.mit.edu/~nitin/NomadicRadio/

feedback was presented. A clock face metaphor was used with 12:00 in front of the user's nose, 3:00 by the right ear, 6:00 directly behind the head, etc. each message was displayed using the audio modality in the soundscape with the position mapped to the time at which the message arrived. The system attempted to calculate the most appropriate amplitude level to present the notifications by recording the background audio level in the user's surroundings. Results of an informal evaluation showed that a novice user could identify notifications in the soundscape successfully whilst attending to other tasks such as reading or typing. This system has shown that using 3D audio spatial locations is an effective way of presenting information for mobile users.

Another common use of 3D audio feedback is for navigation. Jones *et al.* [71] showed that directional cues can be presented successfully by panning music between a pair of stereo headphones. Spatial location is also a parameter that can be used in the tactile domain therefore it may be possible for crossmodal auditory and tactile spatial locations to present alerts and navigation information.

2.1.4 Summary

This section focused on the audio modality with an introduction to audio perception, ways in which to encode data in the audio modality through the use of earcons and Auditory Icons followed by a review of some different applications which employ audio feedback. A review of earcons and Auditory Icons showed the difference between semantically representing information in sound and using abstract encodings. Earcons use abstract mappings that must be learned, as there is no semantic link between the sounds and the data they represent. However, perhaps the most useful aspect of earcons is the fact that the parameters used are based on the basic dimensions of our sense of hearing. This provides a good link to the tactile modality because the basic dimensions of sound are based on vibrations. This review identified the most successful parameters in earcon design as timbre, rhythm (incorporating tempo and duration), and spatial location. Pitch, dynamics and intensity are also possible parameters but require further investigation.

This review has established that audio feedback can be used successfully to encode data and that there are several extremely effective parameters that can be used. The parameters examined in this review will be used as a basis for the design of the audio crossmodal feedback. The next step, detailed in Section 2.2, investigates current work in tactile feedback with a view to establishing the most effective tactile parameters for encoding information in a similar fashion to the audio parameters.

2.2 The Tactile Modality

Before designing tactile feedback it is necessary to gain an understanding of the capabilities of humans to process tactile stimuli. This section begins by providing an overview of the sense of touch and then goes on to present results from the literature regarding perception of the different tactile parameters, drawing conclusions about the implications of these results for the design of crossmodal tactile feedback.

2.2.1 Perception and Parameters

The term haptics means "*sensory and/or motor activity based in the skin, muscles, joints and tendons*" [1]. Under this umbrella term, however, there are several sub-categories as shown in Table 2-1.

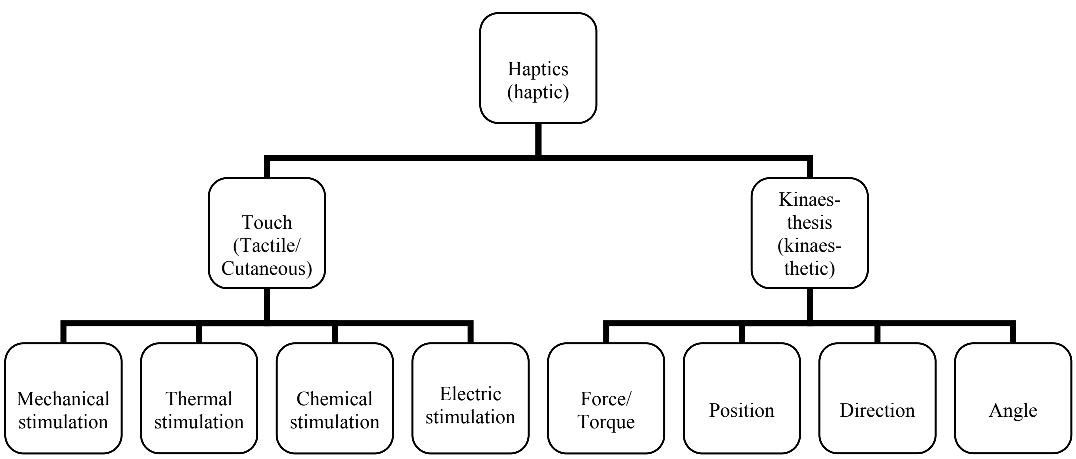

Table 2-1: Definitions of Terminology (adapted from [1]).

The work in this thesis makes use of tactile/cutaneous feedback through mechanical stimulation because kinaesthetic feedback is not so appropriate for mobile usage.

The skin has an area of 1.8 m^2, a density of 1250 kg/m^3, and a weight of 5 kg [128]. It is classified as either glabrous (i.e., non-hairy) skin, which is found only on the plantar and palmar surfaces, or hairy skin, which is found on the rest of the body. These divisions are relevant to tactile displays because they vary in sensory receptor systems and measures of tactile sensitivity [33]. Four types of mechano-receptive fibres have been identified in glabrous skin: Meissner corpuscle (RA), Merkel cell (SAI), Pacinian corpuscle (PC), and Ruffini ending (SAII). Table 2-2 shows a list of specific characteristics for each fibre.

	QuicklyAdapting	**Slowly Adapting**
Superficial skin	Meissner corpuscle (RA) • small receptive field • non-Pacinian (NP) I channel, not sensitive to temperature • 10 to 100 Hz • temporal summation: no • spatial summation: yes • local vibration and perception of localized movement	Merkel cell (SAI) • small receptive field • NP III channel, sensitive to temperature • 0.4 to 100 Hz • temporal summation: no • spatial summation: no • tactile form and roughness
Deeper tissue	Pacinian corpuscle (PC) • large receptive field • P-channel, very sensitive to temperature • 40 to 800 Hz • temporal summation: yes • spatial summation: yes • perception of external events	Ruffini ending (SAII) • large receptive field • NP II channel, sensitive to temperature • 15 to 400 Hz • temporal summation: yes • spatial summation: ? • not in glabrous skin

Table 2-2: Characteristics of the four types of mechano-receptive fibres in the human skin (adapted from [146]).

Each mechano-receptive fibre has a specific role in the perception of vibration that ranges from 0.4 to more than 500 Hz [128]. The Meissner corpuscles are high-density fibres that are abundant in the fingertips. The majority of the tactile feedback used in this thesis research is presented to the fingertips. In contrast, the Pacinian corpuscles are less dense than the Meissner corpuscles, and are numerous in the distal joints. Since the four fibres overlap in their absolute sensitivities, a vibration stimulus will seldom stimulate a single fibre in the skin but several fibres because the energy applied to the skin will move throughout nearby skin tissues [128] [146]. Within most of the vibrotactile literature, the fibres are grouped into two systems: the Pacinian and the non-Pacinian systems.

"*The Pacinian system has a large receptive field excited by higher frequencies and the non-Pacinian system consists of a small receptive field thought to be excited by lower frequencies*" [129]. Bolanowski *et al.* [9] found threshold sensitivities in the range of 0.4 to 500 Hz between these two systems. The Pacinian system exhibited a U-shaped function at higher frequencies (40 to 500 Hz) where maximum sensitivity occurred between 250 and 300 Hz [9]. Therefore, the majority of the stimuli used in this research have a frequency of 250Hz. Verrillo [149] also reported a similar function for hairy skin, where maximum sensitivity occurred at 220 Hz.

Understanding the features of specific skin fibres and their response characteristics when stimulated can help to inform the design of any tactile feedback to ensure that the stimuli are compatible with the characteristics of the skin structures on which the feedback will be presented. According to Kandel and Jessell [74], Meissner's corpuscles and Merkel's cells respond to touch, Pacinian corpuscles respond to vibration, and Ruffini's corpuscles respond to rapid indentation of the skin. Vibration is detected best on hairy, bony skin and is more difficult to detect on soft, fleshy areas of the body [50].

The dimensions or attributes of our sense of touch are detailed below:

2.2.1.1 Frequency

As mentioned in Section 2.1.1, humans can hear sounds in the range 20-20,000Hz; however, the frequency range of the skin is much smaller, ranging from 10Hz to 400Hz, with maximum sensitivity [136] and finer spatial discrimination at around 250Hz [36]. Investigations by Goff involving the stimulation of the subject's finger with a single probe showed that for lower frequencies (< 25Hz), the discrimination threshold was less than 5Hz. For frequencies greater than 320 Hz, discrimination capacities were also degraded [53]. Measures for discrimination thresholds of frequency are problematic, as perception of vibratory pitch is dependent not just on frequency, but also on the amplitude of stimulation. Geldard [48] found that subjects reported a change in pitch when frequency was fixed, but amplitude of stimulation was changed. Sherrick [127] found that combining frequency and amplitude redundantly allowed a greater number of identifiable levels to be created. He found that people could distinguish three to five different levels of frequency, but that adding amplitude as a redundant parameter could increase this range. Therefore, this interaction between frequency and amplitude should to be taken into account or perhaps avoided when designing tactile stimuli.

2.2.1.2 Duration

Geldard [48] reports that the temporal duration just noticeable difference (JND) rose from 50 to 150 ms. when duration was increased from 0.1 to 2.0 seconds Gescheider (as reported in [139]) measured the time difference between two tactile "clicks" on the fingertip, necessary for them to perceived as two separate sensations and found that the minimum threshold reported was 10 ms. Interactions between duration and perceived amplitude should be considered when using duration as it has been shown that short intense signals can be confused with longer, lower intensity signals. Gunther [54] suggests that stimuli lasting less than 0.1 seconds may be perceived as taps or jabs, whereas longer stimuli may be perceived as smoothly flowing tactile phrases. Craig and Sherrick [36] warn that very short durations may result in sensations such as pokes or jabs, which might be undesirable.

2.2.1.3 Rhythm

Rhythms are created by grouping together pulses to create temporal patterns in a similar fashion to rhythms in music. Rhythm is very important and useful in the design of tactile systems. For example, Summers [135] encoded speech information by modulating vibration frequency and amplitude, and by presenting the temporal pattern of the speech using rhythm. The results of an evaluation showed that users obtained the most information from the rhythmic pattern compared to the frequency/amplitude modulation.

2.2.1.4 Location on the Body

As far as our spatial senses go, touch comes in second after vision [81]. Different body locations have different levels of sensitivity and spatial acuity. The most sensitive part of the human body is the fingertip. When applying tactile stimuli to multiple points on the body, the distance between points is extremely important. Two-point discrimination is a measure that represents how far apart two pressure points must be before they are perceived as two distinct points on the skin [50]. The point of contact discrimination threshold for two points is 0.9mm when the stimuli are placed against the subject's finger in the absence of any movement lateral to the skin's surface. It is not possible for two points of contact closer than this threshold to be distinguished as separate stimuli. Experimental evidence suggests *"active exploration marginally increases sensitivity, decreasing the threshold to 0.7 mm"* [107].

It must be noted that there is some controversy surrounding the two-point method. It has been stated that there are several problems with the method including setting appropriate criteria and the fact that many studies have shown that participants were able to discriminate two points at much shorter distances than the two-point threshold [35].

An alternative to the two-point method is called grating orientation where participants are presented with a grating made up of alternating grooves and ridges. The grating can be presented in two different orientations at right angles to each other and the participant must identify the orientation. This method uses stimuli with identical spatial structures; only the width of the grooves and ridges is varied [88]. Using the grating orientation method, Johnson and Phillips [69] measured the sensitivity of the index finger to square wave gratings showing that discrimination improves for gratings greater than 1mm. Discrimination rates of 75% or higher were achieved for gratings of 2.25mm, 1.84mm and 1.68mm.

The body sites involved in tactile parameter estimation in the literature are also those areas of the body that have been identified as most sensitive to pressure and stimulus discrimination:

- Finger, [31]; [114]
- Hand, [9]; [31];
- Arm, [30]; [149];
- Thigh, [31];
- Torso, [29] [147].

Cholewiak, Brill, and Schwab [29] investigated the vibrotactile localisation accuracy for the abdomen using 12, 8, and 6 equidistant actuators, 72 mm, 107 mm, and 140 mm, respectively. Their results showed that the ability to correctly identify which actuator was presenting a stimulus increased as the number of actuators decreased. Study participants were correct in their identification for an average of 74%, 92%, and 97% of the trials for 12, 8, and 6 actuators, respectively. The results also showed that when participants labelled areas on their abdomen, for example the navel at 12 o'clock and the spine at 6 o'clock, they were better able to localise stimuli. Accuracy rates were much lower when labels were not available. This suggests that accuracy can be increased if a label is provided which is mapped to the locations to be identified.

2.2.1.5 Intensity

As indicated in Section 2.1, our sense of hearing is capable of processing a large range of intensities (or amplitudes): up to 130dB above the detection threshold. It is also capable of discriminating small differences at 115dB above the detection threshold. On the other hand, our sense of touch is much more limited, with an intensity range of approximately

55dB above the detection threshold. Any vibrations above this threshold feel unpleasant or even painful [150].

2.2.2 Encoding Strategies

Geldard was one of the earliest researchers to investigate the possibilities of using the skin to communicate messages [49] stating, "*for some kinds of messages the skin offers a valuable supplement to ears and eyes.*" He outlines the basic steps needed to build a cutaneous language with a focus on stimulus properties and the mechanical dimensions for encoding information.

2.2.2.1 Tactons

Tactons [20] are used as the vibrotactile counterparts of earcons in the design of crossmodal icons. These are structured vibrotactile messages which can be used to communicate information non-visually. They are the tactile equivalent of earcons and visual icons, and could be used for communication in situations where vision is overloaded, restricted or unavailable. Tactons are created by manipulating the parameters or dimensions of cutaneous perception (like those detailed above) to encode information. The concept of using tactile parameters to encode information when designing tactons is based on Geldard's notion of mechanical dimensions. The most important dimensions (or parameters) are detailed below:

Locus: the body is a large area on which tactile actuators can be placed making locus (or spatial location) an important consideration. In his lab study, Geldard found that participants could reach levels of 100% recognition using seven actuators placed on the rib cage and the same results for five actuators on the chest. One issue that should be taken into account is the fact that, with standard vibrotactile actuators, the vibration emanates across the body and is not simply confined to underneath the actuator. Furthermore, when two or more actuators are activated simultaneously it can often feel as though there is only one actuator.

The waist has been used as a body location for presenting tactile feedback in many research applications including waypoint navigation as demonstrated by van Erp *et al.* [147]. The authors conducted two experiments to investigate whether navigational information can be encoded in a tactile display. The eight vibrotactile actuators were attached to a belt positioned on the waist of participants. The tactile display was used to encode information on direction and distance.

The first experiment was conducted with 12 participants navigating a route outdoors in a field; distance was encoded in rhythm and direction in vibration location. It was found that mapping waypoint direction on the location of vibration is an effective coding scheme that requires no training, but that coding for distance (increasing intensity or rate of the rhythm was mapped to decreasing distance) does not improve performance compared to a control condition with no distance information.

The tactile modality has also been combined with audio using spatial location in order to produce musical compositions. Gunther *et al.* [54] introduced the notion of *tactile composition*. The authors created a system that facilitates the composition and perception of intricate musically structured spatio-temporal patterns of vibration on the surface of the body. Thirteen vibrotactile actuators were placed on the body with three on each limb and one on the lower back. An initial test of the system was conducted in a performance context which found that the body locations were suitable for presentation of tactile music and that music can be composed for the sense of touch.

Psychophysical studies have demonstrated that using anatomical points of reference when positioning the tactile display enhances localisation accuracy [31] [147]. Given the range of body sizes that these displays can be mounted on, it is important to determine whether it is better to use the available sensory area by adjusting inter-actuator distances to cover the skin surface or to maintain the same dimensions of the display for all users.

Cholewiak *et al.* [32] measured vibrotactile localisation on the forearm and abdomen to investigate the spatial resolution and information transfer abilities for vibratory stimuli. Experiments showed that stimulus frequency did not affect localisation on the arm, but when placed both on the arm and trunk and presented with vibrotactile patterns, fewer than half of the sites were uniquely identified.

Craig and Sherrick [36] suggest the back, thigh and abdomen as suitable body locations. Cholewiak and Collins [30] investigated tactile spatial locations using seven actuators on the forearm. When a stimulus was close to an anatomical reference point, and in particular a point of mobility such as the wrist or elbow, the authors found that higher levels of performance were achieved. Cholewiak *et al.* [29] conducted a study on the abdomen, where the main anatomical references are the spine and navel, and found again that location identification was most accurate when the stimuli was presented at these reference points. They also found that people were less likely to confuse stimulation at another point for stimulation at one of these reference points.

Tactons can be presented to different locations on the body. Brown *et al.* [20] used three locations on the forearm. It is suggested that for accurate localisation of three locations, two actuators should be located at anatomical reference points, with the remaining actuator located at a point between these two. The arm is a practical location on which the reference points (wrist and elbow) should be accurately localised, and the third point should not be confused with either of the two reference points.

Intensity: Geldard [49] states that, when using intensity as a parameter, the stimulus can be between 50 and 400 microns (1 micron is 1⁄1000000 m) but should not be more than 400 microns as this causes discomfort to the user. In lab-based studies it was found that participants could distinguish fifteen different levels of intensity within the 50 – 400 microns range but realistically, Geldard recommends that three levels should be used. In terms of decibels, the intensity range of the skin reaches about 55 dB above the threshold of detection, beyond which vibrations may become unpleasant or painful [150].

Guidelines by van Erp [142] already showed that observing the absolute intensity of a vibration signal is difficult; however users are able to observe changes in intensity. Brown *et al.* [22] successfully made use of intensity change over time as a tactile parameter in their investigation into the possibilities of applying musical techniques to tactile icon design. Tactile versions of musical dynamics were created by manipulating the amplitude of vibrations to create increasing, decreasing, and level stimuli and an experiment was carried out to test perception of these stimuli. Identification rates of 92%-100% indicate that these tactile dynamics (namely increasing and decreasing intensity) can be identified and distinguished from each other.

As mentioned, in earcons, amplitude/intensity is not used as a parameter because users find loud sounds annoying and report annoyance when the volume level is out of their control [18]. Using intensity as a parameter in tactons is equally problematic as reducing the amplitude could degrade perception of other parameters, or render the signal undetectable, while increasing it too far could cause pain [49]. Therefore, it is best to leave amplitude under the control of the user instead of using it to encode information.

Duration: The duration parameter explored by Geldard [49] is an extremely effective dimension. In his study, durations ranging between 0.1 and 2 seconds were used. 100% identification rates were achieved when 4 or 5 levels with intervals of at least 0.15 seconds were used. Vibrotactile stimuli lasting less than 0.1 seconds are perceived as taps or jabs against the skin [54]. Differences in duration enable rhythmic structures to be created. However, it must be noted that stimulating an area of skin for an extended period of time can result in adaptation or even pain.

Frequency: as for tactile frequency, unfortunately humans cannot literally 'hear through the skin' as the detectable frequency ranges for each modality are different (although with some overlap). Using frequency as a parameter has been difficult in experiments with issues rising from its influence on intensity perception. Reports on the frequency discrimination abilities of the skin are dependent on the experimental paradigm and tend to vary somewhat [54]. Sherrick proposes that the results suggest that between 3 and 5 values of vibration rate can be distinguished between 2 and 300 pulses per second [127]. Rovan and Hayward report that ranges broadly divided into 8 to 10 discrete steps are perceptible over a range of 70 to 800 Hz [120].

Frequency has yet to be used as a parameter in tacton research but MacLean and Enriquez [92] used multidimensional scaling techniques to determine how haptic icons can be created from signal parameters such as waveform, frequency, and force. They found that for the ranges of parameters that they implemented in a handheld knob, frequency played a dominant role in distinguishing between the multidimensional stimuli and that waveform and force were less salient.

Given the range of conflicting results in terms of frequency as can be seen in the examples above, it appears as though further research is required in this area. As mentioned in Section 2.2.1.1, there is a perceptual interaction between frequency and amplitude. This may be a contributing factor to the varying levels of success achieved when using frequency as a parameter.

Waveform: Geldard [49] suggests that it may be possible to distinguish between tactile waveforms provided the frequency of the stimuli is low and does not interfere. In musical composition studies, it has been suggested that waveform can be correlated to the "texture" of tactile stimuli [54].

Four different stimuli varying in roughness can be created in tactons using: a 250Hz sine wave, a 250Hz sine wave modulated by a 20Hz sine wave, a 250Hz sine wave modulated by a 40Hz sine wave, and a 250Hz sine wave modulated by a 50Hz sine wave, or a 250Hz sine wave, a 250Hz sine wave modulated by a 30Hz sine wave, a 250Hz sine wave modulated by a 40Hz sine wave, and a 250Hz sine wave modulated by a 50Hz sine wave sine.

Rhythm: Rhythm is an extremely important parameter in earcon design [18] and is the primary parameter used in tactons with recognition rates of over 90% achieved when three different rhythms are used [21]. Rhythms can be created by grouping together pulses of different durations. The rhythms used in tactons are based on Brewster's guidelines for rhythms in earcons [18]. In order to make each rhythm feel as different as possible a different number of notes (pulses) are used in each rhythm. In addition to following Brewster's guidelines, these rhythms also follow advice given by van Erp and Spapé [145] who identified tempo (speed) as an important parameter in the identification of tactile melodies. Therefore, all rhythms used in tacton design are created using the same tempo.

2.2.2.2 Other Haptic Icons

Another approach to developing tactile or haptic icons involves identifying the basic elements, called haptic phonemes, and using these to create different haptic icons. With this method, Enriquez, MacLean and Chita [42] created a set of nine haptic icons that varied in terms of waveform and frequency. They then trained participants to associate each haptic icon with an arbitrary concept, such as the name of a fruit. They found that participants learned these associations after about 25 minutes of training and achieved higher identification rates with stimuli that varied in frequency (81% correct), compared to those that varied in waveform (73% correct).

Rovers and Essen [122] also mention the use of icons with haptic feedback. They state that the message can be designed as a real-world signal such as a heartbeat or can be based on an abstract design. An abstract design requires the use of a set of common rules for example, 3 pulses is equal to 'off'. In this case, variability can be represented in glyphs for example, changing intensity based on running speed: the faster the speed, the higher the intensity.

This section of the review has established that tactile feedback, namely vibrotactile feedback, can be used successfully to encode data and that there are several extremely effective parameters that can be used. The parameters examined in this review will be used as a basis for the design of the tactile crossmodal feedback used in this research. The next step, detailed in Section 2.2.3, investigates currently available hardware with a view to establishing the most effective platform and actuators for use in crossmodal interaction.

2.2.3 Hardware

Tactile devices generally appeal to the cutaneous senses by skin indentation, vibration, skin stretch and electrical stimulation [15]. A number of tactile stimulation devices are available, each of which stimulates a specific tactile response. These include pressure, thermal, slip, electrocutaneous and vibration displays. Vibrotactile actuators were chosen as one of the types of hardware in this research for a number of reasons: firstly, vibration devices are generally easiest to work with and in particular, to control; secondly, the work in this thesis is aimed at mobile interaction and most mobile devices already include a vibrotactile actuator and lastly audio feedback is also, in simple terms, a vibration. This should aid in the crossmodal design of audio and tactile displays if both are based on vibrations (see Chapter 3). Vibrotactile actuators can provide sustained feedback and allow many different textures to be presented. By using the actuator already in commercial devices, the tactile feedback is not restricted by expensive or rare technology and does not require any hardware to be added to the device which could increase its size or weight which may be inappropriate for mobile devices.

Most vibrotactile actuators use electromagnetic actuation to drive a mass in either a linear or rotational manner to stimulate the skin. The main vibrotactile actuator used in this research, the EAI C2 Tactor[3], is shown in Figure 2-2. This device is resonant at 250Hz with much reduced response at other frequencies (which is another reason for the reduced usefulness of frequency as a parameter for vibrotactile interfaces). The advantage of vibrotactile cues is that they can exert high levels of force (so can be felt through clothing) and they can also be distributed over the body to give spatial cues (often attached to a user's belt around the waist). For a more detailed review of vibrotactile devices see Summers [136].

---

[3] www.eai.com

Figure 2-2: Engineering Acoustics Inc (EAI) C2 vibrotactile actuator.

The other type of actuator used towards the end of this research in Chapter 7 involves the use of piezo-electric feedback. Piezo-electric actuators (Figure 2-3) can create short more display-localised tactile bursts, by moving touch screen display modules within the device [79]. The piezo-electric actuator is also able to generate quick pulses and the tactile feedback is concentrated to move the display mass, which is commonly 20% of the whole device mass, providing large displacement with rapid responses, but with less kinetic energy compared to traditional vibration motor systems.

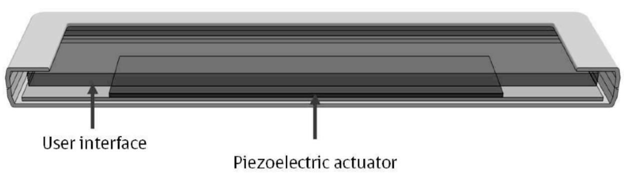

Figure 2-3: Touchscreen device with integrated piezo-electric actuator.

Koskinen *et al.* [78] ran three laboratory-based studies to determine which tactile click (from a set of various different designs) is most pleasant to use in fingertip interaction with a mobile touchscreen device. Using two different types of actuator: piezo or vibration motor, the experiments allowed the authors to find the most pleasant tactile feedback as perceived by participants. The results show that feedback from piezo-electric actuators is perceived as more pleasant than feedback from vibrotactile motors.

2.2.4 Key Applications Using the Tactile Modality

Vibrotactile feedback is already very common within video game systems and handheld controllers. Actuators are often used to provide feedback representing weapon fire or environmental effects. There is commercial interest in this area too, as most mobile telephones include tactile feedback to accompany ring tones. For example, Immersion's

VibeTonz[4] attempt to extend this simple feedback to enhance games and ring tones. Vibrotactile displays have been incorporated into canes used by visually impaired people. The UltraCane[5] uses ultrasound to detect objects in a user's environment and presents the location and distance to targets by vibrating pads on the handle of the cane.

Originally, work on vibrotactile displays was driven by tactile-audio substitution for profoundly deaf people, and was developed in the late 1970s and early 1980s. One of the earliest devices was the Tacticon, a commercial device that adjusted the perceived intensity of 16 electrodes, each of which corresponded to a range of frequencies in the auditory spectrum, in order to improve speech comprehension, auditory discrimination and the clarity of the users speech [73]. For a full review of work in this area see Summers [136].

Cockburn and Brewster [34] looked at combinations of different feedback modalities, including vibration feedback from a Logitech iFeel vibrotactile mouse[6], for selecting small targets on a computer desktop. They found that, in simple Fitts' law type tasks (where discrete targets are used, so there are no distracters), tactile and audio feedback both reduced targeting time (confirming Akamatsu's results [3]), but the combination of audio plus tactile was not as good as when each was used alone. However, in a more realistic task (choosing items from drop down menus) the tactile feedback caused problems and actually increased targeting time over a standard graphical display. The reason for this was that the close proximity of many tactile targets caused a feedback overload.

Jacko and colleagues have looked at how tactile displays (and more generally multimodal ones) can help older adults with age-related macular degeneration (AMD) (which is a leading cause of visual impairment in individuals of 65 years and over). Their evaluations use drag-and-drop type interactions with the Logitech Wingman force- feedback mouse[7] which vibrates to produce tactile feedback. When different combinations of audio, tactile and visual feedback were added to drag-and-drop, there was little benefit from the tactile feedback over a standard visual display, except when it was in combination with audio [68]. Results from this work appear to conflict with those of Cockburn and Brewster [34] as they showed audio and tactile feedback were more beneficial on their own. However, it is difficult to compare the two studies as different users, devices and stimuli were used.

---

[4] www.immersion.com/vibetonz

[5] www.soundforesight.co.uk

[6] www.logitech.com

[7] www.logitech.com

As described earlier, Brewster and Brown [14] have investigated an alternative encoded form of tactile presentation: the tacton, or tactile icon. Tactons have been implemented as alerts on handheld devices such as pagers and mobile phones. Brown and Kaaresoja [72] evaluated nine tactons that were used to communicate the type of alert (voice call, text message, or multimedia message, represented by rhythm) and the priority of the alert (low, medium, or high, represented by roughness or intensity). Overall recognition rates of 72% were achieved.

Tactons can also be employed in the context of large-scale supervisory control environments. For example, Hameed *et al.* [56] developed an interface to support water control engineers in task scheduling and prioritisation. They encoded the nature, urgency, and duration of a pending task by mapping this information to the spatial location, frequency, and duration of a tactile signal, respectively. The information encoded in these signals was correctly identified by participants in an experiment with rates of 94%, 90%, and 83%, respectively. It was found that, by using the information about a pending task from the tactons, participants were able to make more informed and appropriate decisions regarding attention switching compared to traditional interruption cues.

Other research has shown that complex tactile signals are feasible and useful but that their success is highly context dependent. Chan *et al.* [25] found that seven haptic icons could easily be learned in the absence of workload and with minimal training. The authors also demonstrated that an increase in workload resulted in detection times that were significantly longer but still acceptable in most task contexts. The specific designs of the different icons used in this research did, however, influence their susceptibility to workload effects.

2.2.5 Summary

This section of the review has summarised the basics of the sense of touch focusing on human perception. More specifically, this research focuses on the tactile aspect of haptics as opposed to kinaesthetic or proprioceptive aspects.

There are many dimensions of the tactile modality and these have influenced encoding strategies such as Braille, tactons and Haptic Icons. Haptic phonemes are used in the design of Haptic Icons with variations in waveform and frequency. However, in tactons research, frequency has been shown to be a poor parameter in which to encode data. The most effective parameters in tacton design are rhythm, roughness and spatial location. A small amount of research has been conducted investigating the use of intensity and waveform as parameters with results showing that they also have the potential to become useful tacton parameters.

The parameters used when encoding data in the tactile modality are not only dependent on human perception but also the capabilities of the vibrotactile actuators used to produce the vibrations. Therefore, a brief overview of the hardware used in this research was included.

Lastly, the review of key applications show that the tactile modality is a viable modality of communication, and like earcons, tacton parameter design is based on the basic dimensions of touch which in turn are very similar to the basic dimensions of our sense of hearing.

2.3 Audio and Tactile Touchscreen Applications

The purpose of the research in this thesis is to use both the audio and tactile modalities together on touchscreen devices. Much of the current research investigates only one modality at a time. Some related research, which does make use of both audio and tactile feedback, is detailed below along with recent research on the use of mobile touchscreen devices (the chosen platform for this thesis work).

2.3.1 Using Touchscreens

Recent research has focused on the technology used to provide the feedback from touchscreens on mobile devices. Poupyrev *et al.* [112] propose using a piezo element stack to provide tactile feedback for PDAs. Fukumoto *et al.* [44] propose using voice coils to provide tactile feedback of button pushes and found an increase in dialling speed when compared to using audible beeps for button push feedback.

The use of touch screens in mobile devices is a logical step as they have several advantages over many other pointing devices. Shneiderman outlines some of the features of touch screen interaction in his paper ‘Touchscreens now Offer Compelling Uses’ [131]:

- Touching a visual display of choices requires little thinking and is a form of direct manipulation that is easy to learn;
- Touchscreens are the fastest pointing device;
- Touchscreens make hand-eye coordination easier than mice or keyboards.

At the same time, there are also some limitations in interaction with these touchscreens:

- Users' hands may obscure the screen;
- They cost more than alternative devices;
- On a small display it can be difficult for user's to select a specific target with no physical feedback.

Shneiderman suggests using a visually appealing metaphor that reacts predictably. When touching a target in the real world, not only would a user expect to feel the target when it is selected but would expect to experience different tactile sensations when the widget is not activated or partially activated. Furthermore, current telephone keypads provide orientation information by making certain keys, for example '5', feel different compared to the other buttons. These are not features of current touchscreen interface designs. By providing audio and tactile feedback to a touchscreen keyboard, users could feel and hear the widgets they are interacting with.

2.3.2 Unimodal Touchscreen Feedback

There has been research into the addition of single modalities to touchscreen output. Nashel and Razzaque [103] added tactile cues simulating real buttons to virtual buttons displayed on mobile devices with touch screens. Some existing techniques give audio feedback for a button "click" or press but they are not designed to provide the user with information regarding button location or any errors such as slips. Nashel and Razzaque describe their system which provides tactile feedback representing the button location and activation when the user's finger is on the display. As the user's finger moves over a virtual button: a 'pop' is presented as the finger enters a button region; a low amplitude vibration is presented when the finger pushes the button; a short pulse is presented as the finger leaves the button area and no feedback when the finger is between buttons.

The experiments conducted found that all participants were able to differentiate between vibration (finger over a button) and no vibration (finger not over any button). Some were not able to differentiate between the vibration frequencies used for each row of buttons, but most users were able to differentiate between the rows of buttons by touch alone.

Brewster *et al.* [16] designed sonically enhanced graphical buttons using earcons for the audio feedback. Timing, error rates and workload measures were used. Error recovery was significantly faster and required fewer keystrokes with the sonically enhanced buttons than with standard ones. The workload analyses showed participants significantly preferred the sonically enhanced buttons to standard ones.

Once again, these research projects employed a unimodal approach using only the tactile modality or the audio modality alone. Crossmodal feedback may be able to provide a greater amount of feedback using different combinations of tactile and audio. Crossmodal feedback will also allow the user to choose whatever modality is most appropriate given their situation or preference.

Kaaresoja *et al.* [72] presented a touchscreen mobile device augmented with piezo-electric tactile feedback. The actuators are positioned under a resistive touchscreen, and can provide tactile feedback to a stylus or finger. The authors suggest four applications for the touchscreen tactile feedback: numerical keypad, text selection, scrolling, and drag and drop. When using the numerical keypad, button clicks change the colour of the button and a tactile click is presented; when the button is released the colour changes back to the original colour and a second tactile click is presented much like physical interaction with traditional buttons. In text selection tasks, a gentle tactile click is represented as each character is selected. Or, if the text is selected line by line a stronger click is presented. The intensity or amplitude is mapped to the amount of text selected. Scrolling also produces tactile feedback mapped to the scrolling speed. Lastly, when using drag and drop functionality, users receive several occurrences of tactile feedback, for example: when an item is picked up, dragged, and the item is dropped into a folder or application.

There was no formal evaluation of the augmented touchscreen; therefore few conclusions can be made. Further studies incorporating audio feedback and crossmodal parameters as described in Chapters 4 and 7 could perhaps show whether this is an appropriate style of device for use in crossmodal interaction.

Brewster *et al.* [10] presented an initial basic study investigating the use of tactile feedback for touchscreen keyboards on PDAs. The tactile feedback added to the standard touchscreen buttons was made up of simple tactons [20]. The design used two stimuli: one to indicate a successful button press and one to indicate an error. The success tacton was played when a button was correctly pressed and then released. The error tacton was played when a slip or double tap error occurred. The tactons were made using an 800 ms, 250Hz sine wave success cue, and a rough (amplitude modulated) sine wave for the error cue.

A laboratory study was conducted to compare standard buttons to buttons with additional tactile feedback. Results showed that with tactile feedback users entered significantly more text, made fewer errors and corrected more of the errors they did make. The study was also conducted with users seated on an underground train to see if the positive effects transferred to realistic use. There were fewer beneficial effects, with only the number of errors corrected significantly improved by the tactile feedback. The study described by Brewster *et al.* was very small and only made use of the tactile modality not the audio modality. To fully understand the effects of crossmodal feedback an evaluation of a real physical keyboard and one with artificial crossmodal audio and tactile feedback is needed.

A similar study was conducted to investigate the addition of audio feedback to touchscreen buttons. In his paper titled 'Overcoming the Lack of Screen Space on Mobile Computers', Brewster [11] describes a small pilot study and two formal experiments that investigate the usability of sonically enhanced buttons of different sizes. The underlying hypothesis being that presenting information about the buttons in sound would increase their usability and allow their size to be reduced. An experimental calculator-style interface was created and the buttons of the calculator used a range of different types of sound from basic to complex. Results showed that more data could be entered with sonically enhanced buttons and subjective workload was reduced. More sophisticated sounds that encoded more information about the buttons were shown to be more effective than the basic PDA sounds. Results also showed that when a mobile device was used in a realistic situation (whilst walking outside) the usability was significantly reduced than when used in a lab setting.

These two studies have separately investigated the addition of audio and tactile feedback to touchscreen buttons. Both studies have produced successful results for each modality showing that performance in typing tasks can be improved with the addition of such feedback. Therefore a touchscreen keyboard designed with crossmodal audio and tactile feedback appears to be viable and should result in a touchscreen keyboard which is usable despite situational impairments.

Lee *et al.* [82] created a system for providing tactile feedback for stylus-based touchscreen displays called the Haptic Pen (Figure 2-4). The Haptic Pen provides personal tactile feedback for multiple simultaneous users and can operate on large touchscreens as well as ordinary surfaces. A pressure-sensitive stylus was combined with a small solenoid to generate a range of different tactile sensations.

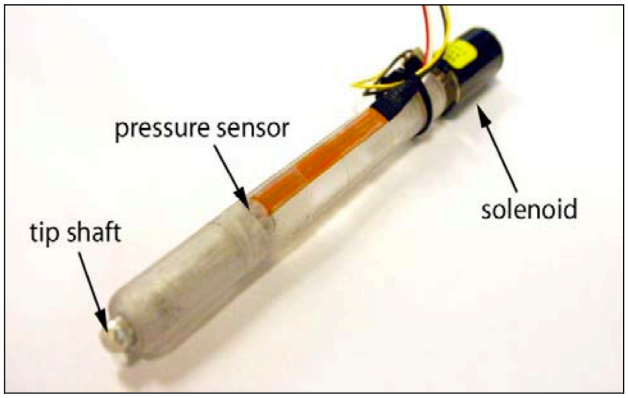

Figure 2-4: Haptic Pen Prototype.

Responses to informal usage experience interviews indicated a high degree of believability in the tactile simulations generated by the Haptic Pen (the feedback when pressing a button appeared to feel realistic). The tactile feedback was intended to simulate the sensation of pressing a physical button or dragging a physical object. There was no formal study of the Haptic Pen so there can be few conclusions drawn but the initial findings indicate that, once again, the use of tactile feedback is an effective approach to simulating the sensation of pressing a physical button. Furthermore, the tactile feedback does not necessarily have to originate from the screen itself but can be incorporated into a stylus. This thesis mainly concentrates on fingertip interaction as opposed to stylus interaction but the design principles behind the Haptic Pen could be transferred to fingertip interaction and there may be a crossmodal audio equivalent to the sensations produced by the Haptic Pen.

Audio feedback has also been combined with gesture input in mobile music players [109]. The non-speech audio feedback (Earcons [8]) allowed users to control the music player without having to look at the screen. Experiments also showed significant usability improvements for the gesture/audio-based interface over a standard visual/pen-based display. An equivalent study using tactile feedback could provide insight into the potential of crossmodal feedback for gesture input. It could also be said that fingertip interaction is a basic form of gesture - an onscreen gesture.

Some of the first researchers to investigate the use of tactile feedback on touchscreen devices were Poupyrev and his colleagues [112]. The TouchEngine™ – a thin, miniature low-power tactile actuator designed specifically for use in mobile interfaces – was embedded in a PDA. The TouchEngine is a piezo-electric actuator that bends when a signal is applied (Figure 2-5). Unlike the actuators in Active Click [44], by using piezo-electric actuators, the device could provide localised feedback to the fingertip instead of vibrating the whole device. In this case, Poupyrev *et al.* used the tactile feedback as an ambient background channel of information. The authors investigated several applications using touch as the ambient, background channel for mobile communication and conducted a formal user study into the use of tactile feedback with tilting devices. Participants were required to scroll through a text list using gestures. The results of the study showed that, on average, participants could complete the tasks 22% faster when provided with tactile feedback.

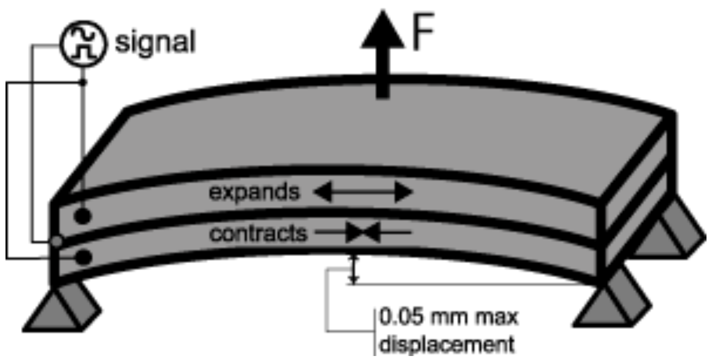

Figure 2-5: Piezo-electric bending motor actuator used in TouchEngine[8].

2.3.3 Multimodal Touchscreen Feedback

Sharmin *et al.* [126] created a stylus with an embedded vibrotactile actuator for use on touchscreen displays with an aim to provide tactile representations of graphical information for visually impaired users. The authors conducted a pilot study to compare performance with audio or tactile feedback when following a graphical trail (Figure 2-6).

---

[8] from http://ivanpoupyrev.com/projects/tactile.php

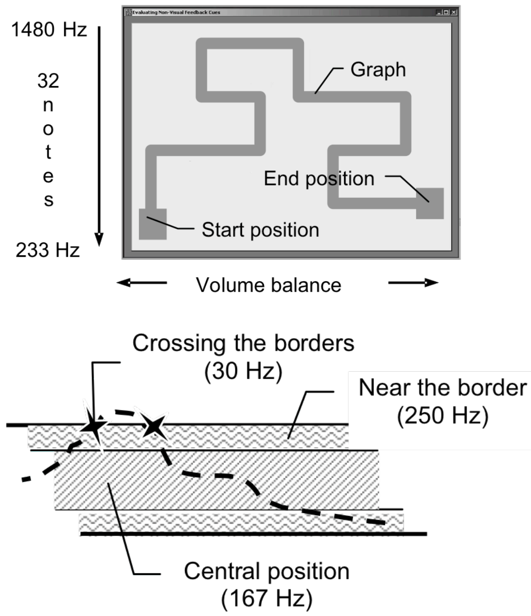

Figure 2-6: Graphical trail used in experiment.

The tactile frequency produced by the actuator varied at different points on the line. For instance, crossing the borders of the trail was indicated by a 30 Hz tactile cue, central positions on the trail were represented by 167Hz and when approaching the border and veering off the centre, 250Hz was used. The audio version also used frequency deviation through the use of different pitch intervals.

Participants were asked to follow the graphical trail using the stylus with audio or tactile feedback. The results show that the speed of following a graphical trail was significantly higher when using tactile feedback compared to audio feedback. Almost all the participants preferred tactile feedback to sound feedback at the end of the study. There was no significant difference between the error rates in the audio or tactile conditions.

The results indicate that the tactile modality was significantly faster than the audio modality. Unfortunately, there was no control condition in the experiment i.e. visual feedback only. Therefore it cannot be determined whether or not audio feedback was still significantly faster than visual feedback alone despite being slower than the tactile condition. However, the results so far are promising and show implementations of both types of modality feedback for touchscreen interaction.

Although this research indicates potential in the tactile modality for use with touchscreens, it focuses on the use of a stylus. The research described in this thesis revolves around fingertip interaction. A stylus can be awkward to use and easy to lose. Styli were originally designed for high precision applications such as drawing, not for everyday use. They also break the metaphor of direct manipulation [131] which is one of the most important features of touchscreen interaction.

Although not specifically crossmodal, one of the few research applications that includes the use of audio and tactile is ComTouch, a sensory augmentation tool, by Chang *et al*. [27]. ComTouch (Figure 2-7) is a device that augments remote voice communication with touch, by converting hand pressure into vibration intensity between users in real-time. It is a vibrotactile sleeve using small commercial acoustic speakers to transmit vibrations that can be fitted over the back of a mobile phone. The overall aim of this research was to assess the potential of creating a tactile language.

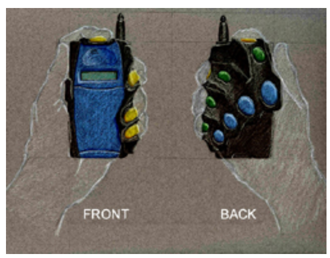

Figure 2-7: Concept drawing of ComTouch[9].

A study was conducted using ComTouch to investigate the possible uses of the tactile channel when used in conjunction with audio and to test the mapping between pressure and vibration. Pairs of participants had to complete two tasks: a chatting task and a desert survival problem. The chatting task required participants to have a conversation for five minutes (with access to the audio channel too) and the other task gave the participants a context in which to use the device; they are stranded in the desert and need to get to safety together (with limited access to the audio channel). By recording and examining both audio and tactile data, the authors found strong relationships between the two modalities when used as communication channels.

Touch communication was shown to enhance an audio conversation by providing redundant and independent information in the form of tactile gestures. This allows communication of nonverbal cues that can be lost or overlooked when only the audio channel is present. The results showed that users developed an encoding system similar to that of Morse code, as well as three original uses: emphasis, mimicry, and turn-taking.

[9] http://tangible.media.mit.edu/projects/comtouch/

In this case, the audio feedback came in the form of speech. Furthermore, intensity was used as a tactile parameter yet in studies of tactons [21] it has been shown to be an ineffective design parameter. The participants created their own language throughout the experiment. The crossmodal icons described in this thesis (Chapter 3) use a predefined language that depends on specific mappings between modalities and information and must be learned by the user.

Chang and O'Sullivan [28] are some of the small number of researchers who have used both the audio and tactile modalities in a mobile device. In the most basic terms, tactile feedback is added to enhance the audio feedback in a standard mobile device. The authors argue that by using integrated stimulation of the five basic senses, the sense of cognition is engaged more fully. The authors present techniques for audio manipulation to create simple vibrotactile feedback based on the fact that both the audio and tactile modalities are made up of vibrations. A filter is applied to split the sound into its constituent parts, i.e. vibrotactile and audio. In this case, any frequencies under 300Hz were amplified and presented through the tactile actuators. Frequencies over this level were presented through audio.

Although the tactile feedback in this case is used purely as an enhancement to the audio modality, the crossmodal similarities between the modalities are exploited through the use of frequency. In this thesis it is proposed that crossmodal similarities such as frequency can be used to allow users to switch between modalities.

After Ambient Touch, Poupyrev and Maruyama presented another design, implementation, and informal evaluation of a piezo-electric tactile interface for small mobile touchscreens [111]. Once again, a PDA was augmented with four embedded custom-designed TouchEngine piezo-electric actuators. Poupyrev and Maruyama classified all tactile feedback for touchscreen interaction into five basic types: tactile feedback provided when the user starts a gesture by touching a GUI element, when the user then either drags or holds the pen/finger, and, finally, when the user lifts it off either inside or outside the GUI widget. In their prototype design, Poupyrev and Maruyama augmented basic GUI elements with tactile feedback, including several variations of buttons, scroll bars and menus. The parameters of amplitude and frequency were manipulated to create the different sets of tactile feedback.

Through informal testing using three conditions: visual feedback only, audio feedback and tactile feedback, it was found that tactile feedback was enjoyed by users who stated that the sensations felt very realistic.

This study is of benefit to crossmodal research as it compared the audio and tactile modalities in a crossmodal manner i.e. both modalities were used separately but to present the same information. However, the feedback design was not produced using crossmodal parameters (see Chapter 4 for details) and the study was informal.

Audio and tactile feedback are often used together in interfaces to complement each other. As in nature, where we can hear the 'thud' of an object falling on our foot at the same time as we feel it. Williamson *et al.* [155] created an interface for sensing data within a mobile device called Shoogle. It is based around active exploration: the user can shake the device to feel or hear the contents moving "inside". The system uses both tactile feedback and audio feedback through impact sonification to present the information. For example, users could shake the device to feel and hear how many messages they have. This is a form of redundant crossmodal feedback.

2.4 Discussion

This chapter has presented a review of audio and tactile perception with particular focus on the parameters and current methods of encoding information in each modality. In addition, it has reviewed existing applications of audio and tactile feedback to touchscreen applications. This section discusses how these findings can be applied to the area of crossmodal interaction.

Research Question 1 will be answered in detail in Chapter 3, where the suitability of established audio and tactile parameters for use in crossmodal feedback is discussed, and in later chapters where the results of evaluations of these parameters are presented. However, from the information on audio and tactile perception presented above, along with the information on previous touchscreen research, it is possible at this stage to draw some preliminary conclusions about the parameters which may be suitable for use in crossmodal displays.

Intensity

Intensity changes over time appears to be a usable parameter in tactile displays however using intensity in combination with roughness could be problematic since both parameters are created by manipulating the amplitude of the vibration. Furthermore, high intensity levels could cause pain to users.

Unfortunately, intensity is not a usable parameter in the audio domain as users have reported loud sounds to be annoying and distracting. Dynamics could be used, much like intensity changes over time in the tactile modality but there has been no investigation of this.

Frequency

Frequency appears to be problematic for use in tactile displays because the interaction between frequency and amplitude means that controlling the perceived vibratory pitch can be extremely difficult. The results in the majority of the literature indicate that frequency should definitely not be manipulated in a tactile communication system if intensity is to be manipulated independently. However, it might be possible to combine intensity and frequency redundantly to improve identification. Sherrick [129] suggests that it might be better to use another parameter such as spatial location in place of frequency in tactile communication systems.

Using different audio frequencies is much like using different pitches. There has been little research into the use of pitch as a standalone parameter in earcons. Brewster [17] suggests that register (or pitch) is a poor choice when absolute recognition is required and therefore it would be better to use it in combination with another parameter. If register must be used then large differences between the different levels will be required.

Waveform

There has been very little work on identification of different waveforms in the literature on tactile perception. One potential solution for the creation of distinguishable tactile waveforms is the use of “rough” and “smooth” waveforms. Several papers have mentioned that distinctions can be made between these two types of sensation. Enriquez *et al.* achieved recognition rates of 73% for nine Haptic Icon designs varying in waveform [42].

Waveform is much like timbre in audio. Timbre is the most important and effective parameter used in earcon design, as they are easy to distinguish. It has been found that pianos, organs, xylophones and drums can be particularly well identified by listeners with recognition levels of the sounds over 80% for each instrument [118].

Duration

Duration seems to be a reliable parameter, with the results from the literature suggesting that at least three different durations can be uniquely identified in the tactile modality.

Duration is also applicable in the audio modality. Earcons with up to six notes played in one second have been shown to be usable [8] [18].

Rhythm/Temporal Patterns

Several guidelines on the use of rhythm can be extracted from studies of tactile rhythm perception, which have shown that people are able to identify and reproduce vibrotactile rhythms: musical principles should be applied to the design of temporal patterns, and tempo (speed) can be used as a distinguishing factor. Rhythm has been investigated as the main parameter in tacton design with recognition rates of over 90% achieved when three different rhythms are used [21]

Rhythm has also been a successful parameter in earcon design. McGookin [96] found that users were able to achieve absolute identification rates of over 90% for rhythm (melody) when used in three-dimensional Transformational earcons in combination with timbre and register

Spatial Location

Tactons can be presented to different locations on the body. In tactons research [20], three locations on the forearm have been used. Other research has also shown successful spatial location discrimination on the abdomen, arm and back [33] [142] [147].

Spatial location has not been used much in earcon design but 3D audio could perhaps be used to present feedback individually with each location in the soundscape representing a different piece of information or type of earcon.

2.5 Conclusions

This chapter has presented an overview of the aspects of sound and touch perception relevant to the design of audio and tactile messages such as earcons and tactons, and a review of current audio and tactile feedback solutions for mobile touchscreen applications.

Research Question 1 asks:

RQ1: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

The review of audio and tactile perception allows some conclusions to be drawn about the parameters of sound and vibration which could be used in crossmodal icons. These findings suggest that the most promising parameters for encoding information in tactile displays are spatial location, roughness, duration and rhythm. Rhythm achieves high rates of recognition when using three different rhythms with varying duration and tempo. Spatial location offers a wide range of identifiable values on various different areas of the body. Intensity appears to be usable, but needs to be considered carefully as the subjective magnitude perceived by the user is also dependent on a range of other factors. Intensity change over time has been investigated with promising results indicating that this is a more appropriate choice than static intensity. Frequency is likely to be a poor choice due to interactions between frequency and subjective magnitude, although there has been some success in Haptic Icon work. Another conclusion from the review is that further investigation should be carried out into the possibilities of using waveform, and Chapter 4 presents a study of this tactile parameter for use in crossmodal icons.

The findings of the audio literature review suggest that the most effective parameters are timbre, rhythm, spatial location and duration. Timbre is the main parameter used in earcon design and many different types of instrument can be distinguished. Rhythm is also an important parameter in earcon design and can result in recognition rates of over 90% in combination with other parameters. Initial studies indicate that spatial location is effective at presenting simultaneous audio cues in a 3D soundscape. The results of these studies suggest that spatial location may also be a promising parameter for use with consecutive audio cues. Intensity is not a recommended parameter in the audio domain as high levels can cause annoyance and low levels can go unheard. Frequency or pitch in the audio modality can be difficult to distinguish but there is a possibility that this may be a useful parameter when used in conjunction with others and when large pitch intervals are used.

Lastly, this chapter reviewed related research in the field of mobile touchscreen interaction using different feedback modalities. There has been much research into the use of audio and tactile feedback with touchscreen applications most of which shows that audio or tactile feedback can improve performance with touchscreen devices and can enhance the user experience. However, the majority of the research is unimodal i.e. it investigates audio only or tactile only. When both modalities are studied together the research tends to focus on multimodal interaction not crossmodal interaction (see Chapter 3 for definition). Given that these studies have shown both audio and tactile to be beneficial to touchscreen interaction, it seems logical to consider the crossmodal use of these modalities.

Chapter 3 Crossmodal Interaction

Chapter 2 reviewed related work in the audio and tactile modalities in terms of perception, encoding strategies and applications using these modalities for feedback. The literature review showed that both modalities (in the form of earcons and tactons especially) have been used for a range of purposes, in particular, to improve interaction on touchscreen devices, sensory substitution, providing alerts, communication and enhancing visual user interface elements. There is a growing consensus in current research that fixed allocations of modalities to specific tasks or types of information (i.e. multimodal interaction) is not practical (for example [117]). Instead, interfaces should be flexible and allow for potential changes in the needs and abilities of users, tasks and workload, and the surrounding environment. As Tamminen *et al.* [138] state, mobile devices "*require both hands and visual attention to operate, which is clearly inappropriate for mobile contexts in which some modalities are preserved for other tasks. On the other hand, nomadic user interfaces (designed for interaction while walking) might be too clumsy and awkward for situations where all modalities are available.*" This highlights the need for crossmodal interfaces. Despite the fact that research has shown both audio and tactile icons to be effective means of communication, the area of crossmodal auditory/tactile displays has been studied much less.

In order to explore the possibilities of crossmodal auditory/tactile output, the concept of crossmodal icons is introduced. Crossmodal icons make use of earcons and tactons by exploiting the similarities between modalities to create equivalent sensations in both. The design of these audio/tactile icons centres on the use of the crossmodal dimensions of sound and touch.

This chapter discusses the definition of crossmodal interaction with a focus on initial perceptual studies in the field of psychology. Then, the audio and tactile modalities are analysed in more depth with respect to their potential for use in crossmodal interaction. Lastly the chapter contains a discussion of what audio/tactile crossmodal icons are and outlines the approach used in this thesis to design crossmodal icons.

3.1 Crossmodal Interaction Definition

There are many different uses of the word crossmodal in HCI research and psychology. Most of them are identical to the definition of the term multimodal. This thesis is based on the following definition:

Crossmodal interaction is a subset of multimodal interaction where the different senses are used to receive the same data. This provides a common representation of the data from both senses (in this case, audio and tactile) [52] making them congruent informationally [93]. Crossmodal use of the different senses allows the characteristics of one sensory modality to be transformed into stimuli for another sensory modality. Multimodal interaction, on the other hand, may also use the different senses to receive different information.

The term *crossmodal* originated in studies of perception in psychology where crossmodality discrimination and matching are frequently studied areas [2]. Crossmodality discrimination involves the identification of an object presented in one modality (e.g. audio) using another modality (e.g. touch). Crossmodality matching is a method of scaling where, for example, the loudness of a tone is adjusted so it sounds as loud as a given weight feels heavy.

One of the earliest arguments for crossmodal interaction (although using the audio and visual modalities) was presented by Geldard in 1960 [49]:

"The choice between the eyes and ears as sense channels for the presentation of information to the human operator rests upon the specific demands of various operational situations."

Marks [93] has argued that our underlying ability to integrate information across different modalities is the fundamental process that enables us to perceive similarity. In other words, whenever we integrate data across different modalities we perceive similar qualities regardless of which sensory modality registered the original input. For example, when a person is seen and heard speaking a word, the visible and audible duration and shape of the word is the same. Similarly, when we touch and look at an object we can perceive similar shapes, sizes, and textures in both modalities.

So, for example, in crossmodal interaction with a touchscreen mobile device both the audio and tactile feedback would represent the same data, for example a vibrotactile pulse spatially located on the left of the device indicating an alarm would also be able to be represented through audio using 3D spatial location on the left hand side. Whereas in multimodal interaction, the vibrotactile cue may indicate an alarm while the audio cue may represent a completely different type of information like, for instance, incoming messages.

Crossmodal interaction relates to both synaesthesia and sensory substitution. It has been shown in studies that sensory inputs from the different modalities are directly processed and translated into a common representation (Figure 3-1) which is used for both unimodal and crossmodal comparisons [143]. The research in this thesis is concerned with crossmodal interaction using the audio and tactile modalities. The following sections highlight the use of crossmodal interaction with various modalities in existing research.

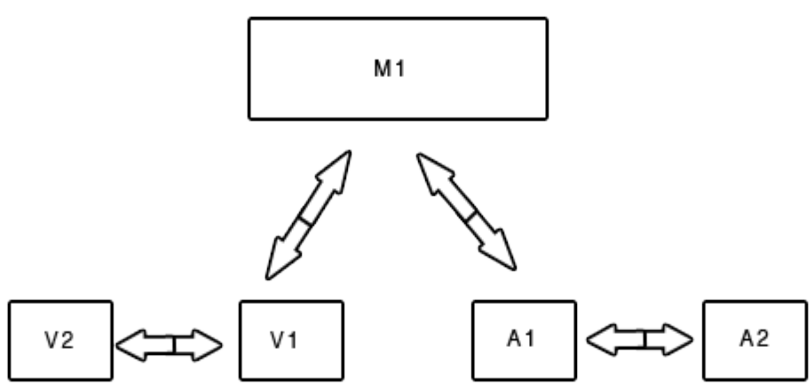

Figure 3-1: Adults can detect equivalent attributes in the audio and visual modalities and integrate them into a unified percept (M1) (from [40]).

3.2 Sensory Substitution

The most popular example of crossmodal interaction can be seen in sensory substitution research. Sensory substitution systems take environmental data which would normally be processed by one sensory system and translate this data into stimuli for another sensory system [84]. The main application of these systems is increasing accessibility for those with sensory impairments. This class of systems include tactile vision substitution, tactile auditory substitution, and teletouch [84]. Sensory substitution research often focuses on the information processing capabilities of the skin and includes the design of tactile aids for visually and hearing impaired people. The body of research from this field has many insights into the relationship between our senses of hearing and touch, all of which provide a basis for the use of crossmodal audio and tactile interaction. One outcome of this research topic has been the development of tactile hearing aids aimed at providing deaf individuals with an additional sensory channel for perceiving spoken language and environmental noise. These devices use vocoding techniques and use frequency-to-place transformation, in which the stimulation location on the skin is mapped to an audio frequency range.

One of the earliest devices in this area is the Tactile Acoustic Monitor (TAM), which was developed by Summers [137]. The TAM uses a vibrotactile actuator to provide information about the loudness levels of the user's speech and other sounds in the surrounding environment. The vibrotactile actuator is turned at a constant amplitude and frequency level if sound levels are above a threshold. Experiments conducted using the TAM showed that the TAM was effective for some lip reading applications. In turn, this led to numerous experiments focussing on speech perception via a vibrotactile actuator [38].

Tactile-vision substitution systems (TVSS) were also some of the earliest of such systems to be developed, with an aim to present visual information to blind people. In a typical system, a camera receives visual information which is then translated into a tactile representation using a two-dimensional pin array (see [36]).

Braille is a very common method used by visually impaired people. Unlike the crossmodal feedback proposed in this research, Braille does not make use of the attributes present in both the visual and tactile modalities. The mappings between characters and the tactile characters must be learned. These mappings are mainly based on a numerical model where each Braille character consists of a three row by two-column cell, with combinations of raised dots allowing 64 individual patterns. The patterns represent the letters of the alphabet and punctuation. Many empirical evaluations have been conducted that indicate that Braille code is more effective than using embossed letters in terms of reading speed and text recognition [125]. The main disadvantage of Braille is that very few blind people can read it. For example, it has been estimated that only 2% of British blind people can read Braille [23].

Some of the most prominent work in this area involved the development of the Optacon (see Figure 3-2) which converted printed letters to a spatially-distributed vibrotactile representation on the user's fingertips, using a miniature handheld camera (detailed in [36]). Although reading speeds were significantly slower than Braille, the Optacon allowed blind people to access any text or graphics without having to wait for it to be translated into Braille.

Figure 3-2: Using the Optacon device to translate text to the tactile modality.

Paul Bach-y-Rita and colleagues also conducted other early research in Tactile Visual Sensory Substitution (TVSS) in the late 1960s. These systems mapped visual navigation information from a tripod-mounted camera to a vibrotactile display attached to the user's back. Unfortunately, due to the limited spatial resolution, tactile masking effects and a low dynamic range, evaluations showed that the system was not appropriate for daily use in navigation. However, studies showed that participants could identify simple shapes and discriminate line orientations. Furthermore, it was stated that more experienced users could complete highly complicated tasks such as facial recognition using the TVSS system [73].

Karam *et al.* [75] developed a sensory substitution technique called the Model Human Cochlea (MHC) used to create crossmodal auditory/tactile interfaces. The purpose of this research is to enable users to 'hear' the musical content of films through the sense of touch. The MHC divides audio vibrations into small bands of frequencies as opposed to presenting a single source. The MHC uses eight audio speakers placed on the user's back or attached to the back of a chair to present the musical excerpts. Studies were conducted to compare the communication of emotional information presented by the MHC (through a Frequency Model and Track Model) and single speaker displays. The Frequency Model is based on the normal distribution of notes in western harmonic music. In other words, notes from the middle of the keyboard occur most often and notes on either side of the keyboard are less prominent. The Track Model expands the Frequency Model by assigning each layer of the musical composition with separate speakers. The experimental results indicate that users can interpret basic emotions (i.e. joy) in music through multiple vibrotactile channels provided by the MHC but it is more difficult for emotions like fear, anger and sadness. Studies also revealed that both the Frequency Model and Track Model are more effective at emotional expression compared to the Control Model with the Frequency Model performing best. This indicates that emotion such as joy, sadness, anger and fear may best be presented using spatiotemporal vibration patterns. Unlike other crossmodal interaction research, the focus appears to be on more than simple data translation but also the recreation of emotions and is based on a model from the human sense of hearing.

The work on crossmodal interaction described in this thesis could be considered a bi-directional form of sensory substitution where the information can be presented to one sense or other depending on the user's particular disabilities or current situation. In most cases sensory substitution occurs in one direction only, for example, Braille is a translation of visual information into tactile information. The bi-directional method in this research not only uses a translation of audio to tactile information but also *vice versa*.

3.3 Audio/Tactile Crossmodal Applications

Despite the fact that research has shown both audio and tactile icons to be effective means of communication, the area of crossmodal auditory/tactile displays has been studied less. Recently, Immersion Corporation has created Vibe-Tonz[10] which could be considered as crossmodal. These are vibrotactile messages which can be used, like personalised ringtones, to indicate the identity of the caller in a mobile phone. However, there have been no empirical tests conducted to determine the effectiveness of these cues or to discover the amount of information that could be encoded in the cues.

[10] Immersion: www.immersion.com/mobility/docs/VibeTonz_Mobile_Player_0305_v1.pdf

van Erp and Spape transformed a set of audio melodies to the tactile domain using a low pass filter [145]. However, they only established one parameter and one dimension for the tactile versions of the melodies (tempo and intrusiveness). Intrusiveness is a dimension or component classified as ranging between soft and polished to loud and aggressive. As can be seen in Table 3-1 in the following section, there are many more parameters that could be investigated.

The Touch Engine by Sony is another system that could be considered crossmodal. Sony's Touch Engine has a vibrotactile screen through which users can feel images and buttons that are on the screen. In the touch engine a heart icon is represented by a heartbeat sensation [113]. This is an example of an iconic relationship between the signified and the signifier. Although this is not a direct translation from vision, it is an intuitive, direct translation from sound and, in fact, touch itself (as you can feel someone's heartbeat). This suggests that although some parameters in a certain modality cannot be directly translated into another modality, it may be possible to use symbolic mappings. Therefore, in this research, abstract mappings are also investigated for some audio parameters such as pitch, which is not directly transferable to the tactile domain, so that the cues are intuitively equivalent instead of a direct translation.

There have been some studies investigating the use of more than one modality in combination and separately. Vitense *et al.* [151] added uni, bi and tri-modal feedback to a simple GUI using the visual, audio and tactile modalities. Then, the participants' performance and perceived workload were evaluated during a 'drag and drop' task using the different types of feedback. The results showed that performance was significantly higher and workload significantly lower using the bimodal tactile/visual condition and unimodal tactile and visual conditions in comparison to unimodal, bimodal or trimodal conditions containing audio. Unfortunately, there was no discussion included in the paper of the actual design of the feedback so it is not clear whether, in the unimodal conditions, the feedback was designed to be crossmodal or equivalent. However, the results do indicate that, for temporal tasks, tactile feedback may be more appropriate than audio feedback or that the audio feedback was poorly designed.

There has been some research into the use of the visual and tactile modalities in crossmodal interaction. van Erp and Verschoor [148] investigated tracking performance with tactile and/or visual presentation of target and cursor; where the tactile display consisted of vibrators in a horizontal linear array on the torso and the visual display consisted of dots projected on a horizontal plane surrounding the observer. Participants performed two different tracking tasks with target and cursor presented to the same modality (either visual or tactile) or to different modalities (a visual target and a tactile cursor or *vice versa*). In the unimodal conditions, the target and cursor were both visual and both tactile. In the crossmodal condition the target was visual and the cursor tactile or *vice versa*. The results from the crossmodal conditions show that there are no costs involved with respect to tracking performance when target and cursor are presented to different modalities. This indicates that the visual and tactile channel can be used in a crossmodal tracking display when both modalities contain qualitatively equal data. The positive results from this research indicate that it is worthwhile to investigate the use of different modalities for crossmodal information presentation, especially in the case of touchscreen mobile devices where the visual sense is usually overloaded with many graphics or is otherwise occupied with the surrounding environment.

The ability to match information between senses has also been studied in cognitive brain research. Saito *et al.* [123] used an MRI scan to study neural representations of crossmodal matching between the visual and tactile senses when presented with 2D shape information. The results of the MRI scans showed that shape information from different senses is indeed integrated in the brain during matching tasks. This shows that humans are naturally able to integrate information and match it and process it the same way despite what sense it was presented to.

Crossmodal priming between the visual and haptic modalities has also been of interest to researchers. The ability to remember stimuli from different modalities indicates an implicit memory representation that is accessible multimodally [116]. Reales and Ballesteros investigated implicit and explicit memory of stimuli using intramodal and crossmodal conditions. The variables measured were the speed of object naming, the level of completeness at which a fragmented picture could be identified and speed of detecting whether a line drawing depicted a real object. The results showed that crossmodal and intramodal priming does occur (faster responses for previously studied objects regardless of the original modality input), and in some cases the speed at which objects were remembered when presented to different senses was the same regardless of the initial modality used. Similar intramodal effects were discovered by Craig and Sherrick [36] whilst investigating the potential of different body locations for tactile stimulation. The authors found that once subjects have been trained in tactile pattern recognition on the back, they can almost immediately recognise the same patterns when they are presented to the thigh or abdomen. This transfer of learning also occurs, somewhat, when patterns are presented to different fingers after training on one finger, but is not so immediate. Similar crossmodal effects can be seen in the study detailed in Chapter 5.

There have been few studies comparing three modalities for feedback in unimodal settings. Modalities are usually combined or used in a unimodal manner with no comparison. Akamatsu *et al.* [3] studied the effect of different types of modality feedback (visual, audio or tactile) on a target selection task using a modified mouse. Tactile feedback was presented using an aluminium pin protruding from a hole in the left mouse button. There were five conditions in the experiment: normal, colour (the shading of the target changed), auditory (a 2kHz tone when in target area), tactile (the pin under the fingertip was raised upwards) and combination (colour, auditory and tactile). The results of the study showed no difference in response time or error rates. Significant differences were found in the final position times (the time between the cursor entering the target and selecting the target) with tactile feedback producing significantly faster times than visual feedback. The authors argue that tactile feedback allows subjects to use a wider area of the target and to select targets more quickly once the cursor is inside the target. Although these results only apply to targeting tasks, they suggest that audio and tactile crossmodal presentation may have the potential to benefit other types of tasks too.

3.4 Amodal Attributes in Audio and Tactile Crossmodal Interaction

Audio and tactile displays are ideal candidates for crossmodal use because our senses of hearing and touch share several important similarities, in particular their temporal characteristics and their ability to perceive vibrations. Moreover, sounds are often described in tactile terms. Mursell [101] observed that tones can contain tactile values as can be seen when we describe a tone as hard or soft, rough or smooth, wooden or metallic.

An attribute that can communicate comparable information across modalities is considered to be amodal. Mendelson [97] provided a scheme or list of such amodal properties. These properties relate to space and time and involve points along a continuum (e.g. location), intervals within continuum (e.g. duration), patterns of intervals (e.g. rhythm), rates of patterns (e.g. tempo), or changes of rate (e.g. texture gradients). Other crossmodal properties such as numerosity or intensity also have been examined.

In the most basic of terms, amodal information is information that is not specific to a particular sensory modality; rather, it is completely redundant across one or more senses [51] [52]. Perceptual experiences can be amodal [52]. That is, different senses may generate very different sensations but the same information about the world. This whole concept of amodal attributes and modality specific attributes is referred to as the 'Gibsonian' point of view [104]. Many researchers have studied whether tactual knowledge of shape and spatial relations is identical to information acquired by the visual modality [58] and have found that there are significant similarities between information acquired visually or tactually. In other words, this thesis addresses the following question: how does one accomplish a translation across the modalities, namely acquire an audio understanding of something that has been touched and *vice versa*?

Walker-Andrews [153] defines amodal information as information not specific to one modality; rather, the same information can be detected by several modalities. Examples include temporal relations such as rhythm or tempo, and properties of an object such as size, shape, texture, and substance. Others have also discussed amodal attributes: see intermodal invariance (see [52]) and common sensibles (see [93]).

Perhaps one of the most important pieces of work in relation to this thesis and crossmodal interaction is 'Development of Intersensory Perception in Human Infants' by David Lewkowicz [86]. He has specifically investigated some of the parameters in audio and tactile that could be used in crossmodal interaction. Like Gibson [52], Lewkowicz states that there are two classes of stimulus attributes: amodal and modality-specific. His examples of amodal attributes are duration, rhythm, shape, intensity, and spatial extent. Modality-specific attributes are those attributes that can be represented only in a single modality because their specification depends on the unique transduction properties of that modality. Examples of modality-specific attributes are colour, odour, and temperature.

The shared temporal and spatial properties between audio and tactile mean that certain audio characteristics may be transformed into tactile stimuli (and *vice versa*). Therefore, the same data may be presented interchangeably via the two different modalities in crossmodal interaction. Table 3-1 outlines the potential amodal parameters available between the modalities.

Parameter	Available in Audio?	Available in Tactile?
Rhythm	Yes	Yes
Pitch	Yes	No
Loudness (Amplitude)	Yes	Yes
Timbre (Texture)	Yes	Yes
Duration	Yes	Yes
Spatial Location	Yes	Yes
Rate (Tempo)	Yes	Yes
Dynamics	Yes	No

Table 3-1: Parameters available in the audio and tactile modalities.

The shared temporal property between audio and tactile means that certain audio characteristics such as rhythm, tempo and duration can be transformed into comparable tactile stimuli (and *vice versa*). This is a bi-directional form of sensory substitution where the information could be presented to one sense or other depending on a user's particular disabilities or current situation.

3.5 Crossmodal Icons

Two icons may be considered to be crossmodal icons if and only if they provide a common representation of data, which is accessible interchangeably via different modalities.

3.5.1 Audio and Tactile Crossmodal Icons

In order to explore the possibilities of crossmodal icons, this research focuses specifically on the crossmodal use of audio and tactile icons with an aim of investigating the best ways in which to design such icons and make use of them in touchscreen applications. A large body of work already exists on the design of audio and tactile icons and, since sound and vibration are both temporal and spatial in nature, it seems likely that work from both of these unimodal domains could also be used to inform crossmodal design.

One novel contribution of this research is the application of earcon and tacton design principles to the problem of crossmodal design, learning from both the structure of earcons and tactons and the parameters used to encode data in them (see Chapter 2 for a description of earcons and tactons). Earcons have been designed by identifying parameters of audio that users can identify multiple levels of, and then combine these to create unique motifs. This is also the approach used in the design of tactons. Therefore it seems appropriate to model the design of crossmodal icons on the common design principles that already exist in both the audio and tactile domains.

As is the case with all types of icon including theearcons and tactons introduced in Chapter 2, for crossmodal icons to convey data successfully, there should be a mapping between the data to be communicated and the stimuli presented to the user. In the area of semiotics (the study of signs) there are different modes of relationship between data and their representation: iconic, indexical and symbolic [26]. Likeearcons and tactons, crossmodal icons use symbolic mappings. These are not based on any pre-existing understanding of the mapping between data and sound or touch. In other words, these mappings are arbitrary and require users to be trained to understand the relationship between data and sound or touch explicitly. According to Saussure (from [26]), a symbolic relationship mode makes use of a signifier (crossmodal icon) that does not resemble the signified (data). The relationship must be learnt (for example, a green traffic light or a character (letter)).

Crossmodal icons allow the same data to be accessible interchangeably via several different modalities. To be able to compare audio signifiers to tactile signifiers in a crossmodal setting, there must be a common representation of the signified from both senses. For example, a set ofearcons/Tactons can be considered to be crossmodal if the information represented can be encoded in both modalities so that users can move from an audio to a tactile presentation of the same data (and *vice versa*). Crossmodal icons are structured, abstract and use a symbolic approach as opposed to an iconic or indexical approach like those found in visual icons, Auditory Icons [46] and Hapticons [92].

Multiple dimensions of data can be encoded in crossmodal icons, with each represented by a different crossmodal parameter. For example, if audio/tactile crossmodal icons were used to represent files in a computer interface, the file type could be represented by rhythm (in audio and tactile), size by duration (audio and tactile) and creation date by intensity (audio and tactile). Each file type would be mapped to a unique rhythm equivalent in both modalities. Therefore two files of the same type, and same size but different creation date would share the same audio/tactile rhythm and audio/tactile duration but would use different levels of audio/tactile intensity.

Research Question 1 in this thesis asks:

RQ1: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

The findings in the review of perception of tactile and audio parameters reported in Chapter 2 indicated that the most successful parameters for encoding information in tactile messages, such as tactons, are spatial location, roughness and rhythm. The review also indicated that the most successful parameters for encoding information in audio messages, such asearcons, are timbre, pitch, rhythm, duration and spatial location. The section in this chapter on the psychological aspects of crossmodal interaction has highlighted that there are amodal attributes available in both the senses of touch and hearing that can be used to represent the same data. These include intensity, spatial location, rate, texture, and rhythmic structure [86]. Therefore, the auditory/tactile crossmodal interaction design described here takes the most successful parameters inearcon and tacton research that have also been identified as amodal attributes: rhythm, texture, and spatial location. Although parameters do not necessarily have to be amodal to be used in a crossmodal manner, they are proposed here because it seems likely that using amodal parameters in both modalities could reduce learning time. Experiments conducted to investigate the best ways to create the crossmodal stimuli using these parameters are described in Chapter 4. These experiments are required because there are many ways to create the different parameters in the stimuli and unfortunately, some of the most effective parameters inearcons cannot be directly transferred to the tactile domain. Geldard points out that the correspondence between vibratory frequency and perceived "pitch" is a tenuous and uncertain one. Vibratory pitch appears to be a combination of both frequency and amplitude [49].

3.6 Conclusions

This chapter has reviewed the concept of crossmodal interaction, specifically the use of audio and tactile crossmodal interaction. It has identified several research applications that make use of different modalities in a crossmodal manner with a focus on the aspects of crossmodal interaction and amodal attributes that have been employed in many sensory substitution systems. These often incorporate the use of vibrotactile displays to present speech to hearing-impaired people by translating the speech signals into temporal or spatial vibrotactile patterns.

Sensory substitution systems are commonly aimed at users with physical impairments such as blindness or deafness. However, by employing the concepts from sensory substitution in crossmodal audio and tactile interaction, it is proposed that users with situational impairments will benefit. In other words, mobile touchscreen users can be in a variety of different situations when using their device. At a loud concert or party, audio feedback could go unnoticed so in this situation it may be more appropriate for information to be translated into the tactile modality instead. However, when the device is not actually in contact with the user's skin, tactile feedback will not be felt. In this case, information could be translated into audio instead of tactile therefore accommodating the user's situational impairment.

The other main aspect of this chapter was the introduction of crossmodal icons and their parameters. By using the psychological reviews of amodal attributes, three audio/tactile crossmodal icon parameters were established for further investigation in answer to Research Question 1. These amodal parameters are also established parameters in earcon and tacton research: rhythm, texture and spatial location. The analysis of amodal attributes also identified other potential parameters such as duration, intensity and rate. These parameters have had little to no specific attention in earcon and tacton research so will require more intensive investigation. Through experimental evaluations, Chapter 4 explores the three main parameters mentioned: rhythm, texture and spatial location in order to investigate the best ways in which to design the feedback so that data encoded using these parameters can be perceived as synonymous in the audio and tactile modalities.

This approach to the design of crossmodal icons has not been used before as most research has focused on a particular sensory modality and although the research in earcons and tactons has some similarities, they have never been combined and their amodal attributes have never been exploited to aid in mobile touchscreen use. Using these design principles along with the understanding of crossmodal perception established in this chapter, crossmodal icons can be created and evaluated though empirical studies and incorporated into mobile touchscreen applications.

Chapter 4 Individual Design Parameters

Research Question 1 in this thesis asks:

> **RQ1**: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

The current parameters under investigation have been derived from a survey of related work on the parameters available in the audio and tactile domains, which, in turn, have been derived from psychoacoustics and psychophysics as discussed in Chapters 2 and 3. In order to address the research question, it is important to consider what features are important in a crossmodal icon parameter. The following two factors are very important in the choice of parameters for this particular research.

1. The user must be able to distinguish and identify multiple levels of the parameter;
2. The parameter must be amodal (i.e. both modalities must have an equivalent parameter).

Factor 1 is important because multiple levels of a parameter must be able to be identified so that more than one dimension of data can be encoded. Therefore, two or more levels must be identifiable for a parameter to be successful. For many tactile parameters, Geldard [49] suggested that three levels is the most that can be expected without extended training and similar results apply to the audio modality according to Brewster *et al.* [17] in their study of earcons. Therefore this is the number of levels of each parameter that is aimed for in this research.

Factor 2 is perhaps the most important issue in this research because crossmodal icons must be made up of tactile cues that may be perceived as equivalent to audio cues and *vice versa*. When speaking of equivalence, the work in this thesis focuses on a user's ability to match pairs of crossmodal earcons and tactons based on their parameter similarities. The encoding of data is similar to that of both earcons and tactons but each of their shared parameters is manipulated to develop equivalent cues. A crossmodal parameter could be considered as a coupling of two single modality dimensions (for example, pitch and roughness) rather than a data attribute and the mapping to some concrete physical stimuli. The coupling persists over any data encoded. Although it is not necessary to use amodal parameters in the design of crossmodal icons, these parameters provide an obvious starting point for design when using the audio and tactile modalities given their similarities. Consequently, the parameters must be investigated to determine whether they can be considered amodal and thus can map the same data between the two modalities. After the extensive literature review and initial pilot experiments, it became clear that this is a complicated issue because some of the most effective parameters available in the audio domain do not have direct mappings to the tactile domain and *vice versa*. For example, pitch and melody are some of the main parameters used in earcons [17] but there is no absolute equivalent in the tactile domain so these cannot be recreated in vibrations.

The findings in the review of perception of tactile and audio parameters reported in Chapter 2 indicated that the most effective parameters for encoding data in tactile messages, such as tactons, are spatial location, roughness, intensity change over time and rhythm. In addition to these tactile parameters, the review in Chapter 2 identified parameters such as timbre, rhythm, duration, register, and spatial location as appropriate methods of encoding data in the audio modality. However, there is not a direct equivalent to register in the tactile domain. Frequency is its nearest match and current literature on tactile perception [80] suggests that frequency can be difficult to distinguish so cannot be used effectively as a multidimensional parameter. According to the literature on amodal parameters [87] rhythm, spatial location, texture and intensity change over time are all suitable candidates for crossmodal use. Of these four parameters, three have been successfully used in the design of earcons and tactons separately, namely rhythm, spatial location and texture. These parameters have received little attention with regards to crossmodal use and, therefore, require some experimental investigation in order to understand how they are perceived before they are used in crossmodal icon design (i.e. if they are perceived as synonymous in audio and tactile and if so, what is the best match: does audio timbre match tactile waveforms?).

This chapter reports two studies investigating the different possible parameters and mappings that can be used to facilitate crossmodal auditory/tactile feedback. The experiments conducted have investigated rhythms with texture and spatial location as potential parameters. Given that the outcome of this research is intended for mobile devices, the experiments were conducted in both a lab-based stationary environment and in a simulated mobile environment.

4.1 Rhythm

As discussed in Chapters 2 and 3, rhythm is an extremely important parameter in both earcons and tactons. Furthermore rhythm also has amodal properties that make it a potential crossmodal parameter as demonstrated in this section. There was no need to conduct an experiment to investigate whether rhythms can be perceived as equivalent in the audio and tactile modalities because research in this area has already shown rhythms can be mapped between both modalities. Therefore this section contains a short review of the related work on audio and tactile rhythm recognition.

Changing the rhythm of a motif (a short rhythmic structure) can make it sound very different. Blattner *et al.* [8] describe rhythm as the most prominent characteristic of a motif. Earcons are based around different rhythms and this is one of the most important methods for grouping sounds into sources (Figure 4-1).

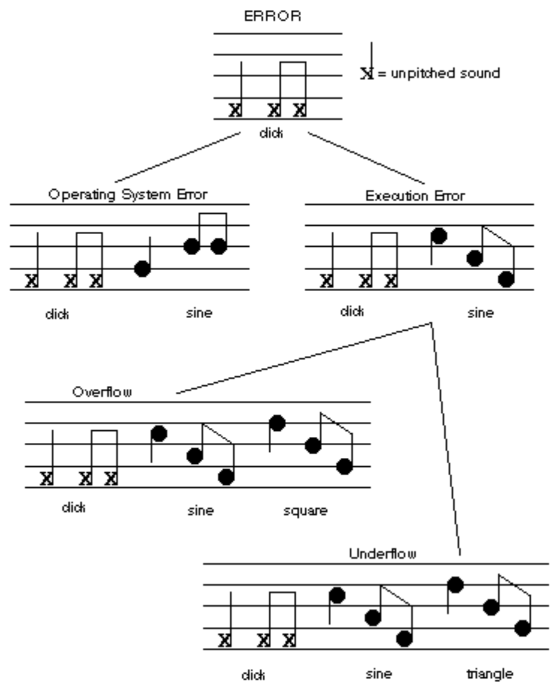

Figure 4-1: Example rhythms used in hierarchical earcons.

In his work on earcons, Brewster uses rhythm (short motifs) to represent objects or actions [17]. An example is shown in Figure 4-2 where items of the same *type* share the same rhythm. For example, the programs all have the same rhythm, the folders another and the files another.

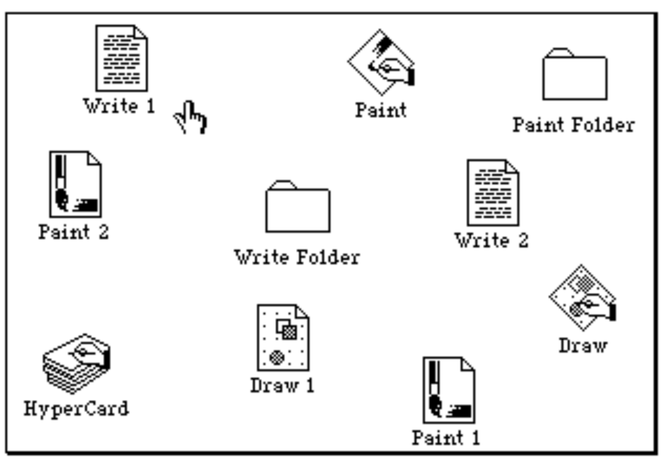

Figure 4-2: Example objects represented by earcons.

Sumikawa *et al.* [134] state that only seven time divisions should be used when creating rhythms, notes should be kept within a range of eight octaves of twelve notes and earcons should be musically neutral.

In the tactile domain, rhythms are created by grouping pulses together to create temporal patterns in the same way as musical audio rhythms. In tactons research, rhythm is created using vibration bursts and gaps of different durations to form temporal patterns. Three rhythms were created by Brown *et al.* [21]: the 7-note rhythm made up of seven short vibrations, the 4-note rhythm made up of four longer vibrations, and the 2-note rhythm consisting of one short vibration and one very long vibration. These rhythms are shown in Figure 4-3. The design of tactile rhythms in tactons research is, in actual fact, based on previous work on audio rhythms in earcons research. The design used by Brown *et al*. [21] follows Brewster's guidelines [17] as well as advice given by van Erp and Spapé [145] who identified tempo (speed) as an important parameter in the identification of tactile melodies. Although all three rhythms are created using the same tempo, they feel faster or slower due to the use of many short pulses (for example, 7-note rhythm), or few long pulses (2-note rhythm).

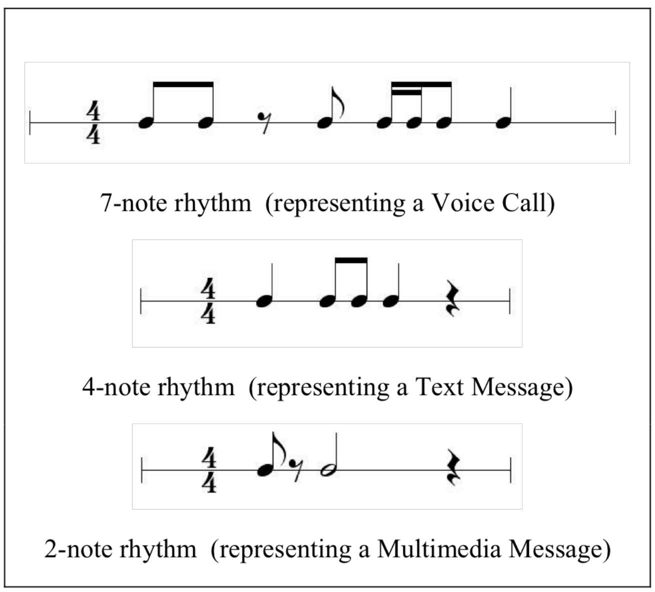

Figure 4-3: The three rhythms used in tactons research.

In an absolute identification lab-based experiment [21], it was discovered that these tactile rhythms could be identified with 96% accuracy, making them a good choice of parameter for tactons.

Rhythm has also been used as a successful parameter in Haptic Icons (brief tangible stimuli with associated meanings) [141]. By using rhythm in combination with frequency and amplitude, 84 distinguishable Haptic Icons can be produced. It has been shown that users perceptually organise rhythms, making such a large set of icons possible. Through evaluation [141], Ternes and MacLean found that the two primary characteristics by which users distinguish tactile rhythms are note length and unevenness.

As mentioned above, van Erp and Spapé investigated the creation of "tactile melodies" to represent information in computer interfaces [145]. Pieces of music were transferred from the auditory domain to the tactile domain and an investigation of how people described and classified certain types of melodies when they were presented via a tactile actuator was then conducted. Unlike Ternes and MacLean, the results from this study showed that the two features most important in tactile melody classification are tempo (speed) and intrusiveness. Intrusiveness is not explicitly defined but the authors state that levels of intrusiveness ranged from soft and polished to loud and aggressive, indicating that users associated intrusiveness with the volume/strength, texture and emotion of the pattern.

Such motifs or rhythmic structures can be used in both audio and tactile displays due to their shared temporal properties [87]. Rhythm is an amodal property in the audio and tactile domain because it can be directly transferred and mapped between modalities. Kosonen and Raisamo [70] investigated the perception of audio, visual and tactile rhythms. Participants were presented with a rhythm in one of these modalities and asked to reproduce it by tapping it out on a mouse button. The audio stimuli were presented via a loudspeaker using a simple tone, the tactile stimuli via the vibration function of a Logitech Wingman mouse (www.logitech.com) and the visual stimuli via a flashing circle on the computer screen. Simple rhythms were created using just two lengths of notes: short (300ms) and long (600ms). The results showed that using the audio modality resulted in the best performance, with only 7.8% of rhythms wrongly reproduced in this modality compared to 14% in the tactile modality and 17.5% in the visual modality. Not only were the results in the audio condition significantly better than those in the visual and tactile conditions but the results in the tactile condition were also significantly better than the visual modality condition.

Buttler and Oravainen (unpublished [24]) conducted absolute identification experiments for rhythm presented through the audio and tactile modalities. The audio feedback was created with 500Hz sine waves while the tactile feedback was created with 159Hz vibrations through a standard vibration motor in a mobile phone mock-up. The aim of the experiment was to compare perception of audio and tactile rhythms, and to compare perception of rhythmic and non-rhythmic temporal patterns. In this experiment the rhythmic patterns all contained six beats of 800ms length, while the lengths varied in the non-rhythmic patterns. The results showed no statistically significant differences in the participants' ability to identify rhythms in the audio and tactile modalities. Also, performance was significantly better in both modalities when rhythmic patterns were used as opposed to non-rhythmic temporal patterns. These results indicate that rhythm can be identified accurately in both the audio and tactile modalities and that using rhythmic rather than non-rhythmic patterns should be used.

Research by Jokiniemi *et al.* [70] showed that absolute identification in both the audio and tactile modalities is possible. Pairs of rhythm patterns (audio, tactile or visual) were presented to subjects who made a same-different judgment. All possible combinations of the three modalities were used. The results showed that the unimodal auditory condition had the highest rate (79.2%) of correct responses. The unimodal tactile condition (75.0%) and the auditory-tactile condition (74.2%) produced very similar results. The average rate remained under 61.7% when the visual modality was involved. In that sense, the tactile modality settles between audio and visual modalities in terms of rhythm perception. Overall, the results confirm that the auditory and tactile modalities are suitable for presenting synonymous rhythmic data, and several participants in the study thought that the tactile modality was almost as pleasant as the audio modality.

Given that rhythm can be perceived as equivalent in the audio and tactile modalities, it appears to be a suitable parameter for use in crossmodal icons. Rhythm could, for instance, be used to encode information about the type of an alert. For example, in a mobile phone, an appointment reminder could be represented by one of the rhythms in Figure 4-3. The audio icon would play this rhythm from a Wave file via a loudspeaker. The tactile icon would transmit this same rhythm via a series of pulses through the vibrotactile device.

4.2 Experiment 1a: Crossmodal Roughness

Roughness has been used as a reasonably effective multidimensional parameter in tactons research [20]. Modulating the amplitude of a tactile pulse creates differing levels of roughness ranging from smooth to extremely rough. It may be possible for users to perceive an auditory equivalent of tactile roughness, given that, sounds are often described in tactile textural terms. Mursell [101] observed that tones can contain tactile values as can be seen when we describe a tone as hard or soft, rough or smooth, wooden or metallic.

Roughness or texture has also been widely studied in the audio domain [106] [67] [140] [43]. Unlike the tactile modality, there are many ways to create multidimensional roughness in audio feedback. An experiment was conducted to determine which of these versions of audio roughness mentioned in the literature (dissonance, flutter-tonguing, amplitude modulation, or timbre) can be perceived as equivalent and maps most effectively to tactile roughness (amplitude modulation).

There were four conditions in this experiment:

Audio roughness created with dissonance [106] – It is widely known that dissonance of musical dyads depends on the frequency ratio of the interval formed by the two tones. Sounds produced by most musical instruments are harmonic complex tones. When two complex tones are played simultaneously, the sound fluctuates in amplitude, due to beats that occur between their harmonics. The beat rate is equal to the difference in frequency between the two beating tones. Beats are perceived differently, depending on their rate. When the beat rate is below about 10 Hz, the beats are heard as loudness fluctuations. As the beat rate increases, the sound becomes unpleasant and is perceived as rough. The roughness sensation reaches a maximal strength when the beat rate is within a range of about 20-60 Hz, and diminishes, as the beat rate is further increased above 60 Hz.

Audio roughness created with flutter tonguing [140] – the musical technique of flutter tonguing is said to create a rough tone. It is accomplished using either the tongue or ventricular folds (false vocal folds) and it simply amounts to the addition of a 15 – 30 Hz signal to the breath pressure.

Audio roughness created with amplitude modulation [140] [43]– exactly the same as tactile roughness for example, a 250Hz sine wave modulated by a 50Hz or 30Hz sine wave as shown in Figure 4-4. Terhardt [140] reported that audio roughness can be created by amplitude modulation, and also by frequency modulation or audio beating (the pulsing sound which occurs when two tones, close in frequency, are played simultaneously). He reported that the most important factors affecting the perceived roughness of amplitude-modulated signals are modulation frequency and modulation depth. He also noted that, below 20Hz, the listener can recognise the individual fluctuations within a signal, whereas above that point the individual fluctuations are no longer perceived as separate events, and the signal sounds "rough" or "harsh". Fastl [43] reports that above 20Hz the perceived roughness increases as modulation frequency increases, until high frequencies where the ear can no longer detect the fluctuations (around 1000Hz), at which point the roughness disappears.

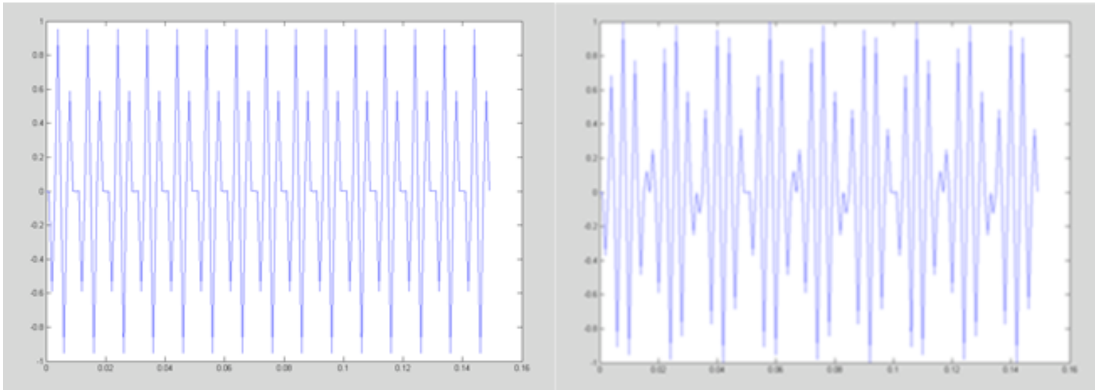

Figure 4-4: Amplitude Modulation - 250Hz sine wave modulated by a 50Hz and 30Hz sine wave.

Audio roughness created with differing timbres [67] – using the different types of sound different instruments make, for example a flute is often considered smooth whereas a saxophone can be considered rough. Timbre is the key parameter in earcons [17] as it was found that users could easily distinguish different timbres and that users were able to achieve absolute identification rates of over 90% for timbre (synthesised piano, violin and trumpet) when it was used as a parameter in three-dimensional earcons with rhythm and register. In Brewster's earcon work [17], different file type attributes were encoded in the earcons. Each *family* of related items shared the same timbre. For example, each menu had its own timbre (violin for menu 1, electric organ for menu 2 and a 'fantasy' sound for menu 3). It must be noted however that some timbres are continuous and some are discrete. The cello timbre is continuous whereas the piano timbre is discrete. This means that timbres should be chosen carefully depending on the required length of the feedback. It must be noted however, when choosing a particular timbre, that some timbres are continuous and some are discrete due to the nature of different musical instruments. For example, the cello timbre is continuous whereas the piano timbre is discrete. A continuous timbre continues to sound until it is turned off whereas the sound of a discrete timbre only lasts a short time [18]. This is due to the nature of certain musical instruments. If continuous sounds are required, discrete sounds would have to be constantly turned on and off in an audio synthesis application to replicate a continuous sensation i.e. the timbre parameter can be augmented with the duration parameter. This means that timbres should be chosen carefully depending on the required length of the feedback.

The aim of the experiment was to determine which of these versions of audio roughness map best to tactile roughness.

4.2.1 Hypotheses

1. One of the versions of audio roughness will be more easily mapped to tactile roughness.
2. That audio roughness can be matched with tactile roughness.

4.2.2 Methodology

16 participants (aged between 18 and 32, 6 females and 10 males, all students at the University of Glasgow and all right-handed with no prior experience of crossmodal interaction) were presented with an audio or tactile cue and then asked to select the equivalent cue from the choices given (Table 4-1). Participants were able to play the choices as many times as they wished.

Modality	Cue Presented to Subject	Answer Choices
Tactile	Rough	Smooth, medium rough and rough timbre
Tactile	Med rough	Smooth, medium rough and rough timbre
Tactile	Smooth	Smooth, medium rough and rough timbre
Tactile	Rough	Smooth, medium rough and rough dissonance
Tactile	Med rough	Smooth, medium rough and rough dissonance
Tactile	Smooth	Smooth, medium rough and rough dissonance
Tactile	Rough	Smooth, medium rough and rough amplitude modulation
Tactile	Med rough	Smooth, medium rough and rough amplitude modulation
Tactile	Smooth	Smooth, medium rough and rough amplitude modulation
Tactile	Rough	Smooth, medium rough and rough flutter tonguing

Tactile	Med rough	Smooth, medium rough and rough flutter tonguing
Tactile	Smooth	Smooth, medium rough and rough flutter tonguing
Audio	Smooth timbre	Smooth, med rough or rough tacton
Audio	Med rough timbre	Smooth, med rough or rough tacton
Audio	Rough timbre	Smooth, med rough or rough tacton
Audio	Smooth dissonance	Smooth, med rough or rough tacton
Audio	Med rough dissonance	Smooth, med rough or rough tacton
Audio	Rough dissonance	Smooth, med rough or rough tacton
Audio	Smooth amplitude modulation	Smooth, med rough or rough tacton
Audio	Med rough amplitude modulation	Smooth, med rough or rough tacton
Audio	Rough amplitude modulation	Smooth, med rough or rough tacton
Audio	Smooth flutter tonguing	Smooth, med rough or rough tacton
Audio	Med rough flutter tonguing	Smooth, med rough or rough tacton
Audio	Rough flutter tonguing	Smooth, med rough or rough tacton

Table 4-1: Cues presented to participants and the corresponding choices available.

For example, the participant was presented with a medium rough tactile cue. The choices presented were three different flutter tongue audio samples created using Soundtrack Pro (www.apple.com/finalcutstudio/soundtrackpro). Participants had to pick the sound he/she thought matched best with the tactile version. Or, a medium rough tactile cue is presented and participants had to choose between three different audio cues with differing levels of dissonance.

First, participants took part in a training session to introduce them to the relevant terminology (roughness, intensity, timbre etc.). Then, participants were shown a worked example (online), which contained example tactile and audio cues, example questions and their correct answers. Next, the participants began the experiment using an online system. The first block of questions displayed by the system is a practice set although the participants were unaware of this. The order of the tasks was randomised. Each question was multiple choice and the participants selected answers using checkboxes.

The online system recorded three dependent variables: the time taken for the participant to answer the question, the number of times each example was played, and the correctness of each answer. The independent variable was the different audio/tactile versions of the tactile/audio parameter.

This experimental method embodies a within-subjects design where all levels of the independent variable are presented to each participant. Therefore, each participant performs tasks related to all conditions. The tasks were displayed in a random order which helps to minimise any learning effect. Also, through the use of a practice set of tasks, all participants start with the same level of expertise. It is also important to minimise the effects of fatigue on a participant's performance. A within-subjects study means that each participant is likely to have to complete a large number of tasks, which could lead to fatigue. To counteract this, the experimental method includes breaks between blocks of questions.

The set of tactons developed in the following studies were presented to users through a C2 Tactor from Engineering Acoustics Inc (www.eaiinfo.com). As mentioned in the literature review, this device is a voice coil transducer with a contact point located outside of the case so that the user feels the vibration through the contact point. The C2 Tactor was attached to the index finger of the non-dominant hand.

4.2.2.1 The Experiment Parameters

- Each task had a time limit of 30 seconds.
- In each condition, each task was shown six times with tactile choices and six times with audio choices.
- A break was given to all participants after every 8 tasks.
- A practice set of 8 tasks was presented to participants at the beginning of the experiment.

4.2.2.2 Online System

The online system used by participants is shown in Figure 4-5.

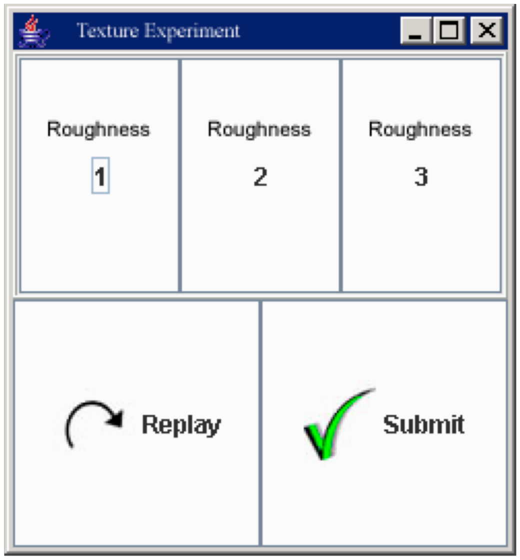

Figure 4-5: Screenshot of experimental system used to present textured tactile and audio cues.

4.2.3 Results

The average number of errors and the average response time for the four audio roughness conditions are shown in **Figure 4-6** and Figure 4-7 (raw data is included in Appendix A).

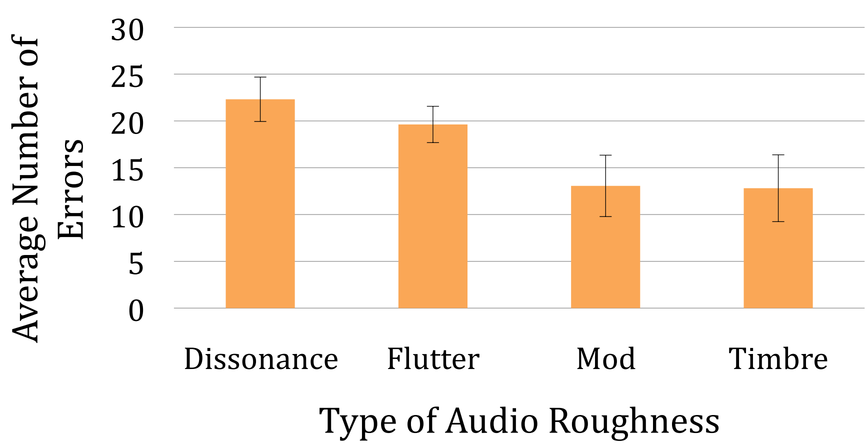

Figure 4-6: Average number of errors out of 36 for each audio roughness condition (with standard deviations).

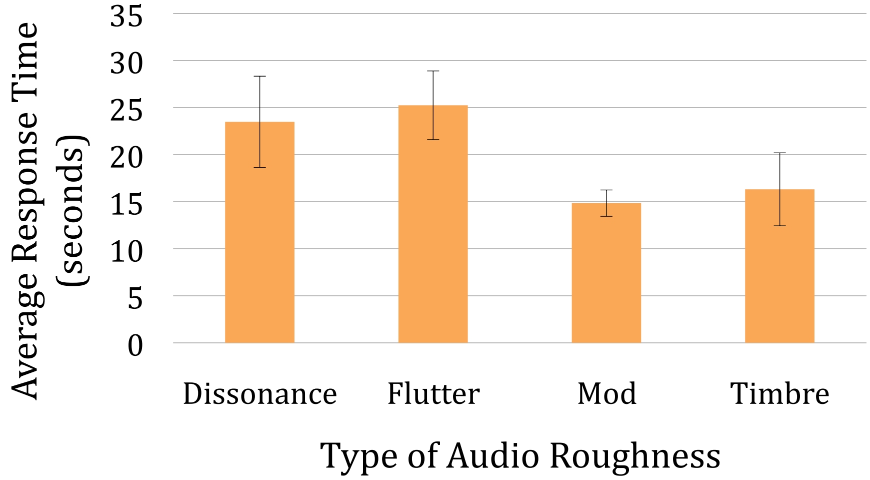

Figure 4-7: Average response time per condition (with standard deviations).

To test the hypotheses the significance of the effects of each audio roughness condition was investigated. The statistical analysis used here is a standard 1-factor repeated measures ANOVA, based on the critical values of the F distribution, with alpha =0.05. In all cases, conservative readings of the critical values of the F distribution were used.

4.2.3.1 Errors

An ANOVA on the mean number of errors made by each participant for each condition showed there are significant differences in the error data between audio roughness conditions ($F(3,60) = 9.027$, $p = 0.05$). Tukey's pairwise analysis showed that the average number of errors for dissonances and flutter tonguing was significantly greater than the number of errors in amplitude modulation and timbre ($p<0.05$). There were no other pairwise differences.

4.2.3.2 Response Time

There are significant differences in the mean response time data between questions ($F(3,60) = 7.897$, $p = 0.05$). Tukey's pairwise analysis showed that the average response time for dissonances and flutter tonguing was significantly greater than the response times for amplitude modulation and timbre ($p<0.05$). There were no other pairwise differences.

4.2.3.3 Qualitative Data

Participants were presented with examples of the four different types of audio roughness and were asked which version that he/she felt matched the tactile feedback best. Participants were also asked to explain their answers for each of these questions. The quantitative results of the questionnaires are shown below:

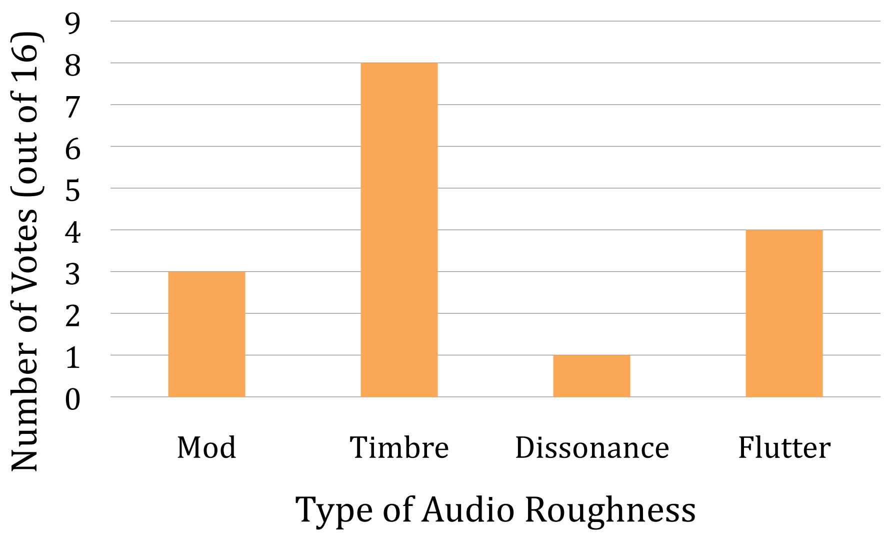

Figure 4-8: Total number of votes of preference from participants per audio roughness condition.

The answers to the questionnaire show that timbre was the form of audio roughness that best matched the tactile roughness with 8 out of 16 votes while dissonance was the least preferred with only 1 vote.

4.2.4 Discussion

It has been shown that using timbre or amplitude modulation produces better results than flutter tonguing and dissonance. This suggests that participants found it easier to match audio and tactile cues when the audio cues used amplitude modulation or timbre. Initial results show that subjects preferred the use of differing timbres in audio. However, the results also show no significant difference in performance between timbre and audio amplitude modulation. These results suggest that crossmodal roughness in the auditory domain should be created using either amplitude modulation or differing timbres.

4.3 Experiment 1b: Tactile Roughness

The next experiment followed up on the results from the initial roughness experiment above with an investigation of different techniques (amplitude modulation, frequency and waveform) for creating roughness in the tactile domain.

Brown *et al.* [20] have conducted experiments showing recognition rates of over 95% for both rhythm and spatial location, indicating that tactons are a successful means of communication through the tactile modality. However, the individual results for tactile roughness (created using amplitude modulation) show a recognition rate of just 57.2%, suggesting that such a design is not effective and an alternative is needed. The study presented here investigates different representations of tactile roughness to see how the recognition rates of these new parameters compare to those achieved by amplitude modulation. If the recognition rates can be increased for tactons, this improvement will also enhance crossmodal icon design.

4.3.1 Stimuli

4.3.1.1 Amplitude Modulation

Tactile roughness as used before in tacton design [20] is created by using amplitude modulated sinusoids. These are created by multiplying a sine wave of a given frequency by a sine wave of another frequency. The roughness levels used previously [20] and also in this experiment were: an unmodulated 250Hz sine wave (smooth), the same sine wave modulated by 50Hz (rough), and by 30Hz (very rough), see Figure 4-4.

4.3.1.2 Frequency

There are conflicting views as to whether using different frequencies to create different textures is an appropriate parameter for tactons. On the one hand, because the frequency range of the skin is only from 10Hz to 400Hz, and the usable frequency range is further reduced by the limited bandwidth of standard actuators, frequency is unsuitable as a parameter in tacton design [14]. However, on the other, studies have shown that frequency can still play a role in tactile texture as subjects in psychophysical experiments have reported a sensation of periodicity or buzzing at low frequencies (below 100Hz) while at higher frequencies a more diffuse, smooth sensation is perceived [150]. Furthermore, different frequencies have been used in experiments with multi-finger tactual displays [139] where it was shown that participants could categorise frequencies into three perceptually distinct groups over the range of DC to 300Hz. Therefore, in the study presented here, the frequency levels used were based on the results from the multi-finger tactual display experiments. The levels used were: 6 Hz (slow motion, very rough), 70Hz (fluttering slightly faster motion, rough), and 250 Hz (smooth) as shown below.

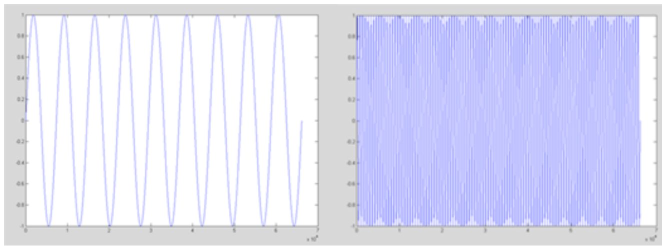

Figure 4-9: 6Hz sine wave and 70Hz sine wave.

4.3.1.3 Waveform

Like frequency, there are also conflicting views as to whether using different waveforms to create different textures is an appropriate technique for tacton design. Originally it was decided that, although users can differentiate between sine waves and square waves, the number of different values that could be encoded in would be limited [14]. However, it has been shown by Miller [99] that it is better to have a small number of values for several attributes of a stimulus set as opposed to having many values for one attribute of the stimuli. Furthermore, the vibrotactile range from pure sine tone to noise is often described as a continuous transition from smoothness to roughness [120]. So, in tacton parameter design, waveform can be correlated to the 'texture' of tactile stimuli. Also, waveform (timbre) is a key attribute in earcon design [8] and, if tactons are to be used in crossmodal applications with earcons, it is an important parameter to investigate.

Initial pilot studies with 6 participants showed they could distinguish between a sine wave (smooth), and sawtooth wave (rough), and a square wave (very rough). Therefore, in this experiment a sine wave, square wave and sawtooth wave were used. The square waves were created using the Fourier series made up of the sum of odd harmonics of sine waves. When adding harmonics, it was ensured that the amplitude levels created by each harmonic were always within the 250Hz resonating frequency range of the actuator used (the C2).

4.3.2 The Experiment

The aim of this experiment was to investigate alternative representations of tactile textures (based on frequency and waveform) for use in tacton design and to examine the recognition rates of these new parameters in comparison with those achieved by tactile roughness (amplitude modulation). The hypotheses were:

1. There will be a difference in participants' ability to recognise three different levels of texture in cues using amplitude modulation, frequency, or waveform.
2. Participants will be able to distinguish between the three different textures created by the three different waveforms and three different frequency levels (over 90% correct identification).

4.3.3 Methodology

Nine people took part in the experiment, aged between 20 – 36 years, 4 female and 5 male, and all members of staff or students at the University of Glasgow. The experimental method used was a within-subjects design where each participant was tested on all three conditions – amplitude modulation, frequency and waveform.

There were 54 tasks in this experiment, 3 different rhythms (see Figure 4-10) were used with each of the three conditions – amplitude modulation (rhythms 1, 2, and 3 made up of a 250Hz unmodulated sine wave, a 250Hz sine wave modulated at 50Hz, and one modulated at 30Hz each repeated twice), frequency (rhythms 1, 2, and 3 made up of a 250Hz sine wave, a 70Hz sine wave, and a 6Hz sine wave each repeated twice), and waveform (rhythms 1, 2, and 3 made up of a sine, square, and sawtooth wave each repeated twice). The tactons each lasted approximately 1 – 1.5 seconds and rhythms contained at least one minim (a longer pulse of 500ms in this case).

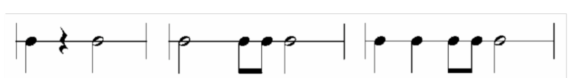

Figure 4-10: Rhythms 1, 2 and 3 used in experiment 1 (crossmodal roughness texture).

The tactons represented cues which might occur on a mobile phone to inform the user of the urgency of incoming alerts. This representation was based on a similar methodology used by Brown *et al.* [20] providing a baseline for comparison. The urgency of the alerts was encoded in the texture with the three different levels of texture (very rough, rough and smooth) created by amplitude modulation, frequency or waveform mapped to the urgency of the alert (very urgent, urgent or not urgent). In each task participants were presented with a tacton and asked to identify the corresponding alert. The stimulus was presented four times and the participant could respond at any time by selecting the corresponding button in the experimental software (Figure 4-11).

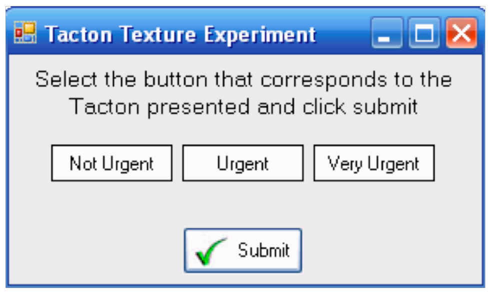

Figure 4-11: Screenshot of experiment interface.

Before beginning the experiment participants took part in a training session to introduce them to the concept of tactons, texture, rhythm, etc. Participants were then allowed to familiarise themselves with each of the different types of tactons for ten minutes before beginning the actual tasks.

4.3.4 Results

In this study, the experimental software recorded data on the participants' responses to each stimulus (raw data in Appendix B). From these results percentage correct scores were calculated for each stimulus (Figure 4-12).

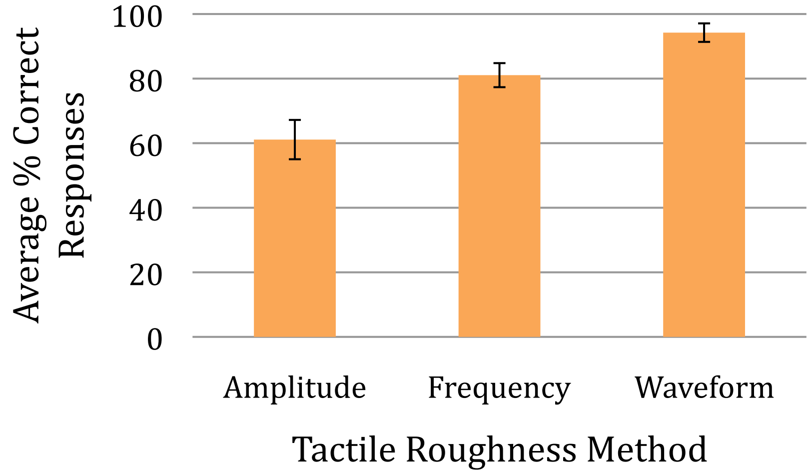

Figure 4-12: Average percentage correct scores for each tactile roughness technique (with standard deviations).

The average recognition rates for waveform were 94.2% with 81% for frequency; whereas the average recognition rates for amplitude modulation were only 61.1% (similar to that found by Brown [20]) therefore Hypothesis 2 can be partially accepted as recognition rates for waveform are over 90% but not for frequency. Participants were interviewed after taking part in the experiment and several indicated that they found frequency quite difficult, as the cues were not long enough to allow them to distinguish the different frequencies. This would suggest that rhythms with longer notes may improve results for the identification of tactile texture using frequency. However, this may not always be practical in a usage context if rapid communication is needed.

To test Hypothesis 1, first the significance of the effects of each representation of tactile texture was investigated. The statistical analysis used is a standard repeated measures 1-factor ANOVA, based on the critical values of the F distribution, with alpha = 0.01. The ANOVA showed significant differences in the error data between tactile texture conditions ($F(2,26) = 31.72$, $p < 0.01$). Tukey's pairwise HSD analysis showed that the average number of errors for amplitude modulation was significantly greater than the number of errors in frequency and waveform. The analysis also showed that the average number of errors for frequency was significantly greater than the number of errors in waveform. Overall, Hypothesis 1 can be accepted.

In a small post-study questionnaire, all participants agreed that waveforms were much easier to recognise than the other designs. The standard deviations shown in Figure 4-12 are also extremely small indicating a level of consistency between participants.

On the whole, the recognition rates of tactons using amplitude modulation are slightly higher than in previous experiments [20] but still produce poor results compared to waveform and frequency. The results of this experiment indicate that using different waveforms to represent tactile roughness as a parameter in tactons would be more effective than using amplitude modulation.

4.3.5 Discussion

This study investigated perception of tactile roughness with a view to identifying the best technique (amplitude modulation, differing waveforms, or differing frequencies) to use when including roughness as a parameter in the design of tactons and in the design of crossmodal icons. The results, with recognition rates of 94.2% for differing waveforms, 81% for differing frequencies, and 61.1% for amplitude modulation, indicate that users can identify and distinguish differing waveforms significantly more effectively than amplitude modulation and frequency. Therefore, different waveforms can be used as the roughness parameter in tacton design.

Previous tacton design has used amplitude modulation to create the roughness parameter but accuracy was not high enough for reliable use especially when being used in combination with crossmodal audio texture. Given that using differing waveforms produces high recognition rates, by changing the technique used to create crossmodal texture, overall recognition rates for 3-dimensional crossmodal tactons could reach levels closer to 100%.

The number of available usable parameters in the tactile domain is limited and using tactile roughness as a parameter produced low recognition rates but the results of this study have shown that tactile roughness created with different waveforms could be a very successful parameter for crossmodal tacton design.

4.4 Experiment 1c: Mapping Tactile Waveforms to Audio Roughness

Given the results of the previous experiment it was necessary to re-run the crossmodal roughness matching experiment with the new version of tactile roughness created with different waveforms. The set-up of the experiment was the same in almost every respect except this time, participants were asked to select the best audio match to different tactile waveforms instead of tactile stimuli with different levels of amplitude modulation. This was to ensure that a good crossmodal match between the modalities was still possible.

As before, there were four conditions in this experiment: audio roughness created with dissonance, flutter-tonguing, amplitude modulation and differing timbres. The aim of the experiment was to determine which of these versions of audio roughness map best to tactile roughness using a sine, sawtooth and square wave.

4.4.1 Hypothesis

3. That audio roughness can be matched with tactile roughness.

4.4.2 Methodology

16 new participants (aged between 22 and 37, 9 male and 7 female, all students at the University of Glasgow and all right-handed) were presented with an audio or tactile cue and then asked to select the equivalent cue from the choices given. Participants were able to play the choices as many times as they wished.

The online system recorded three dependent variables: the time taken for the participant to answer the question, the number of times each example was played, and the correctness of each answer. The independent variable was the different audio/tactile versions of the tactile/audio parameter. The online system used by participants was exactly the same as Experiment 1a and is shown in Figure 4-5.

4.4.3 Results

The average number of errors and the average response time for the four audio roughness conditions are shown in Figure 4-13 and Figure 4-14.

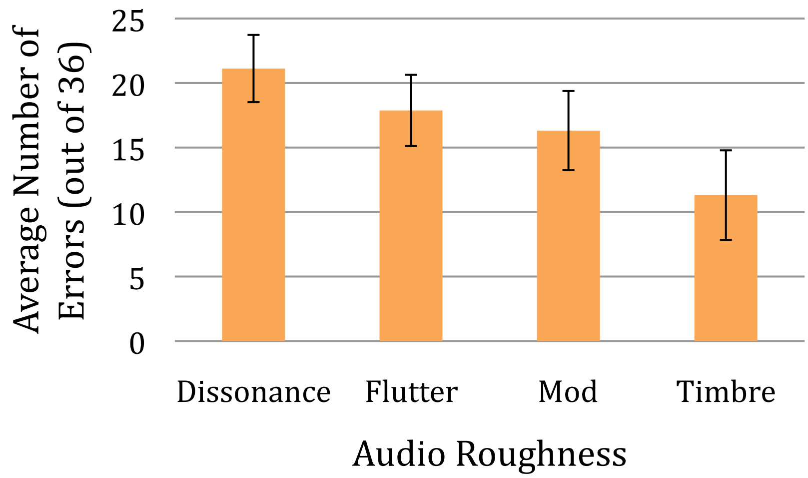

Figure 4-13: Average number of errors for each audio roughness condition (with standard deviations).

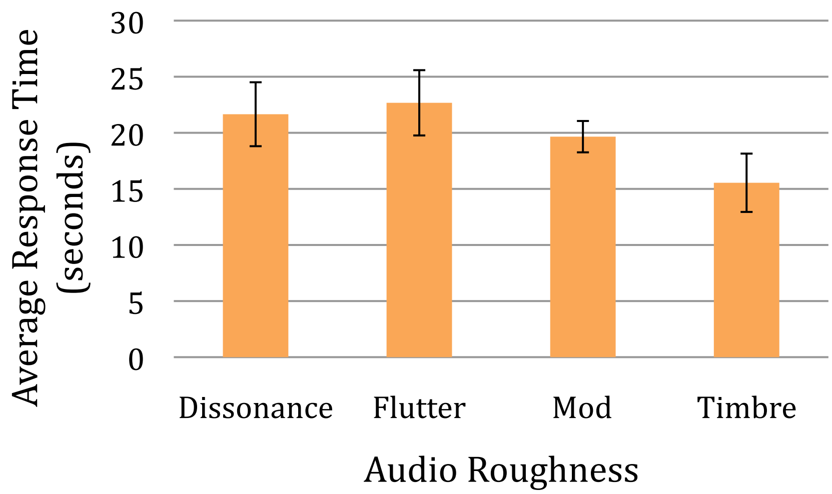

Figure 4-14: Average response time per condition (with standard deviations).

4.4.3.1 Errors

A 1-factor repeated measures ANOVA on the mean number of errors made by each participant for each condition showed there are significant differences in the error data between audio roughness conditions ($F(3,60) = 7.44$, $p = 0.0003$). Tukey's pairwise analysis showed that the average number of errors for dissonances, amplitude modulation and flutter tonguing was significantly greater than the number of errors in timbre ($p<0.05$). There were no other pairwise differences.

4.4.3.2 Response Time

There are significant differences in the mean response time data between questions ($F(3,60) = 8.28$, $p = 0.0001$). Tukey's pairwise analysis showed that the average response time for dissonances, amplitude modulation and flutter tonguing was significantly greater than the response times for timbre ($p<0.05$). There were no other pairwise differences.

4.4.3.3 Qualitative Data

Participants were presented with examples of the four different types of audio roughness and were asked which version that he/she felt matched the tactile feedback best. Participants were also asked to explain their answers for each of these questions. The quantitative results of the questionnaires are shown below:

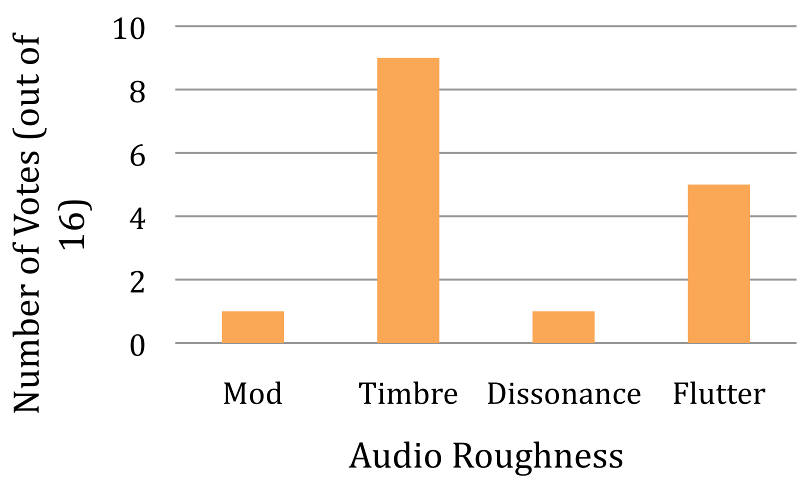

Figure 4-15: Total number of votes of preference from participants per audio roughness condition.

The answers to the questionnaire show that timbre was the form of audio roughness that best matched the tactile roughness with 9 out of 16 votes while dissonance and amplitude modulation were the least preferred with only 1 vote each.

4.4.4 Discussion

This time, it has been shown that using timbre produces better results than flutter tonguing, amplitude modulation and dissonance. This suggests that participants found it easier to match audio cues to tactile cues using waveforms when the audio cues used timbre.

4.5 Experiment 2a: Crossmodal Spatial Location

Spatial location has been used as a successful parameter in tactons research [20]. Three positions on the forearm were used to present different time values to users for example, the actuator on the wrist represented 5 minutes, the middle actuator represented 10 minutes and the actuator at the elbow was 15 minutes. As mentioned in the literature review, the skin has a surface area of 1.8 m^2 meaning that there is a large amount of space on which actuators can be placed. In the audio domain, 3D audio positions have been used with earcons to separate earcons during concurrent presentation [95] but it has not been used as a parameter in which to encode data. Although audio is somewhat more limited than tactile in terms of spatial location, 3-dimensional audio can be used to place multiple sounds around the head.

Given that spatial location is another amodal parameter, it may be possible for users to perceive an auditory equivalent of using body locations for tactile feedback. The next experiment in this PhD research investigated ways to map from a tactile location on the body to an audio location in a soundscape and *vice versa*. This would allow crossmodal cues to be presented via earcons or tactons as in Figure 4-16.

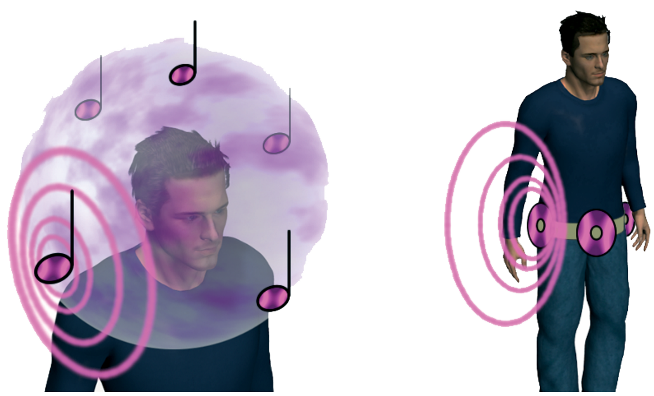

Figure 4-16: Audio cue indicated by audio panned to the right of a 3D soundscape and a tactile cue indicated by a vibrotactile pulse on the right of the waist.

Firstly, the choice of body location for the tactile cues is important. The spatial location of actuators on the body has been studied by many researchers [33] [144]. In order to use spatial location as a parameter in crossmodal interaction, it is important to choose the body locations carefully. Cholewiak and Collins [33] report that tactile localisation is most accurate when the stimulus is close to an anatomical reference point, and in particular at points of mobility such as the wrist or elbow.

A 3D audio system has the ability to position sounds all around a listener. The sounds are actually created by the loudspeakers (or headphones), but the listener's perception is that the sounds come from arbitrary points in space. A sound is placed in the horizontal plane by convolving the sound with recorded head-related impulse responses [45]. Using HRTFs and reverberation, the changes of sound on its way from the source (including reflections from walls and floors) to the listener's ear can be simulated. These effects include localisation of sound sources behind, above and below the listener. Using 3D audio can increase the information content of an audio display and also allow the spatial nature of the audio space to be used.

4.5.1 Experiment Design

An experiment was conducted to determine which body location can be mapped most effectively to locations in a 3D audio soundscape. The aim of this experiment was to investigate whether spatial locations in the audio and tactile domain can be matched and therefore used in crossmodal interaction.

The version of the system (Figure 4-17) used in the experiment took the form of a computer-controlled belt/wrist band/ankle band with four embedded vibrotactile actuators: each of the small actuators were evenly spaced around the circumference of the body area (waist, wrist or ankle) and mapped to spatial audio played through a pair of headphones.

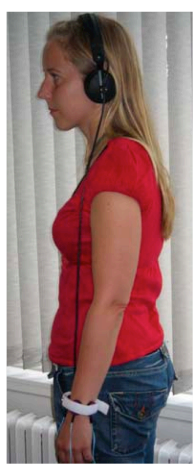

Figure 4-17: Computer-Controlled wristband with four embedded vibrotactile actuators plus headphones.

The audio cues used in this experiment were created using the AM:3D[11] audio engine and were placed on the horizontal plane around the user's head at the height of the ears to avoid problems related to elevation perception. The sounds were located every 90° starting from the nose. This resulted in a 2.5D planar soundscape.

There were three conditions in this experiment, they were:

Waist – four actuators were placed at cardinal points around the waist of the participant, the waist was chosen because it has been identified as an effective body location for tactile perception and studied extensively by researchers such as Cholewiak, van Erp and van Veen [33] [144].

Ankle – four actuators were placed at cardinal points around the ankle of the participant. The ankle was chosen because it is an anatomical reference point with enough surface area to support four actuators and suggested by van Erp in his work on tactile navigation displays [144].

Wrist – four actuators were placed at cardinal points around the wrist on the non-dominant arm of the participant. The wrist was chosen because it is an anatomical reference point as suggested by Cholewiak [33].

4.5.2 Hypothesis

1. Participants will be able to recognise equivalent spatial locations in an audio soundscape when given a body location and *vice versa*.

4.5.3 Methodology

18 participants (aged between 19 and 30, 11 male and 7 female, all students or staff at the University of Glasgow) were presented with an audio or tactile cue and then asked to select the equivalent cue from the choices given (Table 4-2), in a similar fashion to the first experiment in this chapter. Participants were able to play the choices as many times as

[11] AM:3D Positional Audio, http://www.am3D.com/

they wished. Participants were aged between 20 and 31 and were students or staff at the University of Glasgow.

Modality	Cue Presented to Subject	Choices
Tactile	Waist North	Audio north, south, east or west
Tactile	Waist South	Audio north, south, east or west
Tactile	Waist East	Audio north, south, east or west
Tactile	Waist West	Audio north, south, east or west
Tactile	Ankle North	Audio north, south, east or west
Tactile	Ankle South	Audio north, south, east or west
Tactile	Ankle East	Audio north, south, east or west
Tactile	Ankle West	Audio north, south, east or west
Tactile	Wrist North	Audio north, south, east or west
Tactile	Wrist South	Audio north, south, east or west
Tactile	Wrist East	Audio north, south, east or west
Tactile	Wrist West	Audio north, south, east or west
Audio	North	Waist north, south, east or west
Audio	South	Waist north, south, east or west
Audio	East	Waist north, south, east or west
Audio	West	Waist north, south, east or west
Audio	North	Ankle north, south, east or west
Audio	South	Ankle north, south, east or west
Audio	East	Ankle north, south, east or west
Audio	West	Ankle north, south, east or west
Audio	North	Wrist north, south, east or west
Audio	South	Wrist north, south, east or west
Audio	East	Wrist north, south, east or west
Audio	West	Wrist north, south, east or west

Table 4-2: Cues Presented to Participants During Experiment and Answer Choices Available.

For example, the participant was presented with a north waist tactile cue. The choices presented were four different 3D audio samples. Participants had to pick the sound he/she thought matched best with the tactile version. Or, a south 3D audio cue was presented and participants had to choose between four different tactile cues placed around the waist.

First, the participants took part in a training session to introduce them to the relevant terminology (3D audio, cardinal points, tactile etc.) and then shown a worked example (online), which contains example tactile and audio cues, example questions and their correct answers.

Next, the participants began the experiment using an online system. The first block of questions displayed by the system was a practice set although the participants were unaware of this. The order of the tasks was randomised. Each question was multiple choice and the participants selected a checkbox. The online system recorded three dependent variables: the time taken for the participant to answer the question, the number of times each example is played, and the correctness of each answer. The independent variable was the different audio/tactile versions of the tactile/audio parameter. The experimental method embodied a within-subjects design, as above. Therefore, each participant performs tasks related to all conditions.

4.5.3.1 The Experiment Parameters

- Each task had a time limit of 30 seconds.
- Each task was shown twice - once with tactile choices and once with audio choices.
- A break was given to the participant after every 8 tasks.
- A practice set of 8 tasks was presented at the beginning of the experiment.

4.5.3.2 Online System

The online system used by participants is shown below in Figure 4-18.

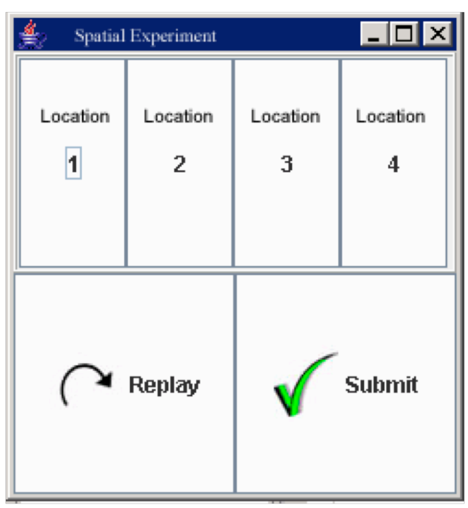

Figure 4-18: Screenshot of experiment interface used to present audio and tactile cues via different spatial locations.

4.5.4 RESULTS

The average number of errors and the average response time for the three tactile body location conditions are shown in Figure 4-19 and Figure 4-20 (raw data in Appendix C).

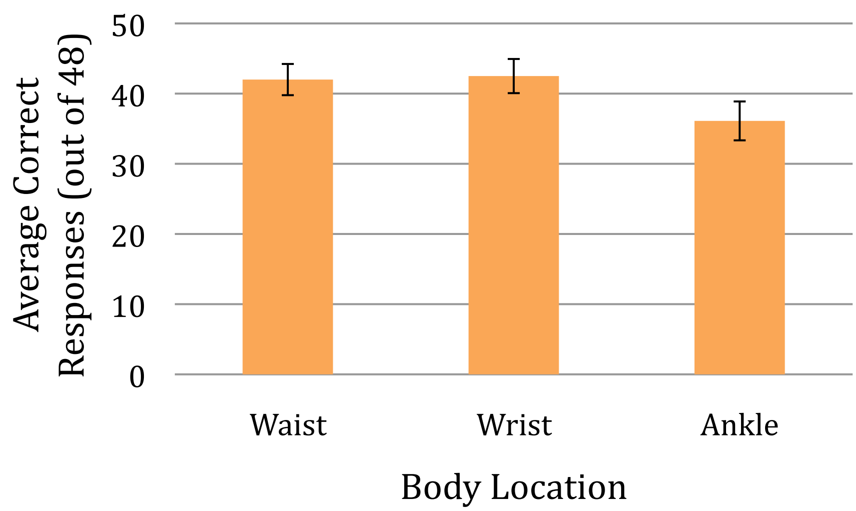

Figure 4-19: Average number of correct responses for each body location (with standard deviations).

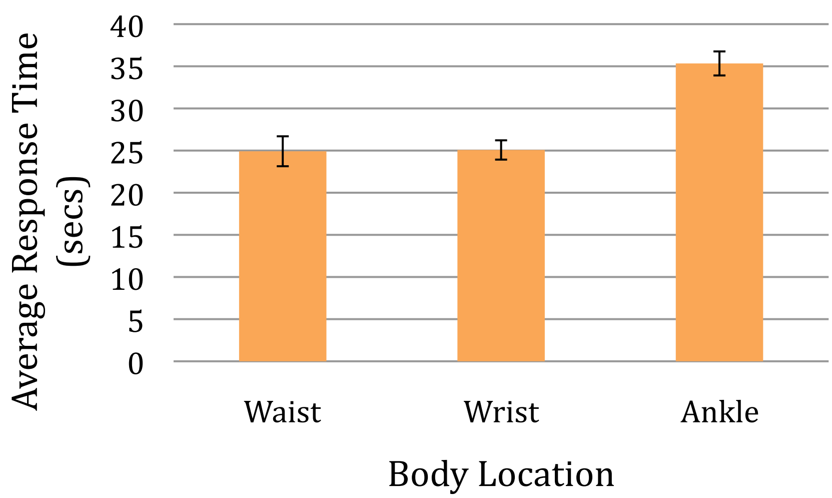

Figure 4-20: Average response time in seconds for each body location (with standard deviations).

To test the hypothesis, first the significance of the effects of each tactile body location condition was investigated. The statistical analysis used here is a standard 1-factor repeated measures ANOVA, based on the critical values of the F distribution.

4.5.4.1 Errors

There are significant differences in the error data between tactile body location conditions ($F(2,51) = 9.2$, $p = 0.0004$). Tukey's pairwise analysis showed that the average number of errors for the ankle was significantly greater than the number of errors in the waist and wrist ($p = 0.05$). There were no other pairwise differences.

4.5.4.2 Response Time

There are significant differences in the response time data between tactile body locations ($F(2,51) = 74.1$, $p = 0.0001$). Tukey's pairwise analysis showed that the average response time for the ankle was significantly greater than the response times in the waist and wrist ($p = 0.05$). There were no other pairwise differences.

4.5.4.3 Qualitative Data

Participants were presented with examples of the three different tactile body locations and were asked which version he/she felt was most comfortable and easiest to match with the audio equivalent. They were also asked to explain their answers for each of these questions. The quantitative results of the questionnaires are shown below:

Figure 4-21: Total number of preference votes for each body location (out of 18).

The majority of participants (66%) found the wrist to be comfortable and easiest to match with the 3D audio soundscape. No participants reported the ankle to be easy to match or particularly comfortable.

4.5.5 Discussion

Results show that participants are able to map the presented 3D audio positions to tactile body positions on the waist and wrist most effectively and that there are significantly more errors made when using the ankle. Although there is no significant difference between the waist and the wrist, participants indicated preference for the wrist.

4.6 Experiment 2b: Mobile Crossmodal Spatial Location

Users of mobile devices are often in motion when using their devices (for example, receiving calls, sending text messages, etc.). Interfaces must be designed to work well under these circumstances too, not just when the user is stationary.

Given the promising results of the stationary spatial location experiment, the same experiment was conducted again in a mobile situation to see if motion affects the results. There are many ways in which motion could affect perception of crossmodal output: mobile environments tend to change frequently, the user's main attention may be on safety whilst crossing a road instead of the mobile device, a user can become physically tired, and during natural motion such as walking, a user's hands are likely to be moving.

The setup of this experiment was identical to the previous one in every respect except that a different set of participants were used and this time participants were asked to walk on a treadmill during the experiment as opposed to sitting in a chair as shown in Figure 4-22. There were 16 participants (8 male and 8 female) aged between 20 and 29, all staff or students at the University of Glasgow with no physical impairments.

Figure 4-22: Experiment set-up.

This mobile experiment used a treadmill set up in a usability lab to simulate mobility because the tactile actuators used were not wireless and were controlled from a PC and therefore inappropriate for use in a real mobile environment. Furthermore, using a treadmill permitted the experimenter to set a standard speed for all participants (in this case, all walked at a speed of 6km per hour). The main hypothesis was that being mobile would increase errors produced during spatial location identification and matching between modalities as compared to being stationary.

4.6.1 Results

The average number of correct responses and the average response time for the three tactile body location conditions are shown in Figure 4-23 and Figure 4-24 (raw data is included in Appendix D).

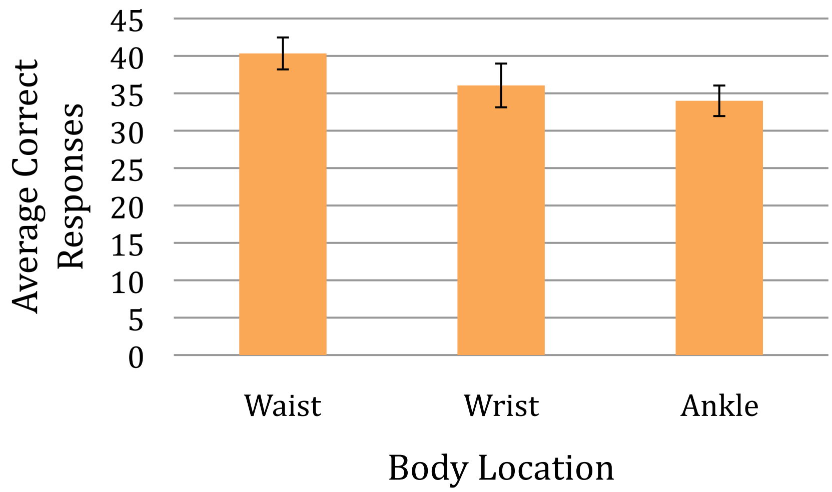

Figure 4-23: Average correct responses for each body location when mobile (with standard deviations).

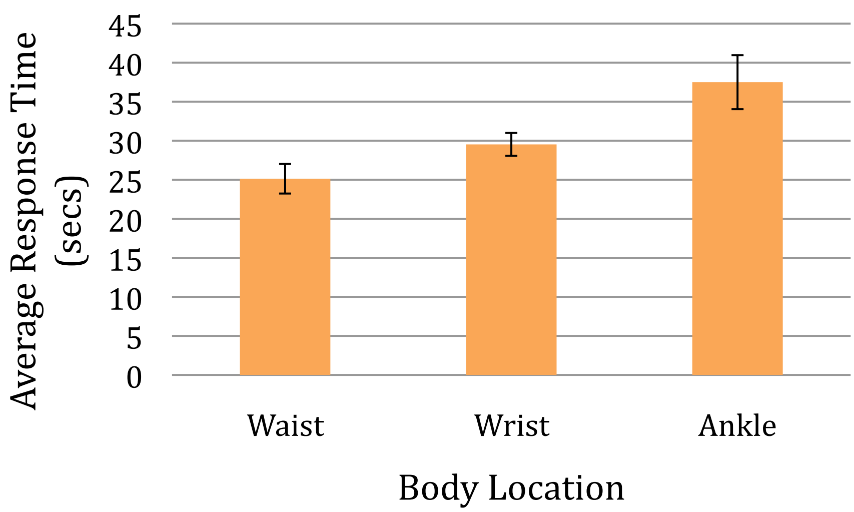

Figure 4-24: Average response time in seconds for each body location when mobile (with standard deviations).

As before, the significance of the effects of each tactile body location condition whilst mobile was investigated.

4.6.1.1 Errors

The average errors for each experiment are shown below in Figure 4-25. In order to establish significant differences in the data between the stationary experiment and the mobile experiment a 2-factor mixed design ANOVA using the three conditions of body location and stationary/mobile as the two factors was applied. The results of the ANOVA show there are significant differences in the error data between tactile body location conditions ($F(2,68) = 12.7$, $p < 0.01$). Tukey's pairwise analysis showed that the average number of errors for the ankle and wrist was significantly greater than the number of errors in the waist ($p = 0.05$). In terms of stationary and mobile environments, the results show a significant difference in the number of errors ($F(1,17) = 9.72$, $p < 0.01$). Once again, a Tukey test was performed and the analysis showed that the average number of errors when mobile is significantly higher than when stationary ($p = 0.05$).

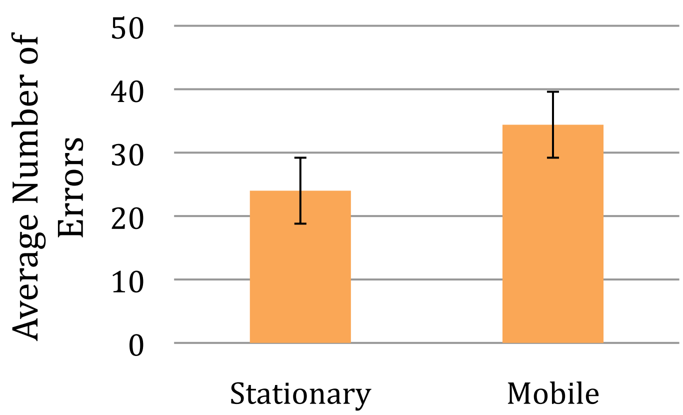

Figure 4-25: Average number of errors in static and mobile conditions (with standard deviations).

4.6.1.2 Response Time

There are significant differences in the response time data between tactile body locations ($F(2,68) = 73.2$, $p < 0.01$). Tukey's pairwise analysis showed that the average response time for the ankle and wrist was significantly greater than the response times in the waist ($p = 0.05$). There were no other pairwise differences.

4.6.1.3 Qualitative Data

Participants were presented with examples of the three different tactile body locations and were asked which version he/she felt was most comfortable and easiest to match with the audio equivalent. They were also asked to explain their answers for each of these questions. The quantative results of the questionnaires are shown below. This time the majority of participants (62.5%) chose the waist.

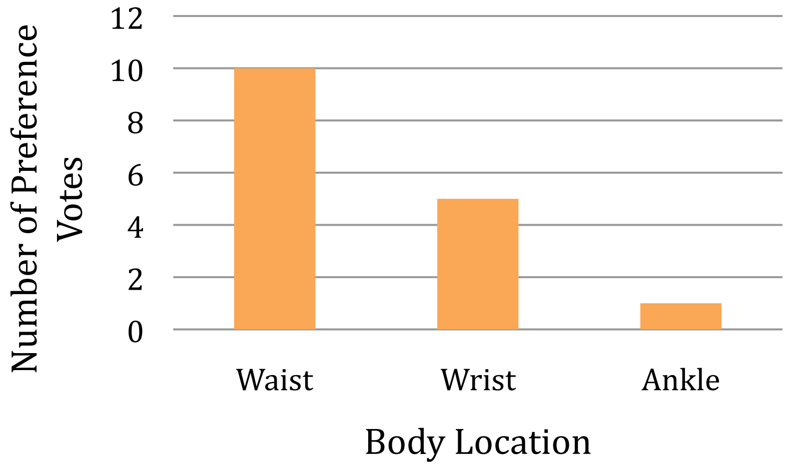

Figure 4-26: Number of preference votes for each body location when mobile.

4.6.2 Discussion

Results show that participants are able to map the presented 3D audio positions to tactile body positions on the waist most effectively when mobile and that there are significantly more errors made when using the ankle or wrist. Unlike the previous experiment, a greater number of participants preferred the waist to the wrist or ankle. However, a greater number of participants still preferred the wrist to the ankle.

The reason why the wrist performed worse in the mobile experiment compared to the static experiment could be that motion naturally changes the orientation of the wrist as the arm swings and therefore it is more difficult to match locations when they are constantly moving. For example, if we take a clock face analogy, an actuator is placed on the left hand side of the wrist to represent 0900 but as the wrist rotates during movement the actuator is no longer at the 0900 position (the wrist position can rotate anywhere between 0800 and 1400 approximately).

4.6.3 Orientation

In order to establish whether the natural rotation of the wrist whilst walking confuses the interpretation of tactile cues presented there, a further condition was tested where the arm was placed in a splint (one designed to immobilise the wrist in case of sports injuries) so that the wrist was unable to rotate (see Figure 4-27). The experiment was otherwise the same as the mobile study. Once again, a new set of 16 participants was recruited from the

University of Glasgow (7 male and 9 female) aged between 19 and 34 with no physical impairments.

Figure 4-27: Wrist splint used to prevent orientation[12].

The results showed that a mobile user with a splinted wrist produced 42% fewer errors than with an unconstrained wrist. Overall, when the wrist was splinted, an ANOVA showed that there was no significant difference between the results of the mobile and static wrist condition with the wrist producing 71% correct crossmodal matches with the audio cues. This suggests that wrist rotation does cause problems, and if spatial location were to be used as a crossmodal parameter such locations would have to be avoided. An alternative would be to track the wrist's rotation and then display vibrations on the actuator pointing in the appropriate direction.

These experiments have established that it is possible for users to perceive spatial location as equivalent in both the auditory and tactile domains. Furthermore, it has been confirmed that the use of the waist as a tactile body location produces significantly better results than using the wrist or ankle. When using crossmodal spatial locations in mobile displays these experiments have shown that the wrist performs badly due to the natural rotation that takes place in motion so it is best to use the waist which, in this case, produces 76% accuracy whilst stationary and 72% accuracy whilst mobile.

4.7 Conclusions

This chapter reported two main studies (with two follow-up studies) investigating some possible parameters and mappings that can be used to facilitate crossmodal auditory/tactile

[12] LP Supports - http://www.lp-supports.com/products/tennis_elbow_and_elbow_supports/

feedback. The experiments conducted investigated rhythms with texture and spatial location as potential parameters for crossmodal icons.

The first experiment detailed in this chapter investigated crossmodal texture, namely roughness. Participants matched audio roughness to tactile roughness significantly more often when using audio amplitude modulation or timbre compared to flutter tonguing or dissonance. Qualitative results showed that subjects preferred the use of differing timbres in audio. However, the results also showed no significant difference in performance between timbre and audio amplitude modulation. Therefore crossmodal roughness in the auditory domain should be created using either amplitude modulation or differing timbres to ensure a suitable design for use with crossmodal icons. This was the first of the experiments that looked at matching and perceived equivalence to answer Research Question 1. Amplitude modulation was expected to perform well in the experiment given that it is a direct mapping to the tactile versions of roughness. However, more interestingly timbre produced comparable results to amplitude modulation and timbre is not a direct mapping to tactile amplitude modulation. This shows that the methods of creating the parameters do not have to be identical in both modalities in order for the icons to be perceived as equivalent.

Following on from this study, a further experiment was conducted to investigate other forms of tactile texture for use in crossmodal icons as an alternative to amplitude modulation. Previous tacton design has used amplitude modulation to create the roughness parameter but accuracy was not high enough for reliable use. The experiment investigated perception of tactile roughness with a view to identifying the best technique (amplitude modulation, differing waveforms, or differing frequencies) to use when including roughness as a parameter in the design of tactons. The results, with recognition rates of 94.2% for differing waveforms, 81% for differing frequencies, and 61.1% for amplitude modulation indicate that users can identify and distinguish differing waveforms significantly more effectively than amplitude modulation and frequency. Therefore, different waveforms (sine, square and sawtooth) can be used as the roughness parameter in tacton design. A small study was conducted to ensure that audio roughness using amplitude modulation or timbre was still perceived as a match to this new form of tactile roughness. The results were comparable to those obtained in the earlier experiment. Therefore, providing a more robust crossmodal parameter than previous versions. This experiment investigated three of the main waveforms. There are many other waveforms that could be created for different textures and it could be possible that similar results would be achieved. In combination with the high number of audio timbres available, this would lead to a huge number of potential crossmodal textures.

The next potential parameter under investigation was spatial location. An experiment was conducted to determine whether tactile locations on the body can be mapped to an audio location in a soundscape and *vice versa.* The results showed that participants were able to map the presented 3D audio positions to tactile body positions on the waist and wrist most effectively and that there are significantly more errors made when using the ankle. Although there was no significant quantitative difference between the waist and the wrist, participants indicated preference for the wrist.

This experiment was also conducted in a simulated mobile setting: walking on a treadmill. Once again, the results showed that participants were able to map the presented 3D audio positions to tactile body positions on the waist most effectively when mobile and that there were significantly more errors made when using the ankle or wrist. The reason why the wrist performed worse in the mobile experiment compared to the static experiment could have been due to motion naturally changing the orientation of the wrist as the arm swings, and therefore making more difficult to match locations when they are constantly moving. However, when adding a splint to the wrist this issue was resolved. Unfortunately, this may not be practical for real world use.

The spatial location experiments have shown that it is possible for users to match spatial locations in both the auditory and tactile domains. Furthermore, it has been confirmed that the use of the waist as a tactile body location produces significantly better results than using the wrist or ankle. When using crossmodal spatial locations in mobile displays these experiments have shown that the wrist performs badly due to the natural rotation that takes place in motion so it is best to use the waist. Given that commercially available mobile devices now incorporate stereo speakers, it may be possible to present 3D audio feedback without using headphones making spatial location even more practical in real world usage while belts or watches could be used to present spatially located tactile feedback. The experiment discussed in this section made use of four spatial locations but given the amount of surface area provided by our skin and the capabilities of 3D audio software to create many feedback sources in a soundscape, this parameter has the potential to provide a very large number of different spatial locations thus increasing the amount of data that may be transmitted.

In answer to Research Question 1:

> **RQ1**: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

This chapter has confirmed that it is possible to create crossmodal icons using the amodal attributes of the audio and tactile modalities. The experiments indicate that rhythm can be used to encode data in the audio and tactile modalities and those rhythms in both modalities can be matched. When crossmodal roughness is created using audio timbre and tactile waveforms, users perceive a match between the data in both modalities. Finally, spatial location can be used as a crossmodal parameter in both static and mobile settings. Experiments showed that spatial location can be perceived as synonymous when using a 3D audio soundscape and tactile locations on the waist. Given the success of these experiments, the next step in this research is to combine these parameters to create a set of multi-dimensional crossmodal icons as discussed in the next chapter.

Chapter 5 Multi-Dimensional Crossmodal Icons

The previous two chapters reported the results of an extensive literature review of crossmodal interaction, outlining the potential of amodal attributes as parameters and several studies investigating the mapping of rhythm, roughness and spatial location between modalities. One of the aims of this research is to include crossmodal icons in various touchscreen applications. These applications may require multiple dimensions of data to be encoded in crossmodal icons. As discussed in Chapter 4, there are now three possible parameters that allow easy mappings between the auditory and tactile modalities. However, these parameters have never been combined to create multi-dimensional crossmodal icons and so far there have been no tests conducted in related research to test whether the concept of crossmodal interaction works i.e. whether users can transfer knowledge of icon meanings between senses. Therefore, this chapter discusses an experiment using a multi-dimensional set of crossmodal icons. The experiment focuses on the extent to which learning can be transferred between the two modalities (for example, can users who have been trained to identify three-dimensional crossmodal earcons transfer their knowledge to the tactile domain and identify the corresponding crossmodal tactons?). The experiment finishes by testing recognition rates during an absolute identification and matching experiment for the resulting crossmodal icons.

Section 5.1 describes the initial icon design with three dimensions of data encoded in three crossmodal auditory/tactile parameters (examples of multi-dimensional crossmodal icons can be found in Appendix I). Section 5.2 covers the details of the experiment conducted to investigate absolute identification of audio and tactile crossmodal icons when a user is trained in one modality and tested in the other (and given no training in the other modality) to see if knowledge can be transferred between modalities. Performance levels in the experiment were compared when users were static and mobile to reveal any effects that mobility might have on recognition of the cues.

This chapter addresses two of the research questions posed at the start of the thesis, in terms of the design of three-dimensional crossmodal icons.

RQ1: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

RQ2: What levels of performance can be achieved when these parameters are combined to create multi-dimensional crossmodal icons?

Research Question 1 is addressed through a discussion of suitable parameters for three-dimensional crossmodal icons and through the evaluation of these individual parameters when used in combination. Research Question 2 is addressed through the transfer of learning and identification rates obtained in the experiments detailed in Sections 5.2, 5.3 and 5.4. Section 5.6 concludes this chapter, drawing general conclusions from this work, discussing how the findings of these experiments answer the research questions posed in this thesis.

5.1 Designing Multi-Dimensional Crossmodal Icons

In this study, crossmodal icons were created to represent alerts which might occur on a mobile phone to inform the user of incoming messages. Three message attributes were encoded in each crossmodal icon using the parameters identified in Chapter 4: the type of message was encoded in the rhythm, the urgency of the message was encoded in the roughness and the sender of the message was encoded in spatial location. These types of information were chosen as they are common alerts provided through the visual modality on current mobile devices and would be familiar to participants. The type of message had three possible values: text, email, or voicemail, the urgency of the message had two possible values: urgent or not urgent, and the sender of the message had three possible values: work, personal or junk. This resulted in a set of 18 crossmodal icons. Therefore, there were 18 earcons representing the message alerts, and 18 tactons representing the same message alerts.

5.1.1 Type of Message

Three different rhythms were used to represent the three types of message: text, email, and voicemail. These rhythms have already been used successfully in tacton experiments [21] with average identification rates of 96.7%. Each rhythm was made up of a different number of beats, with the text rhythm consisting of one short beat and one long beat, the email rhythm consisting of two long beats and two short beats, and the voicemail rhythm consisting of one long beat, three short beats, and two long beats. Using a different number of beats in each rhythm helps to make the rhythms distinguishable [21]. These rhythms are presented in Figure 5-1 using standard musical notation.

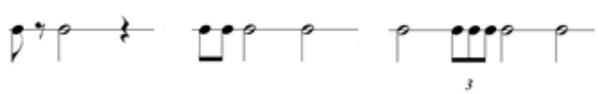

Figure 5-1: Text, email and voicemail rhythms (from [21]).

5.1.2 Urgency of Message

Two levels of roughness were used to represent urgent (very rough) and not urgent (smooth) messages. Different levels of tactile roughness were created as before in Chapter 4: an unmodulated 250Hz sine wave (smooth) and a 250Hz sine wave modulated by 30Hz sine wave (rough). The earcons used differing timbres as levels of roughness based on previous experiments on crossmodal parameters discussed in Chapter 4: a piano (General Midi patch No. 001) was used for smooth whilst a vibraphone (General Midi patch No. 12) was used for rough. It must be noted that, in this experiment, crossmodal tactile roughness was created using amplitude modulation. The results of the experiment detailed in Chapter 4 indicate that tactile waveforms are more effective than amplitude modulation but at the time that this experiment took place, this discovery had not yet occurred. Section 5.4 describes a re-run of the experiment discussed here using tactile waveforms instead.

5.1.3 Message Sender

Three locations on the user's waist were used to encode information about the sender in the tactile crossmodal icons – three vibrotactile actuators were placed on a Velcro belt on the left hand side, the front centre, and the right hand side of the waist (Figure 5-2). One of the experiments in Chapter 4 (Section 4.5) showed that these body locations can be effectively mapped to 3D audio locations in both a mobile and stationary environment, and that the waist was the most effective location for single parameter tactons. The audio crossmodal icons used three locations in a 3D audio soundscape to encode the information about the sender of the message – sounds were placed on a horizontal plane around the users head. A vibration or sound to the left hand side indicated that the message was from 'work', the centre indicated that the message was 'personal', and the right hand side represented 'junk' (Figure 5-2).

As an example, an urgent email from work in a tactile form would be the email rhythm with a rough texture to the left hand side of the user's waist, and the audio version would present the email rhythm played by a vibraphone to the left hand side of the 3D audio soundscape.

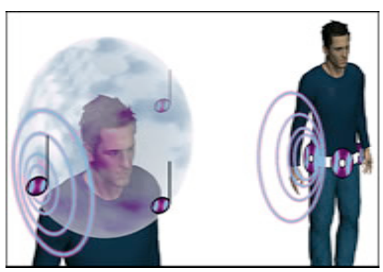

Figure 5-2: 'Junk' message indicated by audio panned to the right (Earcon) and tactile pulse on the right of the waist.

5.2 Experiment 3a – Lab-Based Study of Multi-Dimensional Crossmodal Icons

In response to Research Question 2, this experiment was conducted to investigate absolute identification of crossmodal icons encoding three dimensions of information to see if users would be able to use them and transfer knowledge of messages learned in one modality to the other. Half of the participants were trained in one modality and tested in the other: one quarter of the participants was trained to identify the crossmodal earcons and then tested with crossmodal tactons, the other quarter was trained with tactons and tested with earcons. As a control, the other half of the participants were trained and tested in the same modality (Table 5-1). Data were recorded on the identification of the three parameters of the message the user received– type, urgency, and sender. In addition, participants were informally interviewed about their experiences after the experiment.

Participant Group	**Training**	**Testing**
1	Audio	Tactile
2	Tactile	Audio
3	Audio	Audio
4	Tactile	Tactile

Table 5-1: Experiment conditions.

5.2.1 Aim and Hypotheses

The hypotheses were as follows:

1. If trained to identify the data encoded in audio crossmodal icons, participants will be able to identify the same data in the corresponding tactile crossmodal icons.
2. If trained to identify the data encoded in tactile crossmodal icons, the participant will be able to identify the same data in the corresponding audio crossmodal icons.
3. The rate of identification in the crossmodal training will be the same as that for participants trained and tested in the same modality.

5.2.2 Experiment Set Up

The C2 Tactor, as used in the experiments in Chapter 4, was used to present tactile stimuli. When being tested or trained in the tactile modality, three C2 EAI Tactors were attached to the participant's waist using a belt lined with Velcro (Figure 5-3). The participant also wore headphones to eliminate any inadvertent audio feedback from the actuators. Tactile sensitivity can vary across the waist therefore the vibrations could feel very different in intensity at different points on the waist [33]. To counteract this, each participant was asked to set the levels of the actuators so that they all felt equivalent in intensity at the start of the experiment.

When being tested or trained in the audio modality, the participants again wore headphones attached to a soundcard on a PC through which the audio alerts were played. The audio cues used in this experiment were created using the AM:3D[13] audio engine and were placed on a plane around the user's head at the height of the ears to avoid problems related to elevation perception. The sounds were located in front of the nose (0°) and ±90° to the left and right at each ear. Participants were asked to set the volume levels of the audio to a comfortable level at the start of the experiment.

Figure 5-3: Belt lined with velcro used in experiment with 3 C2 Tactors attached.

5.2.3 Methodology

Sixteen people took part in the experiment, aged between 22 – 38 years, 9 female and 7 male, all members of staff or students at the University. All participants had normal hearing and vision with no interfering medical conditions. The experimental method used a between–subjects design (see Table 5-1). At the beginning of the session participants were presented with a tutorial to introduce them to the concept of crossmodal icons, roughness, rhythm, etc., they were then allowed to experiment with either the crossmodal earcons or tactons (depending on the group to which he/she belonged). Then participants began training using a custom training/testing application (Figure 5-4).

The application is a purpose built experimentation system that can present audio and tactile cues of different types in multiple locations. The system presented the participant with either a tactile or audio cue at the beginning of each task. Then the participant could press the 'replay' button to have the cue presented again. Once participants had identified the information in the cue, they could select the corresponding button and submit the answer (using the tick button). After submitting the answer, a button appeared which the

[13] www.am3d.org

participants press when they are ready to move on to the next task. The system recorded the participant's responses, the time taken to respond, and also the number of times a cue was replayed. Participants were allowed to play each cue up to 4 times per task. Replaying the cues was allowed because the expected usage of these icons is in mobile devices where standard cues such as ringtones for incoming calls are commonly presented several times.

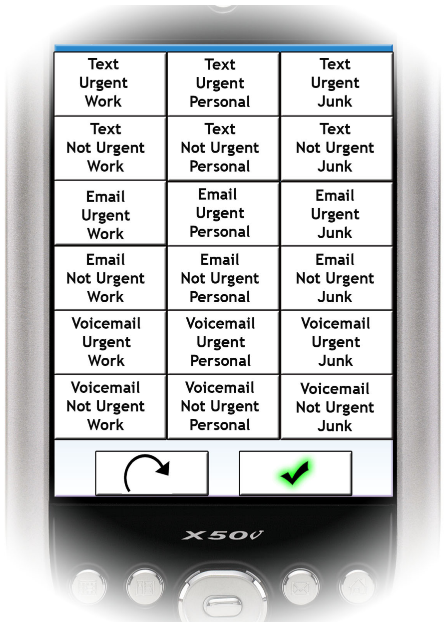

Figure 5-4: Screenshot of training and testing application.

5.2.4 Training

For training and testing, the standard Absolute Identification (AI) paradigm was used and participants received feedback on the correctness of their answers [139]. This paradigm uses a set of stimuli and responses (each of the same size) along with a one-to-one mapping between the sets. The stimuli are presented in a random order and the participant should submit an answer based on the response defined by the one-to-one mapping, i.e., to identify which of the stimuli was presented (message alert, email alert, voicemail alert etc.).

The set of stimuli used to train the participants was identical to the set on which they would be later tested. The application shown in Figure 5-4 was used to record participants' answers. The participants had to identify the information in the cue they heard or felt and then choose the appropriate button on the display shown in Figure 5-4. Each stimulus alternative was applied twice during each training run, resulting in a total of 36 tasks per run. During training the participants were required to repeat experimental runs (in audio or tactile) until a run with >= 90% correct identification was achieved so that the length of time taken to reach a good level of performance could be measured. If a participant did not reach 90% at the end of a training run, he/she received further training before being given another training run.

Originally, for the purposes of this experiment, training was used purely to ensure that all participants reached an appropriate level of understanding. However, the amount of time taken for participants to learn the sets of earcons and tactons is another interesting research issue as there is little data on how long it takes to learn such cues and if the learning required by each of the modalities is different. This allows a comparison of the results of crossmodal training/testing with the results of unimodal training/testing. There have been few other such studies into the training and learning of earcons and tactons. Brewster [13] found that participants could recall 81.5% of 27 earcons after 5 minutes of training through 'active learning'. The only related tactile work available at the time of this study was that of Enriquez and MacLean [41] where Haptic Icon recall rates were examined. It was found that participants could learn the meaning of 20 Haptic Icons varying in rhythm in under 20 minutes at average accuracy rates of 80% and that participants could recall 86% of the icons after 2 weeks.

5.2.5 Training Results

During the training and the experiment itself data were collected on the number of correct responses to the complete crossmodal icons. The learning curves for each participant and each stimulus set during training are shown in Figure 5-5 and Figure 5-6. The amount of time to reach the performance criterion varied across participants. These results show that, on average, it takes 2 training sessions for participants to be able to identify earcons with recognition rates of 90% or higher. They also show that, on average, it takes 3 training sessions for participants to identify tactons with recognition rates of 90% or above.

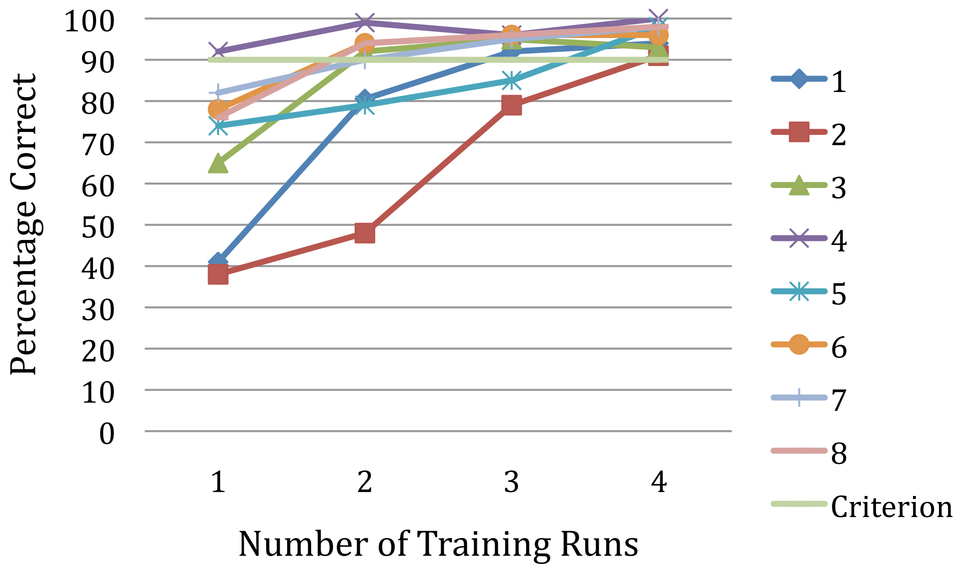

Figure 5-5: Learning curves for audio training.

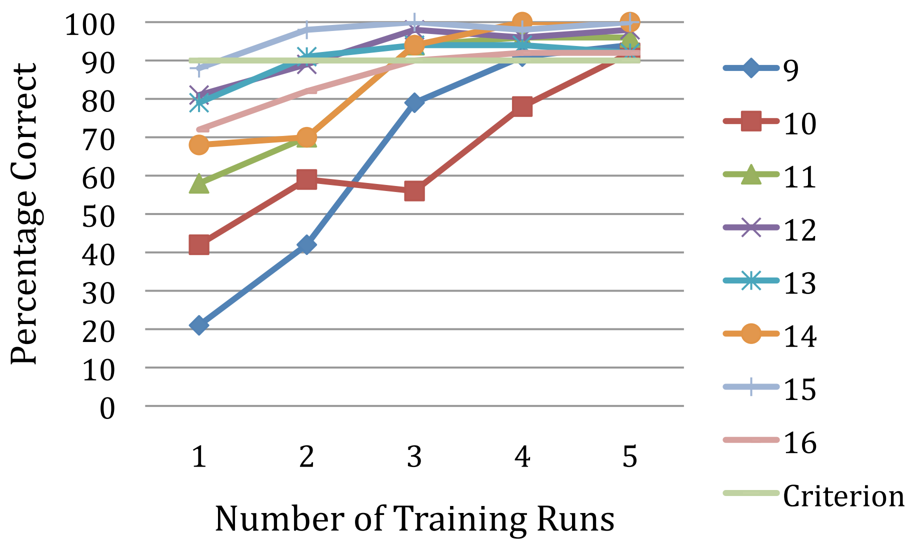

Figure 5-6: Learning curves for tactile training.

These results are promising for using audio and tactile interchangeably and would seem to indicate that the time taken to learn these crossmodal cues in either modality is comparable.

5.2.6 Testing in The Alternative Modality

Once the participants from Groups 1 and 2 (Table 5-1) had achieved the correct level of training, he/she completed the absolute identification test using the same online system and tasks but with cues presented in the other modality. Participants in the control groups (Groups 3 and 4) continued through the absolute identification test using the same tasks in the same modality after training.

In total there were 36 tasks in the experiment, with all 18 crossmodal icons (either audio or tactile) presented twice during the experiment. The order in which the crossmodal icons were presented was random for each participant. In each task the participant was presented with a crossmodal icon which he/she could replay up to 4 times. The participants had to identify the corresponding alert and then select the corresponding button in the dialogue box (Figure 5-4).

5.2.7 Results

The results from the control group in comparison to the crossmodal testing group are shown in Figure 5-7 (raw data can be found in Appendix E).

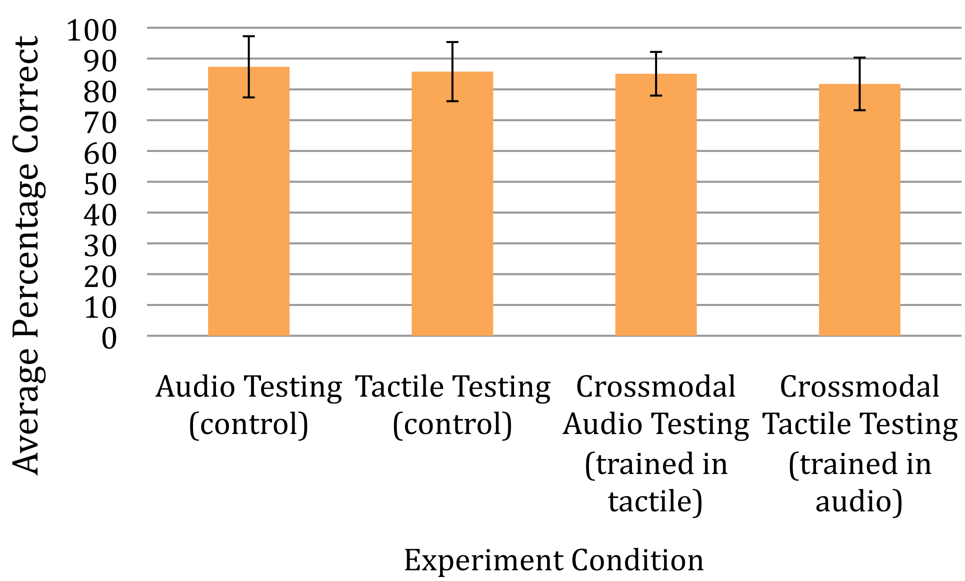

Figure 5-7: Average percentage of recognition rates during testing (with standard deviations).

The results for overall earcon recognition when trained with tactons showed an average recognition rate of 85.1%. The alert 'personal urgent text' achieved the highest recognition rate of 94% while the alert 'work not urgent voicemail' resulted in the lowest recognition rate of 61%. The results for overall tacton recognition when trained with earcons showed an average recognition rate of 81.7%. The alert 'personal not urgent text' achieved the highest recognition rate of 83% and the alert 'work not urgent voicemail' resulted in the lowest recognition rate of 56%.

Having examined the data in depth, there does not seem to be any clear reason for the low scores produced by the 'work not urgent voicemail' cue (the parameter design of this cue is: rhythm 3 (6 beats) with a very rough texture presented on the left-hand side of the waist). All of the individual parameters performed well in general and there was no apparent misunderstanding by the participants.

An independent measures 1-factor ANOVA (using an alpha level of 0.05) on the number of correct identifications showed that there was no significant difference in the recognition rates between the results of the four different Groups (training in audio / tested in tactile, training in tactile / tested in audio, training and testing in tactile, training and testing in audio) with ($F(3,60) = 2.1$, $p = 0.1$). The standard deviations in each condition vary only slightly and the mean scores are very close, thus the analysis suggests that information learnt in one modality can be recovered in the alternative modality in a way which is comparable with recognition of the same information in the trained modality. Thus Hypotheses 1 and 2 can be accepted. In terms of Hypothesis 3, the rate of identification in the crossmodal training was not exactly the same as that for participants trained and tested in the same modality but the results are only slightly lower, especially when trained in tactile.

The results suggest that if a user is taught to understand alerts provided by crossmodal tactons, he/she could be expected to understand crossmodal earcons with no audio training with approximately 85 % accuracy and if a user is taught to understand alerts provided by crossmodal earcons, he/she could be expected to understand crossmodal tactons with no tactile training with approximately 81.7% accuracy. These results are comparable to previous research in 3- dimensional earcons where McGookin's results [96] showed recognition rates of around 70% for identification of complete 3-dimensional messages in audio. They are also comparable with previous tactons research which produced recognition rates of 81% for identification of complete 3-dimensional messages in tactile icons [20].

5.3 Experiment 3b – Mobile Study of Multi-Dimensional Crossmodal Icons

As discussed at the start of the chapter, crossmodal icons are being developed for users of mobile touchscreen devices. Such users are often in motion when using devices so any alerts provided by the mobile device must be designed to be discernible in these situations too and not just when the user is stationary. There are many ways in which motion could affect perception of crossmodal output: mobile environments tend to change frequently with light, volume and vibration levels changing often. Consequently, another experiment in crossmodal identification was conducted which investigated the effects of motion on the results and assessed whether the good results observed in the laboratory would carry over to a more real world situation. The overall experiment involved 16 new participants who were either trained in audio or in tactile and then tested in audio or tactile whilst walking. The participants were aged between 24 and 26 (10 male and 6 female), all staff or students at the University of Glasgow with no physical impairments. Both the methodology and the crossmodal icons used in the experiment were the same as before to allow direct comparisons of the results.

The setup of this experiment was identical to the stationary version above in every respect except that participants were asked to walk on a treadmill during the experiment as opposed to sitting in a chair (Figure 5-8).

Figure 5-8: Mobile condition experimental set up.

This mobile experiment used a treadmill in a usability lab to simulate mobility. This was because the actuators used to present the tactile cues were controlled from a PC and therefore could not be tested in a real mobile environment. Studies show that using treadmills to simulate motion is good for simulating workload [77] when performance measures are of key interest and is a more controllable environment [5]. Furthermore, using a treadmill permitted us to set a standard walking speed for all participants (in this case, all participants walked at a constant speed of 5km/hr during the experiment).

The hypothesis in this experiment was:

4. Being mobile will increase errors produced during crossmodal icon identification and matching between modalities as compared to being stationary.

5.3.1 Results

The average number of errors for audio and tactile identification is shown in Figure 5-9 and in Appendix E. As before, the average recognition rate for both the audio and tactile groups was calculated but this time for the mobile condition as well.

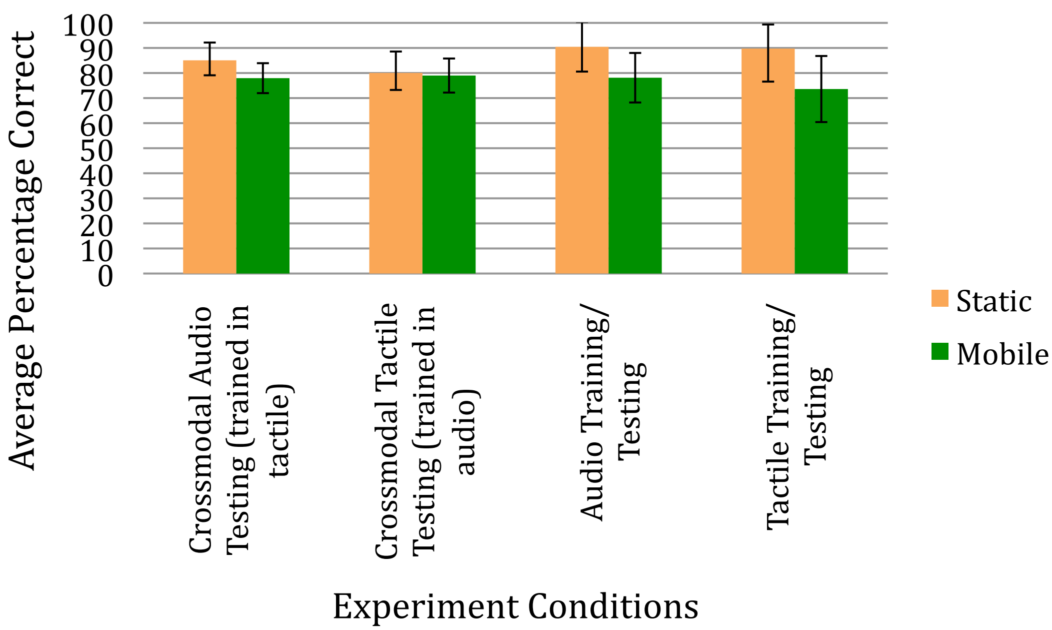

Figure 5-9: Average correct responses in stationary and mobile conditions (with standard deviations).

The results for overall earcon recognition when mobile and trained with tactons showed an average recognition rate of 78%. The results for overall tacton recognition when mobile and when trained with earcons showed an average recognition rate of 79%. To establish whether there is a significant difference between the mobile and stationary results, a 2-

factor mixed design ANOVA (alpha level 0.05) was applied using the two training conditions (audio or tactile) and stationary/mobile as the two factors. The 2-factor ANOVA showed that there was no significant difference in the recognition rates between the results from the mobile and stationary conditions with ($F(1,15) = 3.4$, $p > 0.01$). It also showed no significant differences in the recognition rates when trained with audio or with tactile ($F(1,30) = 0.7$, $p > 0.01$). There were no interactions between the two factors ($F(1,30) = 2.68$, $p > 0.01$). The standard deviations are very small indicating that the slightly higher number of errors was close to consistent between participants.

These results show that if a user is taught to understand alerts provided by crossmodal tactons, he/she could be expected to understand crossmodal earcons with no training when mobile with about 78% accuracy and if a user is taught to understand alerts provided by crossmodal earcons, he/she could be expected to understand crossmodal tactons with no training when mobile with approximately 79% accuracy.

5.3.2 Individual Parameter Results and Discussion

To establish the performance of each of the crossmodal parameters used, further analysis was performed on the data produced by both the audio and tactile versions of each parameter. The average percentage of correct responses in each audio parameter and each tactile parameter are shown in Figure 5-10 and Figure 5-11.

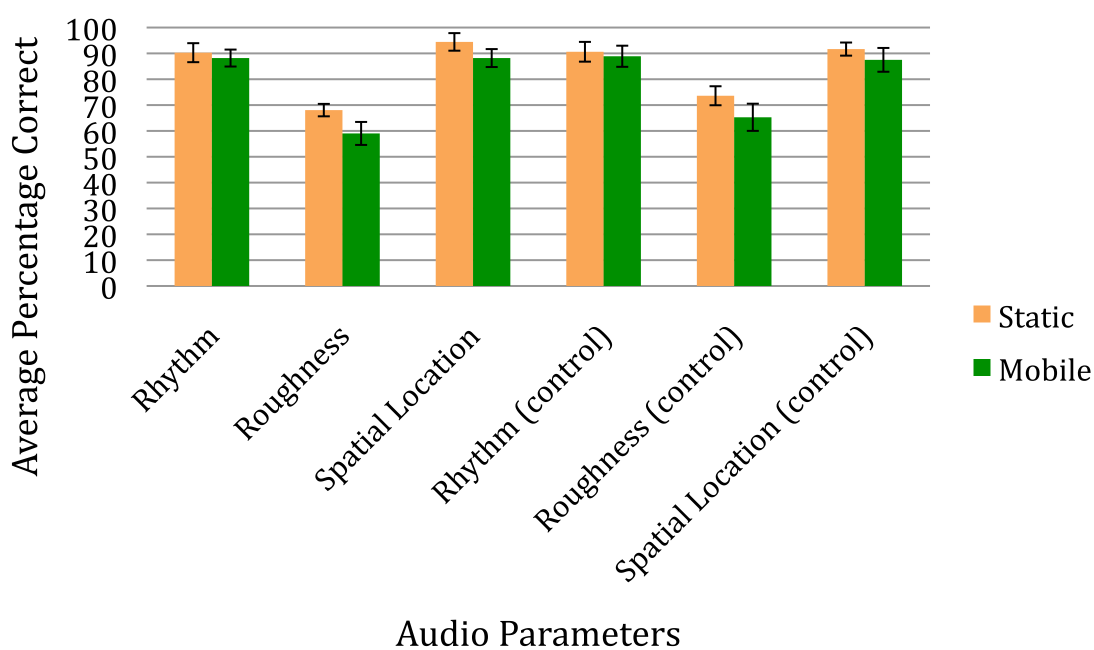

Figure 5-10: Average percentage of correct responses in each audio parameter (with standard deviations).

A 2-factor mixed design ANOVA on the audio recognition rates (alpha level 0.05) showed a main effect for parameter type, $F(5, 18) = 4.01$, $p = 0.09$, such that audio roughness (crossmodal or control) produced significantly poorer identification rates ($p = 0.05$) than rhythm and spatial location (both crossmodal and control versions). The main effect of mobility was also significant, $F(1,18) = 39.23$, $p < 0.0001$, indicating that the recognition rates were significantly higher in the static condition than in the mobile condition ($p = 0.05$). However, the interaction effect was not significant, $F(5,18) = 0.5$, $p = 0.6$).

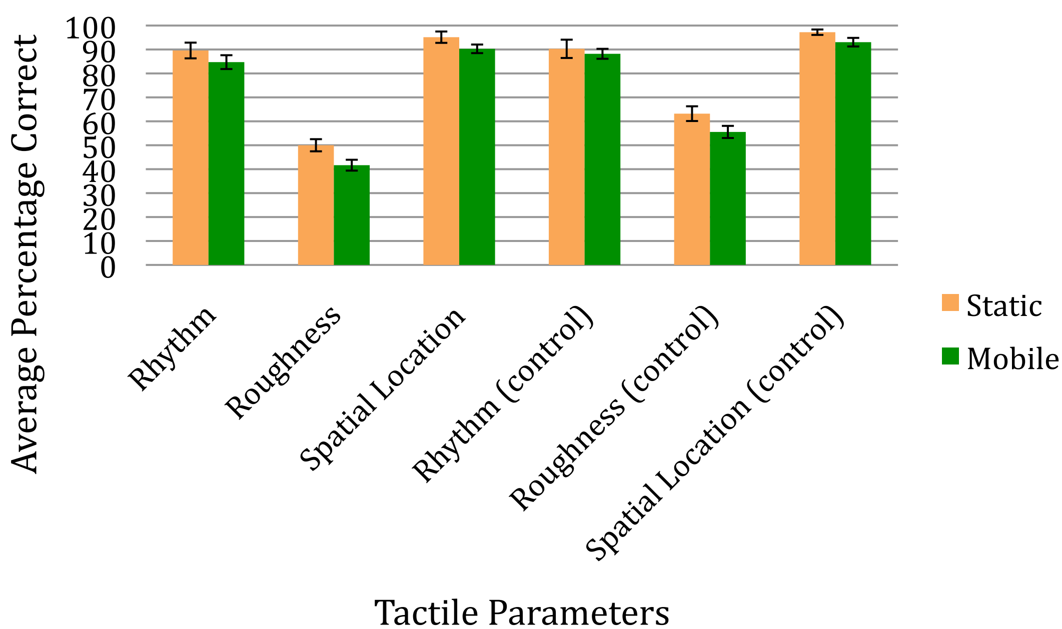

Figure 5-11: Average percentage of correct responses in each tactile condition (with standard deviations).

A 2-factor mixed design ANOVA on the tactile recognition rates (alpha level 0.05) yielded a main effect for parameter type, $F(5, 18) = 6.76$, $p = 0.04$, showing that tactile roughness (crossmodal or control) produced significantly poorer identification rates ($p = 0.05$) than rhythm and spatial location (both crossmodal and control versions). The main effect of mobility was also significant, $F(1,18) = 223.7$, $p < 0.0001$, indicating that the recognition rates were significantly higher in the static condition than in the mobile condition ($p = 0.05$). However, the interaction effect was not significant, $F(5,18) = 0.34$, $p = 0.7$).

There were no statistically significant differences found in the data for rhythm or spatial location. These results show that the rhythm and spatial location parameters produce comparable results. The identification rates of these parameters are also comparable regardless of whether training occurred in the same modality or not. These results suggest that rhythm and spatial location are effective crossmodal parameters and that there are no apparent disadvantages to training in the alternative modality.

In terms of the roughness parameter, these results suggest two different issues: firstly, overall the crossmodal roughness parameter is not as effective as rhythm and spatial location regardless of the training modality indicating that a different parameter may need to be used especially in mobile situations; secondly, when trained to identify roughness in one modality, participants struggle to then identify it in the same modality and in alternative modalities. This implies that roughness is not only a poor parameter for use in crossmodal interaction but is also ineffective in unimodal interaction, most prominently in the tactile modality.

5.4 Experiment 3c: Adding Tactile Waveforms

In view of the fact that amplitude modulation was replaced with tactile waveforms in Chapter 4 and an experiment showed that these waveforms could be effectively mapped to timbre in the audio domain, it was necessary to run the crossmodal identification experiment from Section 5.2 again to see if performance could be improved. As before, the experiment investigated absolute identification of crossmodal feedback encoding three dimensions of information to see if users can transfer knowledge of messages learned in one modality to the other.

The three parameters used to encode data are:

- Audio/tactile rhythm – type of message (text, email or voicemail)
- Audio/tactile spatial location – sender (work, junk, personal)
- Audio/tactile texture (using timbre in the audio and waveforms in the tactile modality) – urgency (not urgent, urgent, very urgent)

Once again, users were trained in the cues in one modality and then tested with cues in the other. Sixteen new participants took part in the experiment: 10 male and 6 female, all staff or students at the University of Glasgow, ranging in age from 19 to 42 with no physical impairments that could impede the study). All experimental conditions were the same as before (see Table 5-1). The only difference was the tactile version of roughness used to represent urgency in the messages. This time, two different tactile waveforms were used (sine and square wave) instead of amplitude modulation.

5.4.1 Training Results

As before, participants were trained to understand the crossmodal icons. The learning curves for each participant and each stimulus set during training are shown in Figure 5-12 and Figure 5-13. These results show that, on average, it takes 2 training sessions for participants to be able to identify earcons with recognition rates of 90% or higher. They also show that, on average, it takes 2 training sessions for participants to identify tactons with recognition rates of 90% or above. This is an improvement on the training results from the previous experiment where 3 training runs were necessary for participants to reach crossmodal tacton identification rates of 90%.

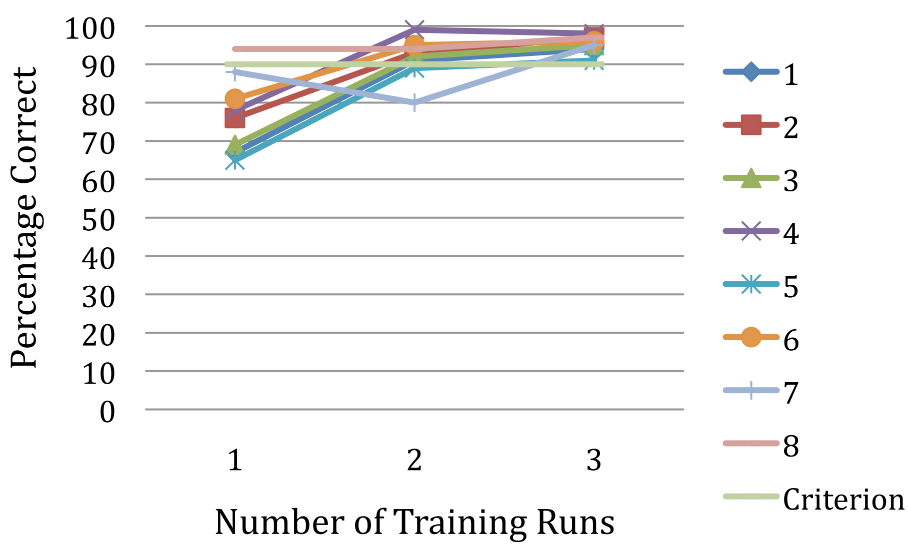

Figure 5-12: Learning curves for audio training.

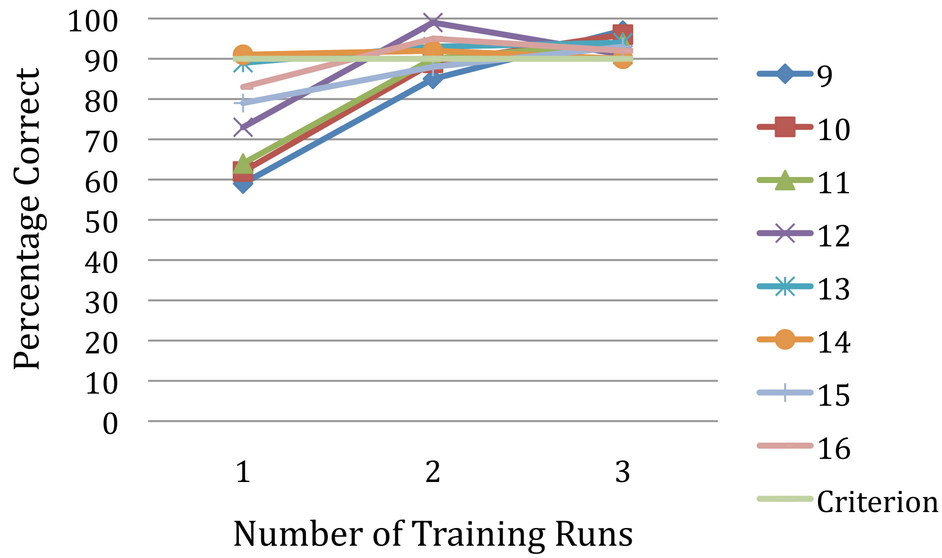

Figure 5-13: Learning curves for tactile training.

5.4.2 Testing Results

The results (Figure 5-14) show an overall recognition rate of 92% for crossmodal audio when trained in tactile and 89% for crossmodal tactile when trained in audio. Analysis using a between-subjects T-Test showed a significant difference (t = 2.06, df = 30, p = 0.05) between recognition rates for the first and second versions of tactile roughness. After users have been trained in the audio modality, tactile waveforms produce significantly higher recognition rates when used in the design of multi-dimensional crossmodal tactile icons compared to tactile roughness using amplitude modulation.

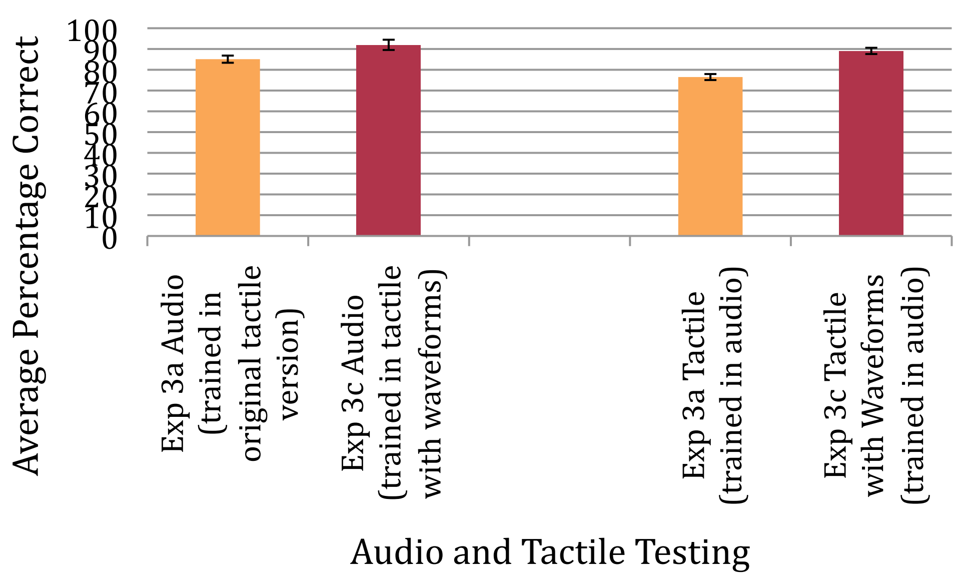

Figure 5-14: Average percentage correct in original experiment and follow-up experiment using tactile waveforms (with standard deviations).

To establish the performance of the new texture parameter design, further analysis was performed on the data produced by both the audio and tactile versions. The average percentage of correct responses in each roughness parameter is shown in Figure 5-15.

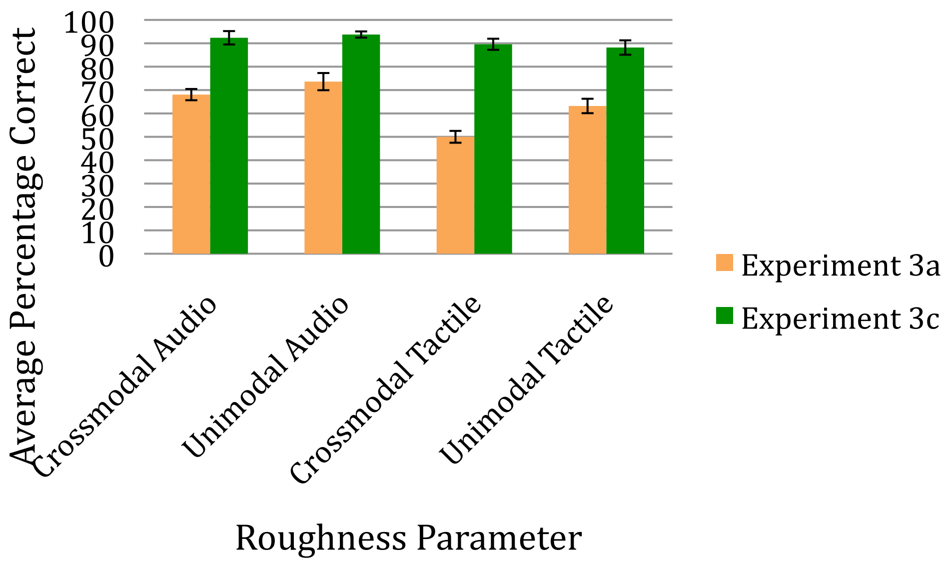

Figure 5-15: Average percentage of correct responses for the roughness (texture) parameter in each experiment (with standard deviations).

A 2-factor mixed design ANOVA (alpha level 0.05) showed a main effect for the experiment version, $F(1, 12) = 475.48, p < .0001$, indicating that audio and tactile roughness (crossmodal or control) produced significantly higher identification rates (p = 0.05) in Experiment 3c than Experiment 3a (both crossmodal and control versions). The main effect of parameter type (audio, tactile, crossmodal or unimodal) was also significant, $F(3,12) = 8, p = 0.0001$, indicating that unimodal and crossmodal audio roughness produced significantly higher recognition rates than crossmodal and unimodal tactile roughness ($p = 0.01$). There was no interaction effect, $F(3, 12) = 0.64, p = 0.3$. Even though the design of audio roughness was not changed (with different timbres representing different levels of roughness), the identification rates were higher in this experiment. The reason for this may be that, by using a more effective form of tactile roughness through waveforms in training, users find it easier to remember the mapping between modalities and then, in turn, find it easier to recognise the same texture in audio. Furthermore, the overall identification rates rose by an average of 9% compared to the multi-dimensional crossmodal icons using amplitude modulation. Therefore, using tactile waveforms lead to an improvement in the design of multi-dimensional crossmodal icons when using texture as a parameter. Interestingly, in this experiment the cue 'work not urgent voicemail' produced much higher recognition rates of 88.7% on average. This suggests that the low recognition rates obtained in the previous experiment were perhaps affected by the roughness parameter and its combination with the 6-beat rhythm and left-hand side spatial location. If this is the case, using the new and improved version of tactile roughness appears to have resolved this issue.

5.5 Discussion

There are many results from the three experiments discussed in this chapter. Three dimensions can effectively be encoded in crossmodal icons using rhythm, roughness and spatial location. Three rhythms were used varying from 2 to 6 notes. Three spatial locations were positioned in an audio soundscape and on a tactile belt to the left, right and centre of the user. In this case, crossmodal roughness was most successful in training and testing when designed using tactile waveforms and audio timbres. Although results in the stationary and mobile experiments show no significant difference in performance with crossmodal icons using rhythm and spatial location, audio and tactile roughness recognition rates are significantly lower when mobile. The mobile results are comparable to the results of the stationary conditions. Overall, the results indicate that, if a user is trained in one modality, the accuracy achieved when he/she is asked to identify the same information in the other modality is comparable even in a mobile situation.

Overall, the experiments discussed in this chapter have generated a large set of data (with 4 participants in each experiment taking part in each crossmodal condition making a total of 12). Each experiment has produced similar results in terms of training and absolute identification. The small standard deviations in the results data indicate consistency among participants across all of the experiments.

The overall findings from the experiment can be summarised as follows:

- Users in a stationary environment can accurately recognise 92% of messages presented by earcons, if they have been trained to recognise the same alerts presented by tactons.
- Users in a mobile environment can accurately recognise 78% of messages presented by earcons, if they have been trained to recognise the same alerts presented by tactons.
- Users in a stationary environment can accurately recognise 89% of messages presented by tactons, if they have been trained to recognise the same alerts presented by earcons.

- Users in a mobile environment can accurately recognise 79% of messages presented by tactons, if they have been trained to recognise the same alerts presented by earcons.

Although the mobile environment used in this experiment was much more controlled than a real world environment, these results give an indication of the sorts of effects that may be seen when a user is in motion. Future experiments will be conducted in real-world situations such as walking and traveling on a train or bus for example as discussed in Chapters 6 and 7.

5.6 Conclusions

This chapter presented an experiment which investigated the crossmodal transfer of information between the auditory and tactile modalities. The results from this experiment can be used to answer two of the research questions posed at the start of this thesis:

RQ1: What are the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons?

RQ2: What levels of performance can be achieved when these parameters are combined to create multi-dimensional crossmodal icons?

Previous research had investigated identification of information in tactons [20] and earcons [17] showing that both could effectively encode information in three dimensions. The research in this chapter investigated whether, if trained to understand multi-dimensional audio alerts, a user can then also understand the corresponding tactile alerts with no additional training and *vice versa*. The results suggest that this is possible or at least, the results failed to show any significantly worse results through crossmodal training and identification.

In relation to Research Question 1, this experiment indicated that rhythm, roughness and spatial location can be used to encode data in crossmodal icons. Identification rates of 52 - 94% have been achieved for these parameters in a lab-based setting, when trained in the same modality or in the alternative modality. In addition, when in a mobile setting identification rates of 43 – 91% have been reached. Furthermore, the sub-study described in Section 5.4 showed that, by using tactile waveforms in the roughness parameter, identification rates rose from 52% to 72%.

In relation to Research Question 2, this experiment has shown that an identification rate of 92% can be achieved for multi-dimensional audio crossmodal icons when trained in the tactile equivalents, and identification rates of 89% can be achieved for tactile crossmodal icons (using tactile waveforms to create roughness) when trained in the audio equivalent.

Users in a mobile environment can accurately recognise 78% of messages presented by earcons, if they have been trained to recognise the same alerts presented by tactons. Similarly, users in a mobile environment can accurately recognise 79% of messages presented by tactons, if they have been trained to recognise the same alerts presented by earcons.

The experiments described here are the first studies to be conducted investigating training and the transfer of training to other modalities in HCI. Like the work of Enriquez and MacLean [41], participants were trained to understand a set of icons. This experiment showed that participants could learn a set of 18 crossmodal earcons or tactons to a level of 90% within 20 to 30 minutes (2 to 3 training runs) whereas Enriquez and MacLean reached levels of 80% after 20 minutes with their Haptic Icons. The identification rates are comparable with those of Enriquez and MacLean with crossmodal tactons reaching levels of 79 to 89% recognition and crossmodal earcons reaching levels of 78 to 92% recall. This PhD research has not only examined the tactile modality alone but has looked at audio and the transfer of learning between modalities too.

The results of this research indicate that it may not be necessary to train users to understand icons in all the modalities a system might use. One concern with using lots of different modalities is the increase in complexity, however crossmodal interaction does not cause this, by eliminating the need for further user training with the addition of more modalities. If crossmodal icons are used to present information, training is only necessary in one modality as results show that users will then be able to understand the same messages in the other modality. Using crossmodal icons to communicate information to mobile device users could therefore reduce the learning time for the user and also increase the number of modalities through which this information may be transmitted.

Mobile technology incorporating audio and tactile output has now become widely available and this research has shown that feedback can be created which exploits users' abilities to transfer knowledge from one modality to another. By taking this into account and designing mobile applications with adaptive crossmodal icons, users may have the ability to interact with their devices even when their situation and surroundings are changing. The following two chapters examine this issue further.

Chapter 6 Applying Crossmodal Icons: Audio/Tactile Touchscreen Text Entry

Chapters 4 and 5 discussed the design of individual crossmodal parameters followed by the combination of these parameters in multi-dimensional icons. The outcomes from these stages of the research indicate that users can easily identify audio and tactile crossmodal icons with three different parameters. To test crossmodal icons in a real-world application and the related issues of usability with unpredictable and ever-changing mobile environments, a touchscreen QWERTY keyboard was developed complete with crossmodal audio and tactile feedback. The reason for this choice of application was based on the fact that a mobile device user's context can be extremely varied, for example, travelling on a train, at a party or in the gym. The user expects to be able to interact effectively with the device in all of these situations. For example, sending emails and browsing the Web on a touchscreen mobile device whilst on the train to work is a common activity. A solution is needed that makes the entry of text efficient and simple in these situations. Therefore the next step in this PhD research focuses on examining the incorporation of crossmodal icons in a mobile touchscreen application with an aim to find out if situationally impaired users can benefit from such crossmodal feedback. The design, implementation and evaluation aspects of this crossmodal mobile touchscreen application each explore the combination of many of the key features discussed in the preceding chapters.

Both Research Question 3 and Research Question 4 are addressed in this chapter:

> **RQ3**: Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?
>
> **RQ4**: Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

Section 6.2 covers the development of crossmodal feedback for a touchscreen keyboard application through a lab-based experiment, followed by a mobile version of the study based on an underground train. The application and related experiments discussed in this section provide a response to Research Question 3 by establishing the effects of crossmodal feedback on user performance levels with a touchscreen QWERTY keyboard during a text entry task.

Research Question 4 is addressed in the final part of this chapter (Section 6.6) which reports a further study. This experiment also took place on an underground train and was conducted to establish how changing levels of vibration and noise in the surrounding environment affect the perception and usefulness of crossmodal feedback. Situational impairments can affect a mobile user's ability to perceive information through the visual sense, but there has been very little research on how situational impairments affect the senses of hearing and touch. By investigating how perception alters as a user's surrounding alters, it is possible to establish exactly when to switch the application feedback to a more appropriate modality. This is the primary purpose of crossmodal icons.

6.1 Why a Keyboard?

As mentioned, touchscreen mobile devices are becoming evermore popular with both manufacturers and users. As there is no need for a physical keyboard to take up space on the device, they can have larger screens which can be used more flexibly, meaning a better display of videos, webpages or games, or reconfiguring the display as required, for example rotating from portrait to landscape. A soft keyboard can be displayed when text must be entered. The most popular such device at the present time is the Apple iPhone (Figure 6-1), but many other manufacturers have also removed the physical keyboards from devices such as PDAs, digital cameras and music players. The use of a touchscreen also allows novel forms of interaction, for example using gestures [109] on the screen to control a device, or more flexible forms of text entry and navigation.

Although the keyboards used on touchscreen devices are based on physical mobile keyboards with real buttons, one important feature is lost: the buttons cannot provide the audio or tactile response that physical buttons do when touched or clicked. Without the natural tactile or audio feedback, users can only rely on visual cues which can be ineffective in mobile applications due to small screen size, social restrictions and the demands of other real world tasks [89]. In an initial, small study [10] it was shown that entering text on a touchscreen when on the move can be problematic and that adding artificial tactile feedback can reduce error rates.

Figure 6-1: The Apple iPhone, a finger-operated touchscreen phone[14].

Another issue is that devices like the iPhone have discarded the stylus as an input device and use the finger instead (and in fact multiple fingers for certain interactions). The previous generation of touchscreen phones used a small stylus for interaction. This is advantageous from the device's point of view as the interaction point is very clear and easy for the device to recognise. Styli are, however, less convenient for users given their small size. Given the necessary limited size of touchscreen displays, widgets are often too small [105] or positioned in an awkward way preventing them from being selected easily with a finger. This makes it even more difficult for users to interact with their devices when, for example, travelling to work on a bumpy train.

In an effort to address these issues, several experiments were conducted investigating the use of crossmodal audio and tactile feedback for a touchscreen mobile phone QWERTY

[14] **www.apple.com/iphone**

keyboard where a fingertip is used to press the keys. One of the aims of these studies was to quantify the effects of crossmodal audio and tactile feedback in mobile and static settings by comparing a device with a physical keyboard to a touchscreen phone with a soft keyboard and then to the touchscreen phone with added crossmodal feedback.

6.2 Experiment 4: Crossmodal Feedback for Touchscreen Typing

The initial part of this research into crossmodal applications focused on each modality individually (audio and tactile) with respect to their effects on performance and usability with a touchscreen mobile QWERTY keyboard. The first experiment was conducted in both a lab and mobile setting investigating text entry on a touchscreen device with and without crossmodal tactile feedback. The aim was to explore the effects of this tactile feedback from keyboard events (confirming that the fingertip is touching a button, confirming that the button has been pressed and highlighting whether the fingertip has slipped off the button or not) to see if performance can be improved.

After initial investigations using currently available mobile touchscreen devices, the Palm Treo 750 and a Samsung i718 (Figure 6-2) were chosen for the experiment. The Palm Treo was chosen for the control condition as it has a physical keyboard, allowing the comparison typing performance between a touchscreen and real, physical buttons. The Samsung i718 was chosen as it has a large resistive touchscreen display, ideal for presenting a full QWERTY virtual keyboard. The i718 phone contains a Samsung Electro-Mechanics Linear Resonant Actuator and Immersion VibeTonz technology to control the actuator and produce tactile effects[15]. This actuator consists of a moving magnetic mass, an electromagnet and a spring. The resonant frequency is ~175Hz. The small size of the mass and the strength of the resonance makes this actuator ideal for short, sharp effects (such as are found in mechanical buttons) because it reaches maximum acceleration in 2 to 3 wavelengths (10 to 20ms).

[15] www.immersion.com

Figure 6-2: A Palm Treo 750[16] and a Samsung i718[17].

6.2.1 Feedback Design

A standard QWERTY touchscreen keyboard was created for the i718 that matched the one on the Treo in terms of button size and keyboard layout. Tactile effects could then be added as required using the built in actuator and Immersion's Vibetonz Studio (their tactile authoring tool). The exact size and spacing of the physical keys on the Treo were copied when designing the touchscreen buttons on the i718. The Treo keys were 50x35mm with a gap of 3mm between each. The i718 has a 2.8inch touchscreen with 240x320 pixels. The touchscreen buttons designed for the i718 were slightly larger than its standard soft keyboard, because they were based on the physical Treo keys and were designed for use with the fingertip not a stylus, but in all other respects were exactly the same as standard Windows Mobile buttons: they highlight when pressed.

In this study, a set of simple crossmodal tactons was created to represent the different keyboard events and keys that exist on a touchscreen keyboard. The events chosen were based on the most commonly occurring events during button use: fingertip-over, fingertip-click and fingertip-slip (defined in Section 6.2). The very nature of fingertip interaction means that a user's finger often covers any visual feedback from the application. These particular user interface events were identified by Brewster *et al.* [10, 16] and Lee *et al.* [83] as events that could benefit from non-visual feedback in fingertip interaction applications. All of the tactile feedback was created using the standard internal vibration actuator in the i718 device (equivalent sound files can be found in Appendix I). The crossmodal tactons were designed using the outcomes put forward in Chapters 4 and 5. Each tacton made use of the rhythm and texture parameters to represent different events:

- Rhythm: importance (1-beat for basic fingertip events and 3-beats for fingertip slip)

[16] **http://www.palm.com/uk/en/products/phones/index.html**

[17] **http://www.samsung.com/he/products/mobilephone/pdasmartphone/sgh_i718.asp**

- Texture: fingertip event (smooth for fingertip over, medium rough for fingertip down, and very rough for fingertip slip)

6.2.1.1 Fingertip-Over Event

In a standard interface, a mouse over event is fired when the mouse pointer is moved over a GUI element such as a button. This was adapted to create a fingertip-over event which is fired when the finger moves over any button in the interface. When the fingertip-over event is triggered, a 1-beat smooth 250Hz tacton is presented.

On traditional physical keyboards it is common to find raised ridges on the 'F' and 'J' keys used for orientation. To recreate this on the touchscreen keyboard, when the 'F or J' key triggers the fingertip-over event a different textured tacton is presented. The tacton is 1-beat amplitude modulated 250Hz sine wave, which feels rough (the experiments in Chapter 4 established that square waves can be used to create a rough sensation but this was not possible with the built-in actuator so amplitude modulation was used instead).

6.2.1.2 Fingertip-Click Event

Tactile feedback was used to confirm that a button had been pressed. When the fingertip-click event was triggered, a 1-beat medium rough tacton was presented. The cue used a sawtooth wave and no ramp up or ramp down time to create a very short and quick 'click' resembling the 'click' felt when depressing a physical button.

6.2.1.3 Fingertip-Slip Event

An event was triggered whenever the fingertip moved over the edge of any button on the screen, indicating a transition or slip from one to the next (fingertip slips can be troublesome for users and can cause errors that are often undetected). This allowed users to run their fingertips over the buttons feeling all of the edges. When the fingertip-slip event is triggered, a 3-beat rough tacton is presented. The rough texture is created using an amplitude modulated 175Hz sine wave. This tacton was designed to be attention grabbing and to feel very different to the other cues allowing easy identification of a slip by the user.

6.2.2 Methodology

The experiment was designed to investigate the effects of incorporating crossmodal tactile feedback into mobile touchscreen buttons. The experiment compared user performance on a typical mobile device featuring a real, physical keyboard (Palm Treo), to a touchscreen soft keyboard with added tactile feedback (Samsung i718) and to the same device with no tactile feedback. The experiment hypotheses were as follows:

1. Participants will be able to enter text with the least errors and greatest speed on the physical keyboard;
2. Tactile feedback will improve speed and accuracy of text entry on touchscreen keyboards;
3. Touchscreens with tactile feedback will achieve comparable accuracy and speed levels to physical keyboards;
4. Tactile feedback will improve the speed and accuracy of text entry on touchscreens when mobile.

It was necessary to recruit participants who had some expertise in text entry on mobile devices, so before beginning the experiment participants were required to complete a questionnaire on their text entry habits. Participants were chosen who send, on average, 1 - 10 text messages on a QWERTY mobile device per day as they can be considered moderate users who have experience of using physical keyboards. It was not possible to get enough participants who had experience of touchscreen keyboards as these devices were not yet common (at the time this experiment was run in 2008). Users were given training with the keyboards as discussed later.

A within-subjects design was used where the conditions were:

1. Standard mobile device with physical keyboard (the control Physical condition);
2. Touchscreen mobile device with tactile feedback added to soft keyboard (the Tactile condition);
3. Touchscreen mobile device with soft keyboard (the Standard condition).

The phrase set used for the text in the experiment was from MacKenzie [90] and has been used successfully in several studies [91] [132]. It is a 500-phrase set with no punctuation symbols and no upper case letters. Due to time constraints from the experimental design, the full set of 500 phrases could not be used so a random set of 30 phrases was selected for each run of the experiment. This resulted in each condition (Standard, Tactile and Physical) lasting approximately 20 minutes. All conditions were tested in a static lab environment and also on the move on a subway train.

6.3 Experiment 4a(i): Crossmodal Tactile Text Entry

Twelve participants took part in this experiment. All participants were students or staff at the University with an age range of 18 to 38 years. There were 3 female and 9 male participants. Two participants were left-handed. All participants were seated during the experiment and asked to hold the device in their hands at all times.

Participants were shown a phrase and asked to memorise it, then type it in using the keyboard for each condition. Participants were asked to enter it as quickly and as accurately as possible. Each phrase was shown at the top of the screen until the participant began typing at which point the phrase disappeared. The interface used on the i718 is shown in Figure 6-3. The Treo had the same display, except the onscreen keyboard was not shown; participants hit the physical 'Enter' key to submit a phrase. Both devices run Windows Mobile so could run exactly the same code. This method sits in between the text creation method and the text copy method. Text creation (where users come up with their own messages), although most realistic, is difficult to use as errors cannot easily be detected. Text copy (users copy messages on the screen) is not very realistic, as most users do not copy their text messages or emails, for example, from a piece of paper onto their device. The method used in this experiment was not text creation, but the participants were not copying text directly onto the device either making it a slightly more realistic scenario (Brewster *et al.* [10] used the copying method in their earlier study). Timing began when the participants hit the first key and stopped when 'Submit' was pressed (or Enter on the Treo). Participants moved on to the next task whether or not the phrase was correct.

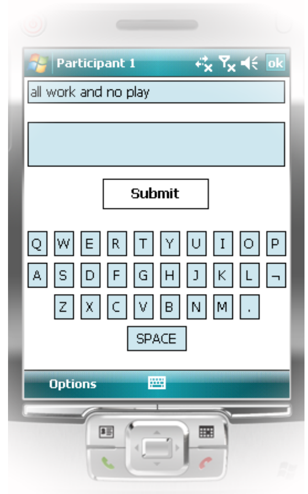

Figure 6-3: Screenshot of the experiment interface.

The mobile part of the study could have been tested in many different ways, for example, with users walking [109]. This has been shown to be an effective way to generate some of the workload of using a device whilst on the move [10]. The interaction was investigated on a subway train (Figure 6-4) on the Glasgow Underground. People use PDAs and phones on trains and buses every day whilst commuting. The underground is a good platform for testing as noise levels are very dynamic, being quiet when stopped at a station, but very noisy when the train is in motion. Light levels again vary dramatically. Vibration and movement are also very changeable. When the train is stopped there is little vibration. However, when it accelerates and decelerates people are subjected to lots of forces and vibration from the engine and general movement. Another important factor for this experiment is that the within-subjects design used meant that participants had to use three different keyboards which took around one hour. This would be too far for some of the participants to walk. The subway allowed testing in a realistic usage situation without fatiguing the users too much.

Conditions in this experiment were fully counterbalanced. Half of the participants completed the lab-based experiment first while the other half took part in the mobile subway train session first. For both the lab and mobile parts of the experiment, the keyboard conditions were also counterbalanced. The first set of conditions was completed on one day and the second set was completed at least one day later, to avoid participant fatigue. A training period was given before each trial (with ten phrases for each keyboard type) to familiarise each user with the interface to be used. Tactile feedback was described and users were given the chance to physically feel the feedback with their fingertips. The dependent variables measured in the experiment were speed, accuracy, keystrokes per character and subjective workload (using the NASA TLX workload assessment [57]). An extra factor was added, annoyance, to the workload analysis to specifically focus on any issues of irritation that the tactile feedback might cause the participants.

Figure 6-4: The mobile condition of the experiment on the subway train. The experimenter (on the left) takes notes whilst the participant (on the right) enters text.

6.3.1 Results

6.3.1.1 Accuracy

The Physical keyboard condition had accuracy levels of 88.25% in the lab and 89.6% in the mobile setting. The Tactile condition achieved scores of 82.7% in the lab and 80% on the train, while the Standard touchscreen keyboard produced scores of 69.6% in the lab and 65.8% when mobile (see Figure 6-5). The raw data can be found in Appendix F.

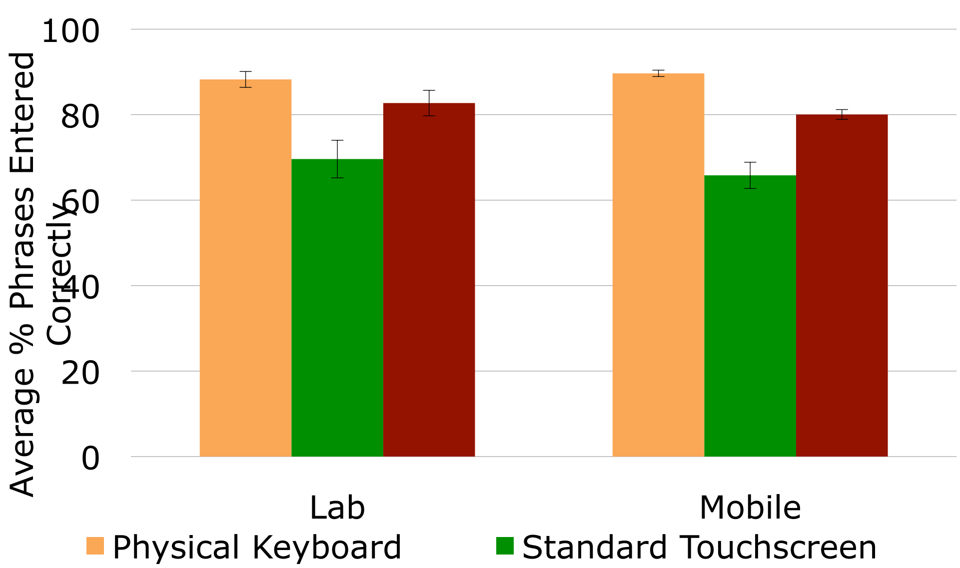

Figure 6-5: Average percentage of phrases entered correctly (with standard deviations).

A 2-factor repeated measures ANOVA (alpha level 0.05) was performed on the mean number of correct phrases entered, comparing the effects of mobility (static and mobile) and the three keyboard types (Physical, Standard and Tactile). A correct response in this case was when the entered phrase (when the 'submit'/Enter button was selected) matched the given phrase completely (regardless of whether corrections were made along the way).

The ANOVA showed a significant main effect for keyboard type ($F(2,22) = 96.9$, $p < 0.001$). Using *post hoc* Tukey's Pairwise Comparisons, it can be seen that a significantly higher number of phrases was entered correctly on both the physical keyboard and the tactile touchscreen than on the standard touchscreen ($p = 0.05$). There were no significant differences in the number of correct phrases entered on the physical keyboard and on the tactile touchscreen. The scores were, on average, 5.5% lower on the tactile touchscreen than the physical keyboard in the lab and 9.6% lower when mobile. Between 1.6 and 2.8 more phrases were entered incorrectly in the tactile condition, suggesting that the performance is comparable with the real keyboard.

There was no main effect for mobility ($F(1,11)=1.79$, $p = 0.18$) and no interaction between keyboard type and mobility ($F(2,22)=1.58$, $p = 0.21$. This suggests that users did not type less accurately when on the move.

These results show that the addition of crossmodal tactile feedback can overcome some of the problems caused by the Standard touchscreen keyboard and enable people to notice and recover from errors he/she makes (which confirms the result of Brewster *et al.* [10]).

6.3.1.2 Keystrokes Per Character (KSPC)

The number of keystrokes per character was recorded for each keyboard type. KSPC is the number of keystrokes required, on average, to generate a character of text for a given text entry technique in a given language with the ideal being one per character [132]. Given that accuracy scores were based on whether or not the submitted phrase matched the given phrase exactly and did not include corrections as errors, KSPC was recorded in order to examine how many corrections users had to make before submitting a correct phrase. The average number of KSPC for each condition is shown in Figure 6-6.

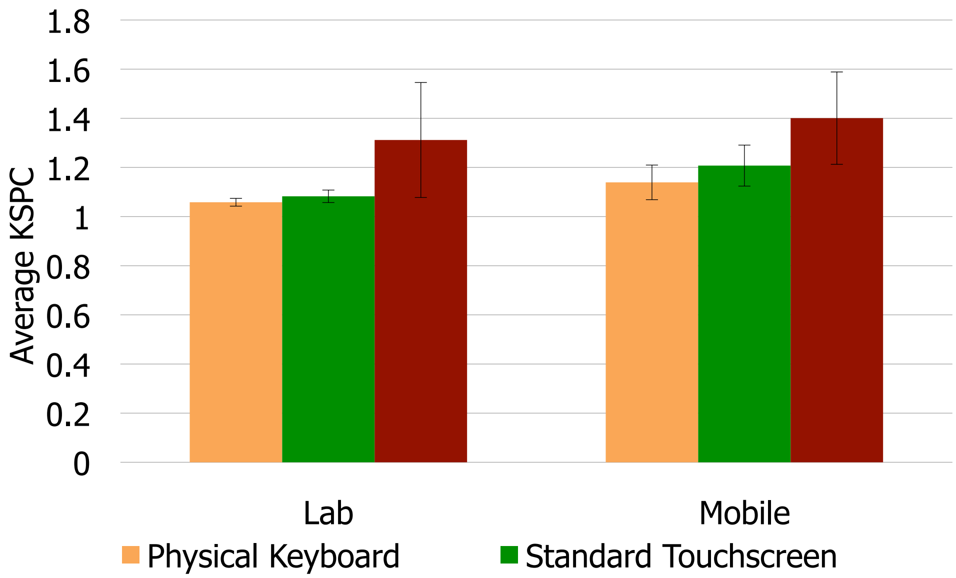

Figure 6-6: Average KSPC for each setting and keyboard type (with standard deviations).

A 2-factor repeated measures ANOVA (alpha level 0.05) was performed on the KSPC data comparing the effects of mobility and the keyboard type. A significant main effect for KSPC for keyboard type was found ($F(2,22) = 6.58$, $p<0.0001$). Tukey tests showed that there were significantly more KSPC when typing on the tactile touchscreen than the physical or standard keyboards ($p = 0.05$). There was no main effect for mobility ($F(1, 11) = 2.53$, $p = 0.11$) and no interaction between keyboard type and mobility ($F(2,22)=0.04$, $p = 0.95$).

The standard touchscreen keyboard had a lower KSPC than the tactile one. The reason for this is that participants corrected fewer of the errors (as can be seen in Figure 6-5). In an experiment like this one this is a reasonable tradeoff for participants as there was no penalty for errors (participants could continue to the next phrase even if the current one was incorrect). In a real life setting, this would result in many mis-typed email addresses or URLs. The physical keyboard was still the best, with the lowest KSPC value. This suggests that the crossmodal tactile feedback added helps some aspects of typing but it is not quite at the level of a real, physical keyboard.

6.3.1.3 Time to Enter Phrases

Figure 6-7 shows the average time taken to enter a phrase for each keyboard condition in the lab and mobile settings. Participants using the physical keyboard entered the phrases with means of between 13 and 17 seconds (lab and mobile). The tactile touchscreen allowed participants to enter a phrase of text in 20 seconds (lab) and 22 seconds (mobile) while text entry on the standard touchscreen took longer with rates of between 25 and 27 seconds.

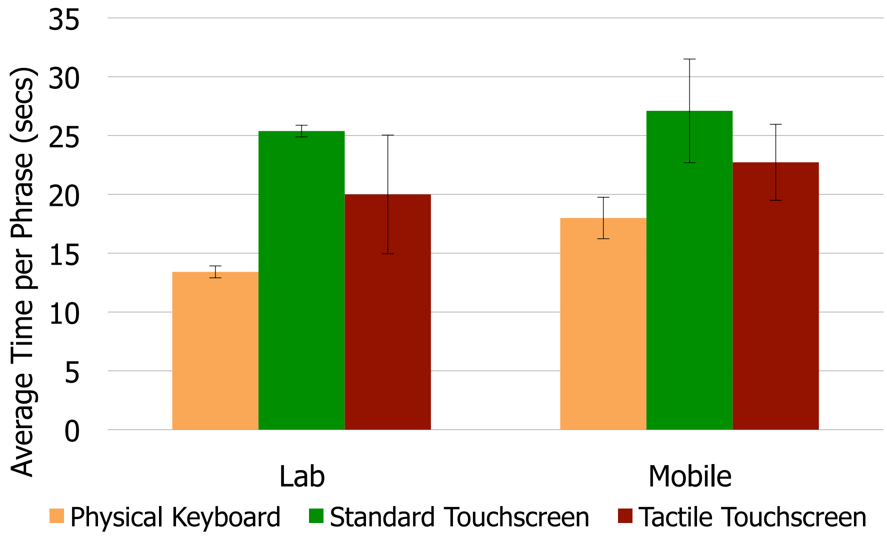

Figure 6-7: Average time per phrase entered in seconds (with standard deviations).

The average words per minute for each keyboard type in static and mobile settings are shown in Figure 6-8. The analysis shows that participants can reach levels of over 19 words per minute when mobile using a physical keyboard. The results for the tactile keyboard are comparable at 15.1 words per minute. The standard touchscreen allows participants to type more than three words less on average - 12.6 words when mobile.

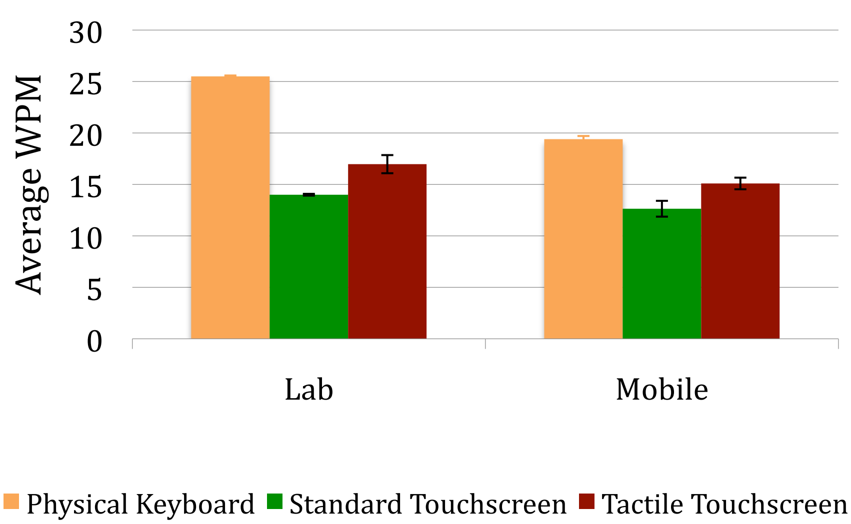

Figure 6-8: Average words per minute for each type of keyboard in the lab and mobile settings (with standard deviations).

A 2-factor repeated measures ANOVA (alpha level 0.05) on the time taken to enter phrases for the keyboard types and mobility showed a significant main effect for keyboard type ($F(2,22) = 69.78$, $p < 0.001$). Tukey HSD tests showed that the time taken to enter phrases on the physical keyboard and tactile touchscreen were significantly lower than on the standard touchscreen keyboard ($p = 0.001$). The physical keyboard was significantly faster than the tactile one ($p = 0.05$), indicating that the physical keyboard allows faster typing speeds than the touchscreen even with tactile feedback. However, overall, this result still indicates that tactile additions to the standard soft keyboard had a significant positive effect on the usability of the device. Combining this with the accuracy results suggests that crossmodal tactile feedback can offer some significant advantages for touchscreen devices.

This time there was a significant main effect for mobility ($F(1,11) = 9.48$, $p = 0.003$), with the mobile condition increasing the time taken to enter phrases over the lab (there was no interaction, $F(2,22)=2.65$, $p = 0.077$). This shows that being mobile does slow down text entry rates due to the movements in the environment even though it did not affect accuracy. This may be because participants chose to maintain accuracy at the expense of input speed.

6.3.1.4 Subjective Workload

The results of the NASA TLX [57] questionnaires are shown in Figure 6-9 (a copy of the questionnaire can be found in Appendix F). A 2-factor ANOVA (alpha level 0.05) on overall workload showed a significant main effect (for the standard six factors) for keyboard type ($F(2,22 = 111.35$, $p<0.001$). There was no significant main effect for mobility ($F(1,11) = 0.19$, $p = 0.66$) and there was no interaction ($F(2,22) = 0.7$, $p = 0.49$). Tukey HSD tests showed that overall workload when using the standard touchscreen keyboard was significantly higher than when using the physical keyboard or the tactile touchscreen keyboard ($p = 0.05$). There was no significant difference between the Physical and Tactile conditions.

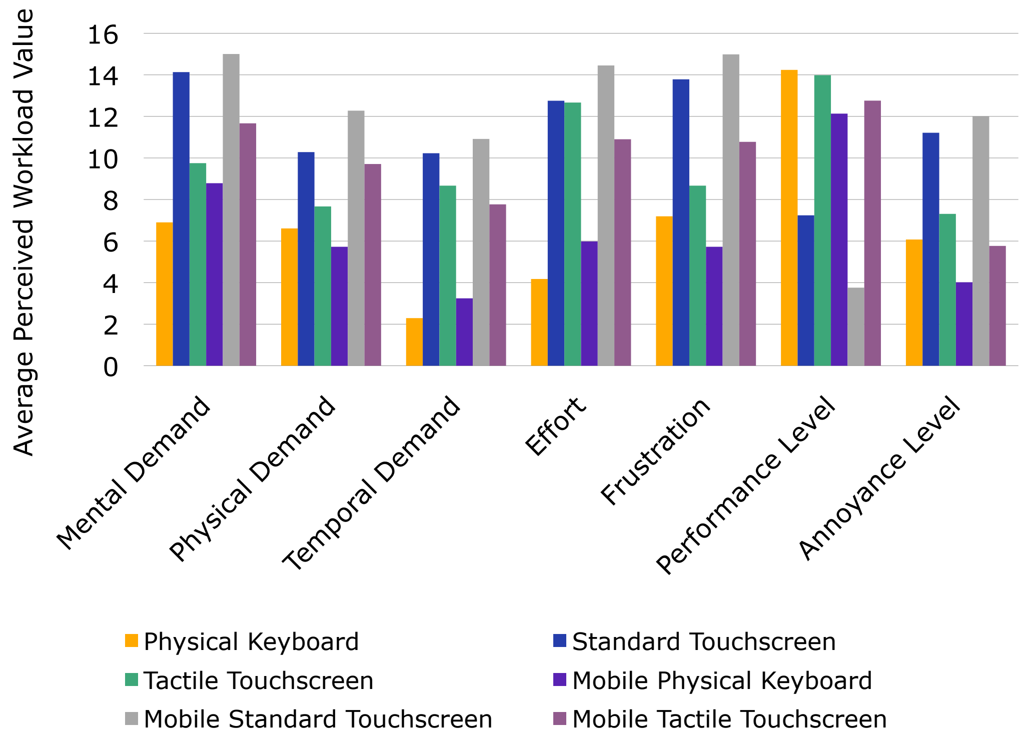

Figure 6-9: Average scores from NASA TLX questionnaires.

Further analysis using a single factor ANOVA on each of the workload factors showed a significant difference in all seven factors of the workload analysis with $p<0.001$. A Tukey HSD test confirmed that the mental demand, physical demand, frustration and annoyance levels when using the standard touchscreen are significantly higher than when using the other keyboard types. It also showed that perceived performance levels are significantly lower on the standard touchscreen.

The analysis of each workload factor also showed that temporal demand and effort was significantly higher when using the standard or tactile touchscreen than when using the physical keyboard. Some participants commented that the standard touchscreen was frustrating as there was no feedback when the fingertip had moved off the edge of the button. Any visual feedback was masked by the fingertip over the button.

6.3.2 Discussion

Hypothesis 1 can be accepted as it has been shown that using the physical keyboard produces significantly fewer errors and the greatest input speed with phrases being entered up to 10 seconds faster than the on the standard touchscreen. The greatest input speed results apply in both the lab and mobile settings.

Hypotheses 2 and 4 can also be accepted as the results show that touchscreen keyboards with tactile feedback produce fewer errors and greater speeds of text entry compared to standard touchscreen keyboards without tactile feedback both in the lab setting and in the mobile setting. This can be seen clearly in the mobile setting where phrases were entered up to 6 seconds faster with tactile feedback and accuracy scores were as high as 74% compared to the poor accuracy scores on the standard touchscreen keyboard. These results indicate that when in a mobile situation on a bumpy noisy train, it becomes even more difficult to use a standard touchscreen keyboard but tactile touchscreens still perform significantly better despite the dynamic environment.

Given that text entry on the tactile touchscreen only took 4 seconds longer on average than the physical keyboard and the accuracy results between both keyboards are comparable, hypothesis 3 can be partially accepted. In the lab setting, participants reached speeds of 14 WPM on the standard touchscreen and 17 WPM on the tactile touchscreen. These results are slightly lower than those of MacKenzie *et al.* [91] where novice users reached speeds of 7 to 10 WPM and experts reached 21 WPM on standard touchscreen soft keyboards. However, all participants in this case were novices and MacKenzie *et al.* did not test in a mobile environment.

6.4 Experiment 4a(ii) – Crossmodal Tactile Text Entry Version 2

Given the promising results obtained in this experiment suggesting that crossmodal tactile feedback could significantly improve the usability of touchscreen text entry, it was decided that this should be further investigated using an alternative actuator (the C2). Using this actuator allowed the inclusion of spatial location as an additional crossmodal parameter to see if performance could be improved even further. In Chapter 5, tactile spatial location was presented to the user's waist but different tactile spatial locations do not necessarily have to be on the body but could be on the actual device. For instance, localised tactile feedback on touchscreen mobile devices can provide spatial information [59]. Therefore, the same experiment was run again in the lab and mobile environments, but this time using a Dell Axim PDA (a small handheld Personal Digital Assistant with a resistive touchscreen). The Dell PDA was used because the C2 actuators could not be connected to the proprietary audio connection in the i718. The aim of this experiment was to investigate whether including the spatial location parameter could increase performance and get closer to that of the real physical keyboard.

6.4.1 Hardware

In the experiment, the PDA was augmented with two C2 actuators attached to the back (Figure 6-10) so that they rested under the user's hand when it was held. Both pressed against either side of the palm of the hand.

Figure 6-10: The Dell Axim PDA with 2 C2 actuators on the back.

6.4.2 Feedback Design

The tactile feedback used in this experiment was identical to that provided in Experiment 4a(i) in that the manipulated tactile parameters are the same. However, the feedback felt very different due to the higher quality of the actuators: they ramp up faster and modulation is clearer. The only other difference was that the actuators provide localised feedback to the hand holding the device as opposed to shaking the whole device. By placing the two actuators on the left and right sides of the device, spatial location could be incorporated into the feedback to give some indication of which button was giving the feedback.

Whenever a fingertip-over, fingertip-click, or fingertip-slip event was triggered, the actuator placed nearest the button would be used to present the feedback. For instance, if the button 'A' was pressed, the actuator on the left was activated, if the button 'G' was pressed, both actuators were activated and if the button ‘L’ was pressed, the actuator on the right was activated. Again, the devices both run the same Windows Mobile code as before.

6.4.3 Methodology

The aims and methodology of this experiment were the same as the previous experiment with one additional hypothesis:

5. Vibrotactile feedback from the C2 actuators will provide better results than the built-in actuator in the mobile device.

A new set of 12 participants was recruited from the University, aged between 18 and 26 (8 male and 4 female) with no physical impairments. Each participant was presented with 30 random phrases from the MacKenzie phrase set and asked to enter them as quickly and as accurately as possible on the PDA (with and without tactile feedback). Participants used the device in the lab and then on the subway. Participants only used this device for the tactile touchscreen and standard touchscreen conditions as these results could be compared back to Experiment 4a(i) for performance on the physical keyboard condition. Once again, speed, accuracy and KSPC were measured.

6.4.4 Results

6.4.4.1 Accuracy

The average number of phrases entered correctly is shown in Figure 6-11 alongside the results from Experiment 4a(i) for comparison. In the lab condition the PDA with actuators scored 23.8 correct answers and mobile 24.5.

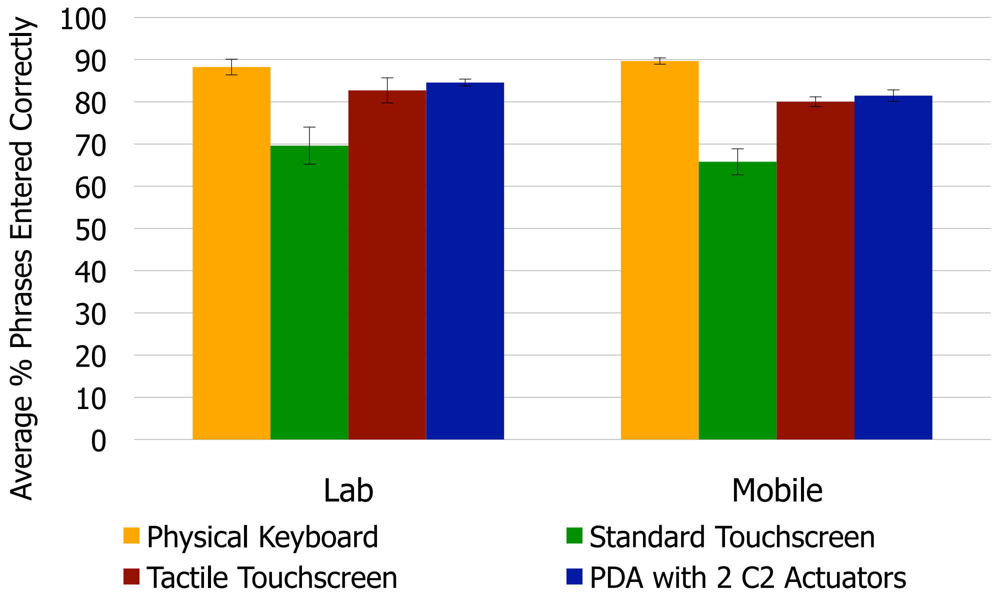

Figure 6-11: Average percentage of phrases entered correctly (with standard deviations).

A 2-factor ANOVA with replication was performed on the mean number of correct phrases entered, comparing the effects of mobility (static and mobile) and the four keyboard types (Physical, Tactile and Standard from Experiment 4a(i) and the PDA with 2 C2 actuators). Although participants in Experiment 4a(ii) had less practice (participants only did PDA with no tactile and PDA with tactile in lab and mobile), the first study was fully counterbalanced so that a valid comparison can be made between both sets of data from each experiment. The ANOVA showed there was a significant main effect for keyboard type ($F(3,33) = 84.6$, $p<0.0001$). *Post hoc* Tukey tests showed that significantly more phrases were entered correctly on the physical keyboard, PDA with 2 actuators and on the tactile touchscreen compared to the standard touchscreen ($p = 0.05$). There were no significant differences between the tactile touchscreen, PDA and the physical keyboard. The results show that the average number of correct phrases on the physical keyboard (26.4 in the lab, 26.9 when mobile) and the PDA (25.3 in the lab, 24.4 when mobile) were very similar. There was again no main effect for mobility ($F(1,11)=2.1$, $p = 0.144$) and no interaction ($F(3,33)=1.39$, $p = 0.24$).

6.4.4.2 Keystrokes per Character (KSPC)

The KSPC data for Experiment 4a(i) and the PDA with the C2 actuators is shown in Figure 6-12. The PDA scored a mean of 1.20 in the lab and 1.24 when mobile.

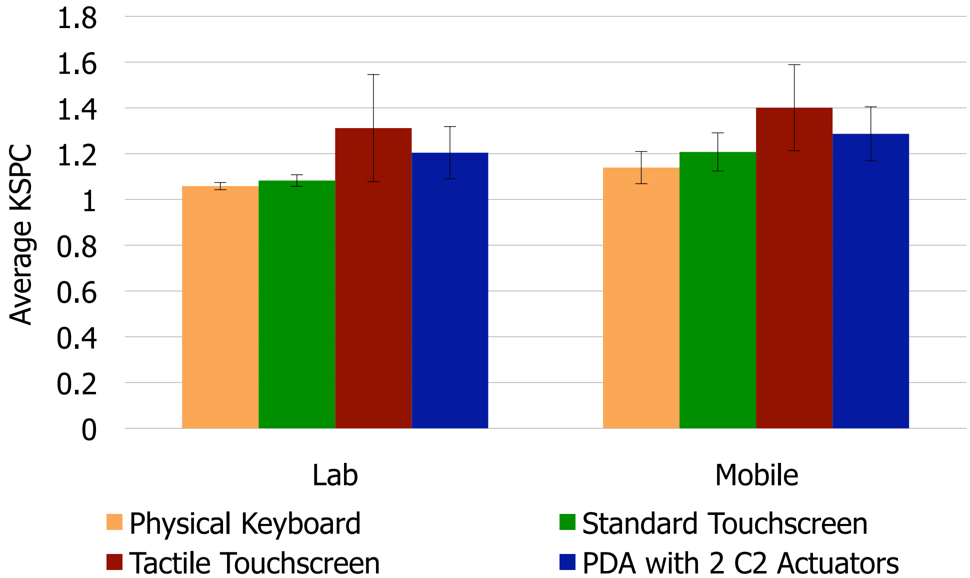

Figure 6-12: Average keystrokes per character (with standard deviations).

A 2-factor repeated measures ANOVA showed a significant main effect in KSPC for keyboard type ($F(3,33)=4.82$, $p = 0.003$). Using *post hoc* Tukey's Pairwise Comparison, it can be seen that a significantly higher number of KSPC occurred when using the tactile touchscreen compared to all three other types of keyboard including the PDA with C2 actuators ($p = 0.05$). There were no significant differences in the KSPC on the physical keyboard and on the PDA with C2 actuators or the standard touchscreen. This suggests that the PDA has an advantage over the tactile touchscreen in that KSPC is reduced, meaning that it is closer to the performance of the real keyboard. On the other hand, the KSPC results could be seen as an indication that participants corrected more of their errors on the tactile touchscreen therefore increasing KSPC. Therefore suggesting that the tactile touchscreen helps users to identify errors more easily by providing fingertip-slip feedback while errors could go unnoticed on the standard touchscreen keyboard with no tactile feedback.

There was no main effect for mobility ($F(1,11)=3.42$, $p = 0.07$) and no interaction between mobility and keyboard type ($F(3,33)=0.03$, $p = 0.98$).

6.4.4.3 Time to Enter Phrases

The average time to enter a phrase is shown in Figure 6-13. A 2-factor repeated measures ANOVA on mobility and keyboard type showed a significant main effect for keyboard type ($F(3,33)=70.41$, $p<0.001$), for mobility ($F(1,11) = 10.24$, $p = 0.001$), with a significant interaction between the two ($F(3, 33)=2.92$, $p = 0.03$).

As before, the average time to enter a phrase was significantly affected by keyboard type. Tukey tests showed that the time per phrase on the PDA was significantly lower than on the tactile touchscreen and the standard touchscreen ($p = 0.05$). There was no significant difference between the physical keyboard and the PDA.

As in Experiment 4a(i), it was found that mobility significantly increased the time taken to enter phrases, but this time there was a significant interaction between keyboard type and mobility. The interaction occurred as there was no change in performance in the PDA condition when static and mobile. All other keyboard types performed worse when mobile, but performance with the PDA went from 17.5 to 17.9 seconds per phrase. It is not clear why this occurred and further investigation is needed to see if there is a real effect. It does suggest, however, that the performance with the virtual crossmodal tactile keyboard is robust, with performance in the mobile condition very close to the real physical keyboard.

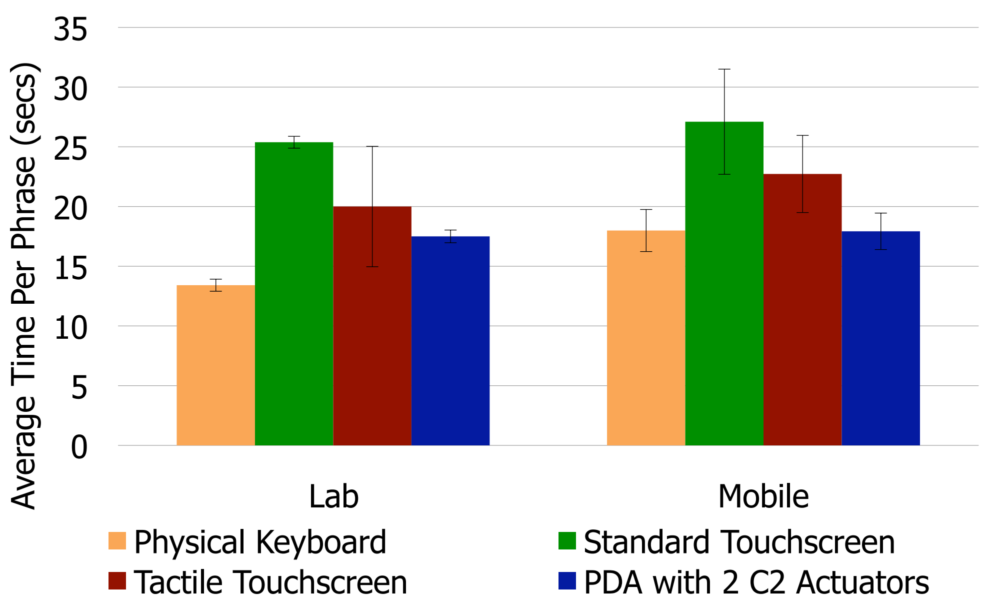

Figure 6-13: Average time per phrase in seconds (with standard deviations).

The average words per minute for each keyboard type along with static and mobile settings are shown in Figure 6-14. The PDA with 2 C2 actuators allowed participants to type 19.15 words per minute whilst mobile. This is slightly better than the original tactile keyboard from the first experiment in Section 6.2 and comparable to the physical keyboard.

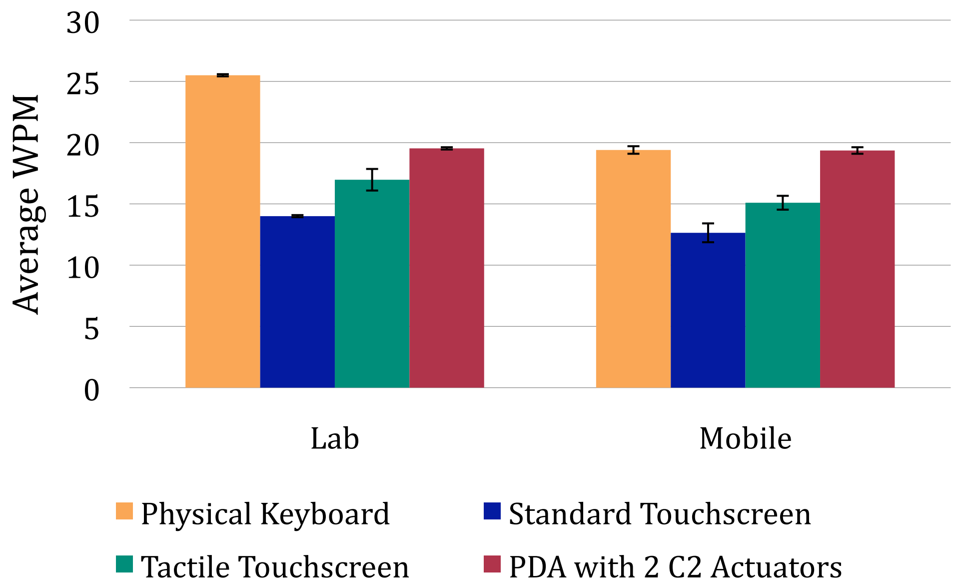

Figure 6-14: Average WPM for each keyboard type in the lab and mobile settings (with standard deviations).

6.4.4.4 Subjective Workload

The results of the NASA TLX questionnaires are shown in Figure 6-15 (a copy of the questionnaire is included in Appendix F). A 2-factor repeated measures ANOVA on overall workload showed a significant main effect for keyboard type ($F(3,33) = 88.62$, $p<0.001$) but no effect for mobility ($F(1,11) = 0.12$, $p = 0.72$) and no interaction between them. A 1-factor ANOVA performed on each of the seven workload factors followed by Tukey tests showed that mental demand, physical demand, temporal demand, effort, frustration were significantly increased and perceived performance was significantly decreased when using the standard touchscreen.

Unlike Experiment 4a(i), the ANOVA and Tukeys showed a significantly higher level of annoyance for the PDA with C2 actuators than with the physical keyboard or with the original tactile touchscreen ($F(2,22) = 35.4$, $p < .0001$). It is not clear why there should be more annoyance in this case, particularly as performance overall was improved. It may be due to the stronger forces that the C2 actuators can apply. Users were not allowed to change the force of the actuators, but that could easily be done in the same way as the volume of the audio can be changed.

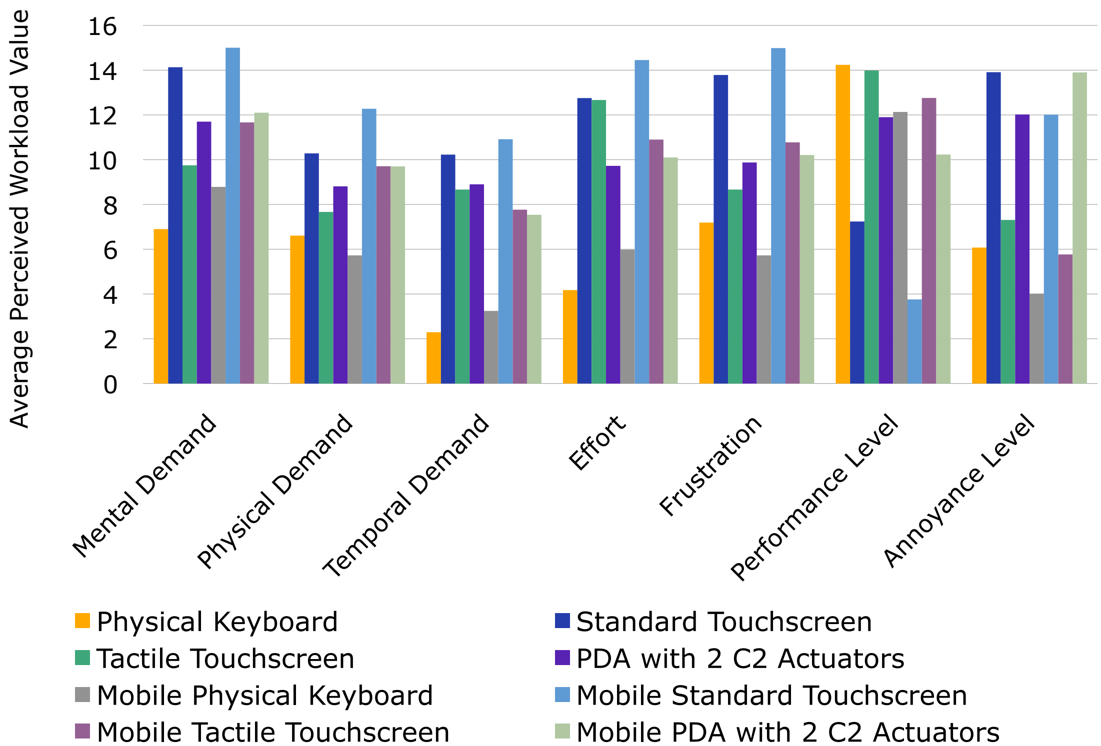

Figure 6-15: NASA TLX Scores Compared with Experiment 4a(i).

6.4.5 Discussion and Conclusions

The two studies reported here have shown that crossmodal tactile feedback can significantly improve interaction and performance with soft keyboards on touchscreen mobile devices. With the addition of this extra tactile feedback the performance of touchscreen keyboards can be brought close to the level of real, physical keyboards. This means that the benefits of touchscreen displays do not come at the cost of poorer text or number entry. It has been demonstrated that crossmodal tactile feedback can benefit touchscreen interaction in both stationary situations and more varying, realistic mobile situations.

Furthermore, a comparison of two different types of tactile actuator showed that text entry can be further improved by using multiple, specialised actuators with the addition of the spatial location parameter. However, the results for both types of crossmodal tactile touchscreen show that user performance is significantly better than when using a touchscreen with no tactile feedback. Therefore, given that the C2 actuators used are expensive and not currently found in standard devices, it would appear to still be beneficial and easier to augment touchscreens with crossmodal tactile feedback using the actuator already present in the phone.

6.5 Experiment 4b: Crossmodal Audio Text Entry

Having shown in Section 6.4 that crossmodal tactile feedback can be beneficial and can be incorporated into a real world application, the next step in this research was to confirm that crossmodal audio feedback could produce comparable outcomes. Therefore, the same experiment was run again in the lab and mobile environments, but this time using the Dell Axim PDA with audio feedback instead of tactile feedback to see if performance can be improved as well as it can be with equivalent crossmodal tactile feedback.

6.5.1 Hardware and Software

In this study, audio feedback was provided through the standard stereo speakers included in the PDA. The audio 3D elements were created using AM:3D (www.am3d.org). The spatial location parameter was encoded in the audio by placing the sounds on a plane around the PDA and participants were asked to hold the PDA at the height of the face to avoid problems related to elevation perception. The sounds were located in front of the nose (0°) and ±90° to the left and right at each ear.

6.5.2 Feedback Design

The audio feedback used in this experiment was identical to that provided in Section 6.4 in that the manipulated crossmodal parameters were the same – rhythm, texture and spatial location. A simple set of crossmodal earcons were developed to match the crossmodal tactons used earlier. There were three different types of information represented by the audio feedback. Rhythm was used to represent the different types of fingertip event and texture was used to indicate successful or unsuccessful fingertip presses. There were three possible values for rhythm as shown in Figure 6-16 representing clicks, slips and fingertip-over events, three possible values for texture representing successful clicks, slips and home keys, and lastly, spatial location simply represented the spatial location of the key on the keyboard.

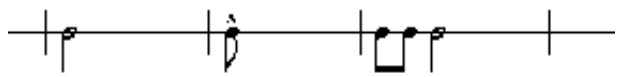

Figure 6-16: Rhythms used to represent fingertip-over, fingertip-click and fingertip-slip events.

6.5.2.1 Fingertip-Over Event

In the tactile version of the experiment, when the fingertip-over event was triggered, a 1-beat smooth 300ms tacton was presented. Therefore an equivalent or crossmodal 1-beat smooth 300ms earcon was created. The smooth texture was represented by a smooth timbre, in this case, a piano (General Midi Patch 01, as determined by the individual parameter experiments detailed in Chapter 4).

6.5.2.2 Home Keys

For each home key ‘F’ and ‘J’ a different textured tacton was presented in the earlier experiment. The tacton was a 1-beat 300ms amplitude modulated 250Hz sine wave, which feels rough. The audio version of this consisted of a 1-beat 300ms rough earcon. The rough texture was created using a different timbre: a tremolo trumpet (General Midi Patch 57).

6.5.2.3 Fingertip-click Event

Previously, when the fingertip-click event was triggered, a 1-beat sharp 30ms tacton was presented. The cue used a 175Hz square wave and no ramp up or ramp down time to create a very short and quick ‘click’ resembling the ‘click’ felt when depressing a physical button. An audio equivalent of this was created using a 1-beat 30ms staccato click played by a glockenspiel (General Midi Patch 10).

6.5.2.4 Fingertip-Slip Event

When the fingertip-slip event is triggered, a 3-beat very rough 500ms tacton was presented in the tactile experiment. The rough texture is created using a square wave. An equivalent earcon was used in this experiment consisting of a 3-beat rhythm lasting 500ms with a very rough texture created by using a vibraphone (General Midi Patch 12).

6.5.3 Methodology

The aims and methodology of this experiment were exactly the same as the previous experiment with the hypothesis for the effects of audio feedback being the same as the predictions of tactile feedback performance:

5. Audio feedback will improve speed and accuracy of text entry on touchscreen keyboards;
6. Touchscreens with audio feedback will achieve comparable accuracy and speed levels to physical keyboards;
7. Audio feedback will improve the speed and accuracy of text entry on touchscreens when mobile.

A new set of 12 participants was recruited from the University, aged between 22 and 34 (7 male and 5 female) with normal hearing and vision. Each participant was presented with 30 random phrases from the MacKenzie phrase set and asked to enter them as quickly and as accurately as possible on the PDA (with audio feedback). Participants used the device in the lab and on the subway. Participants only used this device for the audio touchscreen condition as these results could be compared directly back to Experiment 4a(i) and 4a(ii) for performance on the physical keyboard and standard touchscreen conditions. Once again, speed, accuracy and KSPC were measured.

6.5.4 Results

6.5.4.1 Accuracy

The average number of phrases entered correctly is shown in Figure 6-17 alongside the results from Experiments 1 and 2 for comparison. In the lab condition the touchscreen with audio feedback scored 82.5% correct answers and mobile 70%. The raw data is included in Appendix F.

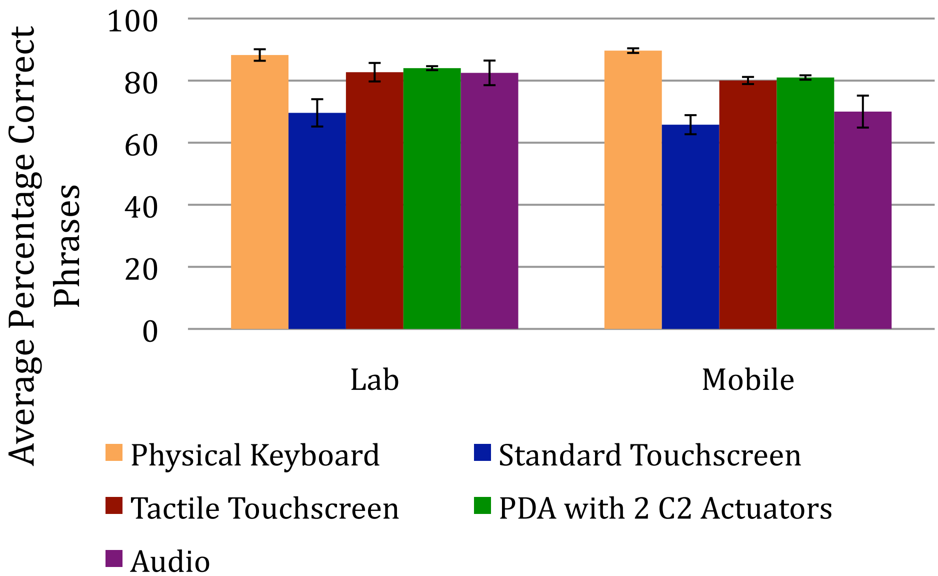

Figure 6-17: Average percentage of phrases entered correctly (with standard deviations).

A 2-factor repeated measures ANOVA was performed on the mean number of correct phrases entered, comparing the effects of mobility (static and mobile) and the five keyboard types (Physical, Tactile and Standard from Experiment 4a(i), the PDA with 2 C2 actuators from Experiment 4a(ii) and the Audio version introduced in this section). Although participants in Experiment 4b had less practice (participants only did PDA with audio in lab and mobile), the first two studies were fully counterbalanced so that a valid comparison can be made between all sets of data from each experiment. The ANOVA showed there was a significant main effect for keyboard type ($F(4,44) = 20.99$, $p<0.0001$). However, there was also a main effect for mobility ($F(1,11)=3.98$, $p = 0.05$) and an interaction ($F(4,44)= 4.58$, $p = 0.05$). *Post hoc* Tukey tests showed that significantly more phrases were entered correctly on the physical keyboard, PDA with 2 actuators, tactile touchscreen and on the audio touchscreen compared to the standard touchscreen ($p = 0.05$) in static settings. There were no significant differences between the tactile touchscreen, PDA, and the audio touchscreen. The physical keyboard, PDA with C2s and tactile touchscreen produced significantly more correct answers than the audio version in the mobile setting.

6.5.4.2 Keystrokes per Character (KSPC)

The KSPC data for Experiments 1 and 2 alongside the audio condition is shown in Figure 6-18. The Audio version scored a mean of 1.08 in the lab and 1.03 when mobile.

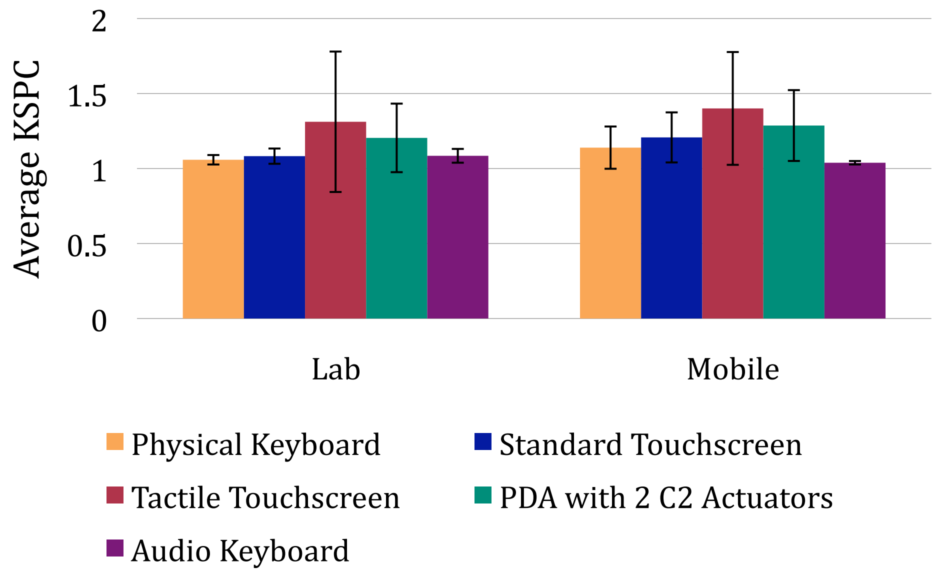

Figure 6-18: Average keystrokes per character (with standard deviations).

A 2-factor repeated measures ANOVA showed no significant differences in the KSPC on the physical keyboard and on the audio keyboard or the standard touchscreen ($F(4,44) = 1.02$, $p > 0.05$). Unlike the results from the tactile touchscreen, when combined with accuracy data it can be seen that there is a lower number of KSPC (i.e. corrections) on the audio keyboard, which correlates with the lower number of correct phrases entered when mobile. However, the number of KSPC is also low in the lab setting but the number of correct phrases entered is high. This suggests that participants did not need to correct many errors in the lab setting with the audio keyboard as they were entering the phrases correctly on the first attempt. In the mobile setting, it suggests that participants did not notice the errors and therefore did not correct them or that they were trying to maintain speed so did not spend time correcting errors (words per minute can be seen in Figure 6-19).

6.5.4.3 Time to Enter Phrases

The average words per minute on the audio keyboard are shown in Figure 6-19. A 2-factor ANOVA on mobility and keyboard type showed a significant main effect for keyboard type ($F(4,44)=70.41$, $p<0.001$), for mobility ($F(1,11) = 10.24$, $p = 0.001$), with a significant interaction between the two ($F(4, 44)=2.92$, $p = 0.03$).

As before, the average time to enter a phrase was significantly affected by keyboard type. Tukey tests showed that the time per phrase on the audio keyboard was significantly higher than on the physical keyboard ($p = 0.05$). In the mobile setting, there was no significant difference between the standard touchscreen and the crossmodal audio keyboard and the chart shows that these levels of words per minute are comparable. This indicates that in the mobile situation, the audio feedback had no effect on typing speed because similar levels of speed were reached when no feedback was provided. As in the other experiments it was found that mobility significantly increased the time taken to enter phrases.

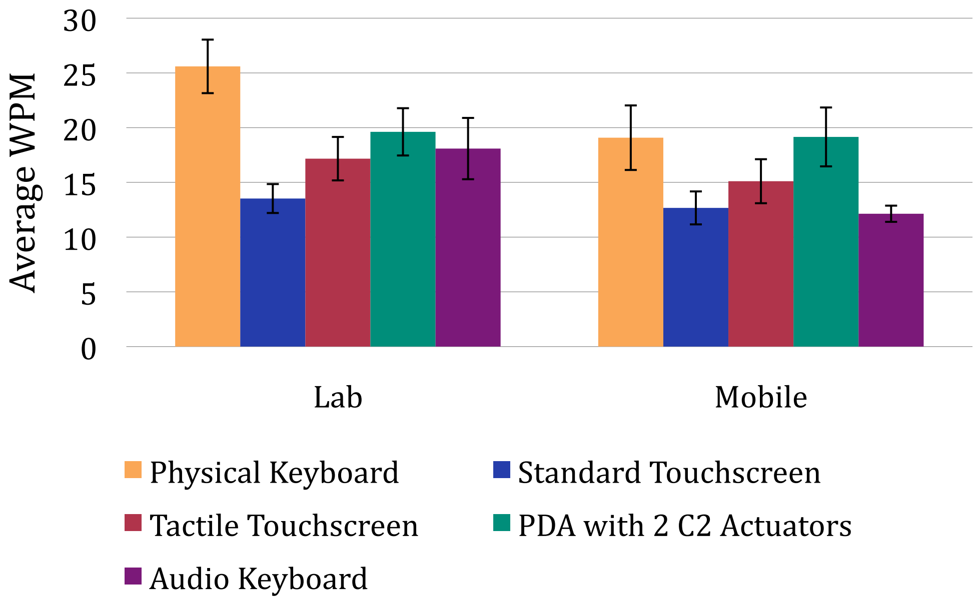

Figure 6-19: The average words per minute for each keyboard type along with static and mobile settings (with standard deviations).

6.5.5 Discussion

The crossmodal audio feedback study reported here has shown that audio feedback can significantly improve fingertip interaction and performance with soft keyboards on touchscreen mobile devices in lab settings reaching comparable levels to those reached on the equivalent or crossmodal tactile version. In the lab, the results for the audio keyboard show that user performance is significantly better than when using a touchscreen with no feedback. This means that Hypothesis 1 can be accepted.

Hypothesis 2 can be partially accepted. The audio keyboard achieved accuracy scores in the lab setting that were comparable to the physical keyboard. However, in both mobile and lab settings, the average number of words per minute was significantly lower than on the physical keyboard.

Hypothesis 3 cannot be accepted because, in the mobile setting, audio feedback was not so effective with levels similar to those achieved on the standard touchscreen showing that audio feedback had no additional benefit. This is most likely because the mobile setting chosen was the underground subway train which is an extremely noisy environment at times. It is probable that the audio feedback was masked by the noises on the train. Other mobile environments with less noise could prove to benefit from audio feedback. These issues are addressed in the experiment detailed in Section 6.6. Furthermore, the results may be improved by providing the audio feedback through personal headphones instead of through the speakers. Speakers were chosen originally because they are present in most mobile devices and users are not required to have additional hardware with them in the form of earphones. It is common for users to have the standard audio feedback turned on from their mobile device and to use the device speakers (for example, audio feedback is provided through the speakers when there is an incoming call). However, headphones are commonly used for listening to music so it is not unrealistic to imagine headphones being used to provide other audio feedback.

6.6 Experiment 5: Crossmodal Text Entry and Environmental Effects

The experiment presented next examines how changing noise and disturbance in the environment affects user performance in a touchscreen typing task with the interface being presented through visual only, visual and tactile, or visual and audio feedback.

This thesis asserts that mobile device users can be in different contexts where one modality may not be as useful as another. It is difficult even to define context let alone measure it. However, it is not so difficult to measure environmental variables such as vibration levels and noise levels which affect the use of audio and tactile displays. Current mobile devices include a variety of built-in sensors such as accelerometers and microphones so can measure environmental values whenever the user interacts with the device [115]. The study discussed here exploits this by using the sensors to establish if it is too noisy for audio or too bumpy for tactile feedback and then allow crossmodal switching to the more appropriate modality.

The experiment described in this section investigated fingertip text entry performance using a QWERTY keyboard displayed on a touchscreen mobile device (with visual, crossmodal audio or crossmodal tactile feedback) in an everyday situation (an underground train). Vibration and noise levels were measured to see if performance on one modality was better than the others at different levels of environmental disturbance. The overall aim was to define the levels at which crossmodal audio or tactile feedback in a real-world setting is no longer valuable to answer Research Question 4.

6.6.1 Experiment Location

The approach used in this experiment involved crossmodal audio and tactile feedback using crossmodal icons. In order to measure extreme vibrations and sounds a controlled environment was needed where high levels of noise and vibration occur naturally. The Glasgow underground was once again chosen as the real-world environment. It is an ideal real-world platform because noise and vibration levels are very dynamic; being quiet and still when stopped at a station, but very noisy and bumpy when the train is in motion. The underground train was previously used successfully in the experiment outlined in Section 6.3 and showed that it was an effective test environment for mobile text entry.

6.6.2 Crossmodal Stimuli and Hardware

A simple set of crossmodal icons were created to represent the different keyboard events and keys that exist on a touchscreen keyboard. This stimuli set was based on the design used in the original text entry experiment detailed in Section 6.4. A fingertip-over event used a 1-beat smooth tacton or earcon, a fingertip-click event used a 1-beat sharp tacton or earcon, while a fingertip-slip event used a 3-beat rough tacton or earcon. On physical keyboards raised ridges are used for orientation. To recreate this, whenever the 'F or J' key triggers the fingertip-over event a different textured tacton or earcon is presented. A Dell Axim PDA was used to display the touchscreen text entry interface.

The C2 Tactor was used again for this study to present the tactile feedback. Audio feedback was created using standard midi wave files designed in an audio synthesis application. The feedback was presented through a standard single earpiece from a set of headphones. Using an earpiece seemed to be a realistic choice as many people use Bluetooth headsets in everyday life. Participants were asked to match the audio volume heard through the earpiece to a given audio file with a sound level of 68dB A weighted (approximately the maximum volume produced by the PDA). This allowed the noise levels to be calibrated and the sound levels heard by users through the earpiece to be estimated before the train journey.

Like the previous experiments in Sections 6.2, 6.4 and 6.5, the visual feedback provided by the QWERTY keyboard was based on the standard built-in feedback found in Windows Mobile i.e. the button highlights when pressed.

6.6.3 Instrumenting the Usability Evaluation

As stated by Crossan *et al.* [37] “instrumented usability analysis involves the use of sensors during a usability study which provide observations from which the evaluator can infer details of the context of use, specific activities or disturbances.” In this case, the factors measured were the accelerations the device was subjected to and the noise level in the environment. To measure movements and disturbances affecting the device that the experiment ran on, the 3DOF linear accelerometer in a SHAKE sensor pack [155] [155]was attached to the back of each participant’s hand holding the device (Figure 6-20).

Figure 6-20: Experiment set-up with PDA, C2 Tactor, bandage securing the SHAKE to hand and text entry GUI screenshot.

The SHAKE logged through Bluetooth to a Samsung UMPC at 90Hz. A handheld sound level meter measured noise levels. To measure device disturbance, the rate of change of acceleration (g/s) was convolved with a rectangular window of one second (90 samples). A Fourier transform was then used to analyse the frequency content of acceleration traces with five minutes of moving train data for participants in each session. The measurable frequency contributions were concentrated between the regions of 5Hz to 20Hz. For 95% of the time, measured accelerations deviated from background gravitational acceleration by < 0.3G. (It must be noted that Andrew Crossan developed the acceleration logging code and aided the author in the analysis).

6.6.4 Methodology

There were 12 participants, 8 male and 4 female, all right-handed, aged between 20 and 25, all staff or students at the University. All participants had experience with QWERTY mobile devices, sending on average 1 to 5 SMS or emails per day. A between-subjects design was used where the conditions were a touchscreen keyboard with: audio, tactile and visual feedback. For each journey, three participants each performed a different condition simultaneously (Figure 6-21). This ensured that all participants were subjected to the same vibration and noise levels at the same time. Overall there were 4 journeys on the underground train.

Figure 6-21: Experiment set-up on underground train.

The methodology and experimental application were based on the previously successful study which measured the effects of crossmodal audio and tactile feedback on touchscreen text entry (see Section 6.2 for full details). The difference here was that the surrounding vibration and noise levels in the real-world environment were measured during text entry to examine their effects on each modality. Instead of having one participant per trial, there were three per journey: one for each condition. Because the participants were all on the same journey they all experienced the same vibration and noise levels at the same time. Therefore speed and accuracy of text entry could be compared in each modality condition in a real world, dynamic environment.

Participants were shown a phrase and asked to memorise it, then type it in as quickly and accurately as possible using the on-screen keyboard (Figure 6-20, right). Each phrase (from a set by MacKenzie [90]) was shown at the top of the screen until the participant began typing at which point it disappeared. A random set of 60 phrases was selected for each train journey. A training period was given before each trial (10 practice phrases) to familiarise participants with the interface and the crossmodal feedback. The dependent variables measured in the experiment were speed, accuracy and keystrokes per character. These were mapped to a vibration and noise level timeline for each train journey (examples of the noise levels experienced during each journey are included in Appendix G).

6.6.5 Results

Before analysing the data from the experiment, it is helpful to put the recorded sound and vibration levels from the underground train into context. Table 6-1 ([102]) shows the average decibels for some everyday situations for comparison (from [19] [76] [108]). In other words, participants in this experiment were subjected to levels similar to a jackhammer and jet engine at times during the journey. Noise levels during the majority of the journey however were similar to the levels of traffic noise.

Source of sound	**Sound pressure level**
	dB re 20 μPa
Jet engine, 100 m distant	110–140
Jackhammer, 1 m distant / discotheque	approx. 100
Traffic noise on major road, 10 m distant	80–90
Moving automobile, 10 m distant	60–80
TV set – typical home level, 1 m distant	approx. 60
Normal talking, 1 m distant	40–60
Very calm room	20–30
Quiet rustling leaves, calm human breathing	10

Table 6-1: Examples of Sound Levels.

6.6.5.1 Accuracy and Keystrokes per Character

To analyse the effects of environmental disturbance, the vibrations and noise were grouped into three blocks of increasing value with the accuracy and speed data for each modality condition mapped to these blocks. The average number of keystrokes per character (KSPC) is shown in Figure 6-23 in parallel with the vibration and noise levels for each modality. The raw data can be found in Appendix G.

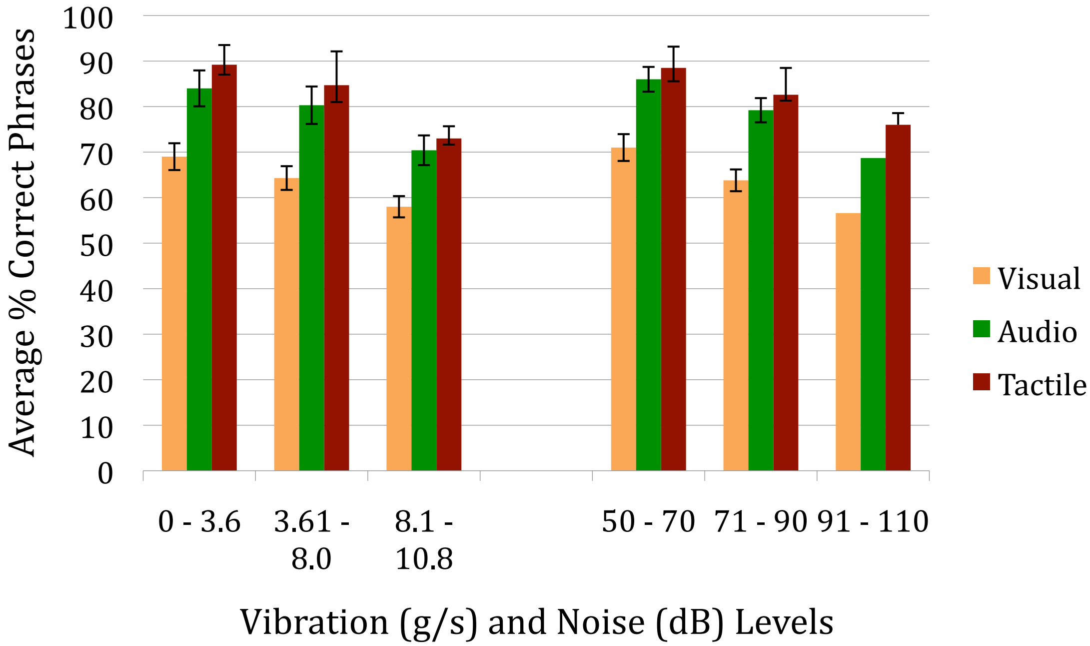

Figure 6-22: Average percentage of phrases entered correctly (with standard deviations).

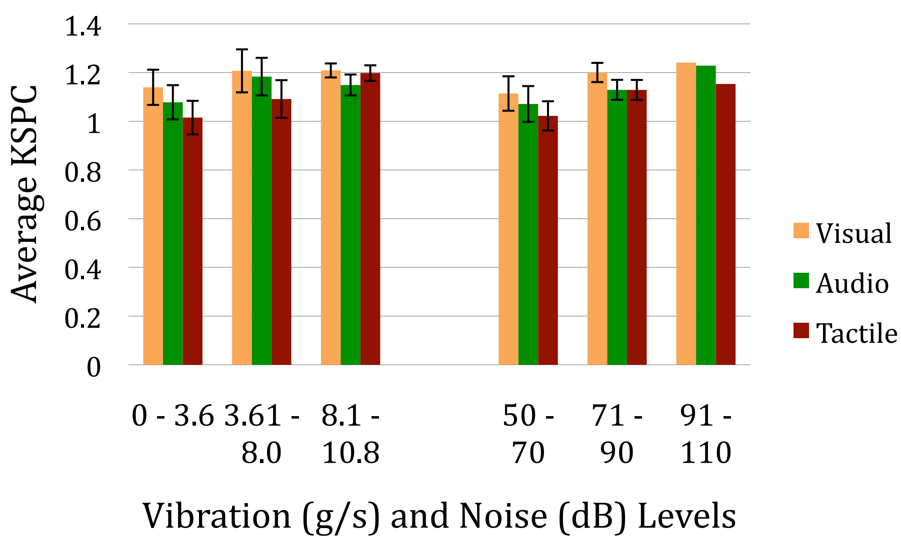

Figure 6-23: Average KSPC for vibration and noise level sets (with standard deviations).

A 2-factor mixed design ANOVA was performed on the mean number of KSPC, comparing the effects of modality (visual, audio and tactile) with three increasing vibration and noise levels. With *post hoc* Tukey's Pairwise Comparisons, a summary of the key results can be seen in Table 6-2 and Table 6-3.

Vibration Level: 3.61–8.0 g/s	Vibration Level: 8.1 – 10.8 g/s
Significantly more KSPC than at level 0 – 3.6 g/s ($F(2,22) = 14.8$, $p < 0.001$)	Significantly more KSPC than at levels 0 – 3.6 g/s and 3.61 – 8.0 g/s ($F(2,22)=14.8$, $p<0.001$)
	Significantly more KSPC in the tactile modality than audio ($F(2,22)=8.22$, $p<0.001$)

Table 6-2: Summary of the KSPC and vibration results.

Sound Level: 71 – 90 dB	Sound Level: 91 – 110 dB
Significantly more KSPC than at 50 to 70dB ($F(2,22)=30.7$, $p<0.001$)	Significantly more KSPC than at 50 to 70 dB and 71 to 90 dB ($F(2,22)=30.7$, $p<0.001$)
	Significantly more KSPC in the audio modality than tactile ($F(2,22)=11.1$, $p<0.001$)

Table 6-3: Summary of the KSPC and noise results.

6.6.5.2 Text Entry Rate (Words Per Minute)

The mean words per minute (WPM) in parallel with vibration and noise levels are shown in Figure 6-24. The raw data can be found in Appendix G.

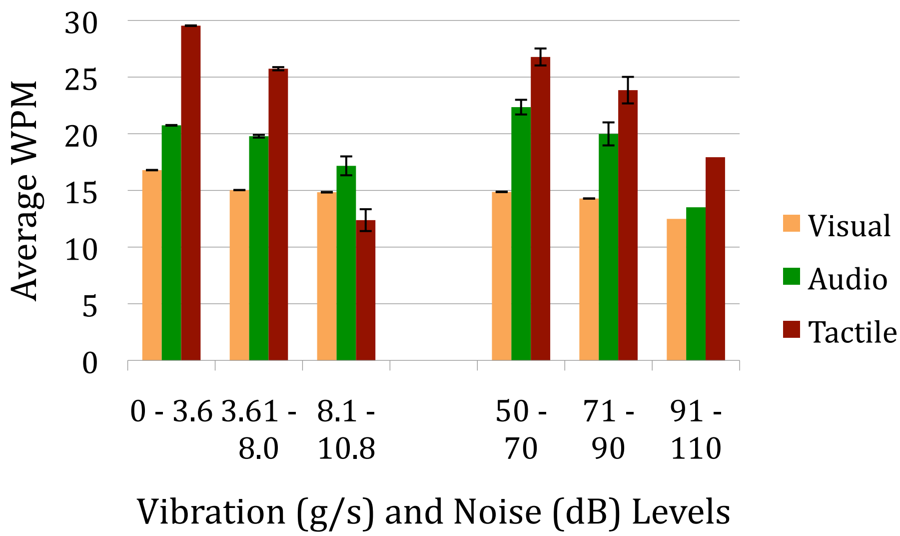

Figure 6-24: Mean WPM for each set of vibration and noise levels (with standard deviations).

A summary of the analysis of typing speed is shown below:

Vibration Level: 3.61–8.0 g/s	Vibration Level: 8.1 – 10.8 g/s
Significantly lower WPM than at 0 to 3.6 g/s (*F*(2,22)=10.9, *p*<0.001)	Significantly lower WPM than at 0 to 10.8 g/s (*F*(2,22)=10.9, *p*<0.001)
	Significantly less WPM using the tactile modality than audio (*F*(2,22)=4.9, *p*<0.001).

Table 6-4: Summary of WPM and vibration results.

Sound Level: 71 – 90 dB	Sound Level: 91 – 110 dB
Significantly lower WPM than at 50 to 70dB (*F*(2,22)=54.3, *p*<0.001).	Significantly lower WPM than at 50 to 90dB (*F*(2,22)=54.3, *p*<0.001).
Significantly less WPM achieved using the audio or visual modality than tactile (*F*(2,22)=2.91, *p*<0.001).	Significantly less WPM achieved using the audio or visual modality than tactile (*F*(2,22)=2.91, *p*<0.001).

Table 6-5: Summary of WPM and noise results.

To determine a more exact point at which these decreases in performance occur, it was necessary to break the data down into smaller blocks of 2dB and 0.2g/s intervals. Statistical analysis using 2-factor mixed design ANOVA followed by Tukey's Pairwise Comparisons show significantly more KSPC at 9.18 to 9.45 g/s in the audio condition compared to lower levels of vibration and at 8.19 to 8.37 g/s in the tactile condition compared to audio ($F(2, 22) = 34$, $p<0.001$). In terms of words per minute, there were significantly less WPM at 9.18 to 9.45 g/s in the audio condition compared to lower levels of vibration and at 8.01 to 8.19 g/s in the tactile condition ($F(2,22) = 23.1$, $p < 0.001$).

Analysis of the noise level data shows that there are significantly more KSPC at 94 to 96 dB in the audio condition compared to lower noise levels and at 100 to 102 dB in the tactile condition compared to audio ($F(2,22) = 4.79$, $p<0.001$). Further analysis shows significantly less WPM achieved at 90 to 92 dB in the audio condition compared to lower noise levels and at 100 to 102 dB in the tactile condition ($F(2,22) = 11.43$, $p < 0.001$).

6.6.6 Discussion

The results show that while crossmodal tactile and audio feedback both improved performance over a visual only interface, they perform differently when the levels of background noise or vibration vary. As expected, as the background noise level increases, the number of KSPC increases faster in the audio condition than other conditions, with a comparable result for background vibration and the tactile condition. Eventually even with high KSPC, the overall accuracy decreases at extreme levels, with performance similar to visual suggesting that participants were not able to use the augmented feedback at these high levels of vibration (for tactile) and background noise (for audio). The high number of keystrokes per character indicates use of the backspace key meaning that users try to correct errors. At the highest levels of vibration, it could be argued that accuracy is lost because it is physically difficult to maintain the finger's position on the screen. For high vibration levels, typing speed and accuracy in the audio modality do not decrease as fast as in the tactile modality meaning that users can continue using audio feedback for longer in these conditions. Again, comparable results occurred for high noise levels and tactile feedback. The analysis shows that typing speed decreases first and then at higher levels, accuracy decreases suggesting that users sacrifice speed first but try to maintain accuracy for as long as possible.

The results of the study suggest that audio feedback becomes ineffective at noise levels of 94 – 96dB and above so tactile feedback should be used instead as there was no significant decrease in performance until 100 – 102dB. Tactile feedback becomes ineffective at vibration levels of 9.18 – 9.45 g/s and above suggesting that audio feedback should be used at these levels. Unfortunately, however, it is often the case that in situations with high vibration levels, there will be high noise levels too. In these circumstances the effectiveness of both audio and tactile feedback will significantly decrease resulting in levels of performance similar to those achieved with visual feedback only.

6.7 Conclusions

This chapter focused on the actual application and evaluation of crossmodal icons in real-world mobile touchscreen systems. The QWERTY keyboard with crossmodal audio and tactile feedback demonstrates one application of crossmodal icons in a mobile touchscreen device for use within unpredictable and ever-changing environments

As mentioned at the start of this chapter, the QWERTY keyboard is a standard touchscreen keyboard making it useful in many applications. The keyboard was augmented with audio and tactile feedback designed in a crossmodal manner as discussed in earlier chapters, so as to gain the advantages of using amodal attributes such as the ability to use modalities interchangeably and present information in the most appropriate modality given the situation or context of the user. The keyboard makes use of very simple multi-dimensional crossmodal icons using rhythm, texture and spatial location. These studies into crossmodal tactile and audio feedback for the keyboard quantified the effects of such feedback and environmental variables in mobile and static settings on user typing performance.

> **RQ3**: Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?

An answer to Research Question 3 is provided by the results from the studies in Sections 6.2 – 6.5 where crossmodal feedback was generally found to be useful when used with a standard touchscreen QWERTY keyboard application. With the addition of crossmodal tactile feedback the performance of touchscreen keyboards can be brought close to the level of real, physical keyboards in both static and more realistic mobile environments. Text entry on the tactile touchscreen only took 22% longer on average than the physical keyboard and the accuracy results between both keyboards are comparable. Overall, the study showed that touchscreen keyboards with tactile feedback produce fewer errors and greater speeds of text entry compared to standard touchscreen keyboards without tactile feedback. Furthermore, a comparison of two different types of tactile actuator showed that text entry on can be further improved by using multiple, specialised actuators which can incorporate the spatial location parameter through localised feedback (the C2 Tactor) as opposed to a single standard actuator which vibrates the whole device. However, the results for both types of tactile touchscreen show that user performance is significantly better than when using a touchscreen with no tactile feedback.

An audio equivalent of the tactile text entry experiment was also conducted showing that audio feedback can significantly improve fingertip interaction and performance with soft keyboards on touchscreen mobile devices in lab settings reaching comparable levels to those reached on the equivalent or crossmodal tactile version. In the lab, the results for the audio keyboard show that user performance is significantly better than when using a touchscreen with no feedback. The audio keyboard achieved accuracy scores in the lab setting that were comparable to the physical keyboard. However, in both mobile and lab settings, the average number of words per minute were significantly lower than on the physical keyboard. Unfortunately, in the mobile setting, audio feedback was not so effective with levels similar to those achieved on the standard touchscreen showing that audio feedback had no additional benefit. This is most likely because the mobile setting chosen was the underground subway train which is an extremely noisy environment at times.

> **RQ4**: Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

The last study in Section 6.6 on the effects of situational impairments on the effectiveness of crossmodal feedback was conducted to resolve Research Question 4. The aim was to determine whether performance with one modality was better than others at different levels of vibration and noise in the environment and at what levels these changes in performance occur. Overall, the data shows that while crossmodal tactile and audio feedback both improved performance over a visual only interface, each modality performs differently when the levels of background noise or vibration vary. As expected, audio feedback was shown to become ineffective in noisy environments and tactile feedback become ineffective in bumpy environments. The novel aspect of this study was that the results revealed the exact levels at which these modalities become ineffective and suggest that manufacturers can use the data obtained from conventional sensors already present in mobile devices to determine the most appropriate feedback modality for users and allow devices to automatically switch between audio and tactile feedback.

Tactile feedback seems to be the most robust and consistently produced the best performance. The only time audio was better was at extremely bumpy times. That being said, almost all of the time, both audio and tactile were better than visual alone. For high vibration levels, typing speed and accuracy in the audio modality does not decrease as fast as in the tactile modality meaning that users can continue using audio feedback for longer in these conditions. Again, comparable results occurred for high noise levels and tactile feedback. The fact that speed decreases first suggests that maybe it is a conscious effort by users who sacrifice speed first but try to maintain accuracy for as long as possible.

Specifically, the results of the study suggest that audio feedback becomes ineffective at noise levels of 94 – 96dB and above so tactile feedback should be used instead as there was no significant decrease in performance until 100 – 102dB. Tactile feedback becomes ineffective at vibration levels of 9.18 – 9.45 g/s and above.

In the real world this would mean switching from tactile to audio feedback when we reach high levels of bumpiness for example when driving on a bumpy road, travelling on a train or jogging. There are still many issues to take in to consideration for crossmodal switching. For instance, when switching between modalities it is important not to distract the user from their task and as Spence [133] points out that there are many crossmodal links in attention between the senses of touch, hearing, and vision. This means that the time taken to shift attention from audio to tactile or *vice versa* should be taken into account and perhaps, when applications automatically switch to the most appropriate modality, the swap should take place in between tasks or when there is enough time for the user's attention shift to occur.

Another issue that should be taken into consideration when using crossmodal switching is user preference; what modality is preferred. Furthermore, the location and social context of the user is also important. For example, the surrounding noise levels in a library are often very low suggesting that the user could hear audio feedback easily. However, it would be socially inappropriate for a mobile device to provide loud audio feedback in such a situation so tactile feedback would actually be a better choice. All of these factors alongside the surrounding environmental noise and vibration levels must be taken into account when using crossmodal interaction, as demonstrated in the following chapter.

Chapter 7 CrossTrainer: Testing The Long-Term Use of Crossmodal Interfaces

As mentioned in the literature review and throughout this thesis, audio and tactile feedback are becoming prevalent features in mobile touchscreen devices and recent studies [11, 72] [83] [111] have indicated that such feedback can be beneficial to users, increasing typing speeds and reducing errors. So far, however, almost all studies have been limited to laboratory-based settings and measurement of performance over approximately one hour. There have been very few long-term studies of earcons and tactons, or of the long-term use of such feedback in mobile applications. The research described in this chapter involved a longitudinal summative evaluation of a touchscreen application with crossmodal feedback for a range of different interface widgets with the aims to investigate the everyday use of crossmodal audio and tactile feedback and to study user performance and preference over time.

In addition to the general examination of the everyday use of crossmodal feedback, this longitudinal study enabled an investigation into the use of such feedback in a variety of different situations. As the thesis statement suggests: as the user's context changes so should the feedback modality. For example, on a building site with high noise levels, tactile feedback may be more appropriate, whereas on a bumpy train ride, audio feedback may be more suitable. The experiments in previous research have involved situations such as the laboratory (Chapter 4), walking on a treadmill (Chapter 5) and travelling on an underground train (Chapter 6), usually with the user's full attention on the experimental task. There are numerous other environments and situations in which users interact with mobile devices. Therefore, another aim of this experiment was to analyse user performance in different situations (in the user's everyday life) to establish whether one modality is more suited than the other and whether crossmodal audio and tactile feedback could be effective in real world applications in different contexts and under different degrees of workload.

Longitudinal studies also allow learning curves to be assessed. The experiments detailed earlier in the thesis tested the identification and use of crossmodal icons after very short training periods commonly around ten minutes (Chapter 5). Although some longer term 2-week studies have taken place [41], 100% performance rates have never been achieved. This study investigated how performance changes after people have been exposed to the crossmodal feedback regularly over an extended period of time. It may prove to be the case that less audio or tactile feedback is required over time as the user becomes more accustomed to the feedback and application, or that in certain situations or types of task, more feedback is required than in others or that overall performance does not improve over time. The results could enable the design of crossmodal displays that adapt according to learning over time.

This study was intended to answer Research Questions 3 and 4:

RQ3: Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?

RQ4: Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

Research Question 3 is addressed in Section 7.2 with an outline of the design of CrossTrainer: a real-world mobile touchscreen application using crossmodal icons and in Sections 7.4 and 7.5 where the methodology and results from the longitudinal study of CrossTrainer are discussed in reference to usability and performance.

Research Question 4 is answered in Section 7.5 where user performance with CrossTrainer with different modalities and the users' choice of modalities is compared in different situations.

7.1 Background

Given that both audio and tactile feedback appear to produce better results than visual feedback alone in terms of performance as demonstrated in Chapters 4, 5 and 6, the question is which modality should be used: audio or tactile?

This research builds on Bernsen's concept of Modality Theory [7] which addresses the mapping of information to different modalities. Modality Theory was introduced to concentrate on the general problem of mapping task domain information into multimodal interfaces. The outcomes of Bernsen's research include a generative taxonomy of output representations and a methodology for information mapping. The methodology focuses on establishing the most appropriate modality given the task whereas the research in this chapter investigates the most appropriate modality for long-term use on a mobile device, and the most appropriate modality for different interface widgets, locations, and situations regardless of task.

Chapter 6 discussed an experiment investigating the most appropriate modality when surrounded by different environmental disturbances on a subway train. The aim of the study was to show at what exact environmental levels audio or tactile feedback becomes ineffective. The results show significant decreases in performance for audio feedback at levels of 94dB and above as well as decreases in performance for tactile feedback at vibration levels of 9.18g/s. These results suggest that at these levels, feedback should be presented by a different modality. The results of the study detailed in Chapter 6 focus on the effects of environmental disturbances on performance not on user preference. In this study, the user's personal modality preference is examined in parallel with surrounding environment levels. Furthermore, the extent to which location and social context affects a user's modality preference is also taken into account.

7.2 The Application: CrossTrainer

Most research tends to focus on design parameters and the type of information encoded in each modality. There are few complete multimodal or crossmodal applications in existence as yet. For this reason CrossTrainer was created: a mobile touchscreen game based on traditional IQ/brain training games. It makes full use of crossmodal audio and tactile feedback allowing modalities to become interchangeable, i.e. to provide the same interaction feedback, enabling users to select the most appropriate modality given their usage context or personal preference.

Crossmodal feedback was incorporated into a game because CrossTrainer requires a great deal of interaction with many different types of interface widget and UI events. Using a game enabled an investigation of a wide range of crossmodal audio and tactile feedback whilst remaining an enjoyable and engaging experience for the test users.

There are 200 questions in CrossTrainer (see Figure 7-1) all of which are designed to test and train the user's IQ. The interface makes use of crossmodal audio or tactile (piezo) feedback for every widget interaction with an additional five random crossmodal audio or tactile (vibrotactile) alerts in each game. Each game of CrossTrainer is made up of a random set of 20 questions each with a time limit of 40 seconds. There are five types of questions involving different audio/tactile feedback: mathematics, true or false, reaction speeds, logical reasoning and general knowledge. Users are required to enter answers via the crossmodal touchscreen widgets (for example, buttons, radio buttons). Upon completion, users are informed of their CrossTrainer IQ score in terms of brain age (similar to many commercial IQ games).

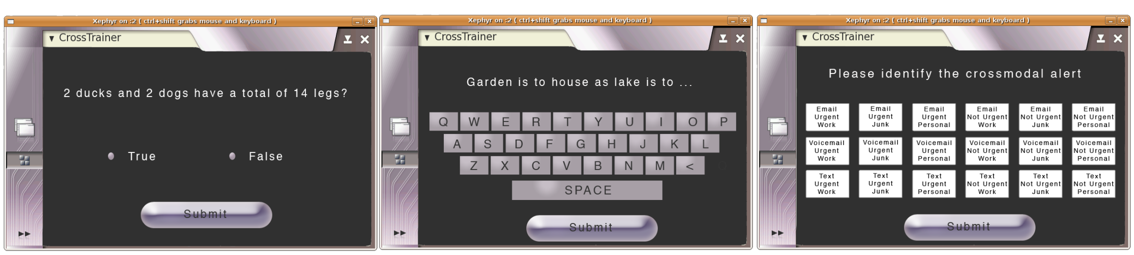

Figure 7-1: CrossTrainer screenshots with example questions a, b, and c.

7.3 CrossTrainer Hardware

CrossTrainer was implemented on the Nokia 770 Internet Tablet, a commercial device that has been augmented with novel piezo-electric actuators [79] (on the left and right behind the touchscreen) and a standard vibration motor. Tactile stimuli were created with a proprietary script language implemented on the device while the audio stimuli use standard midi files played through the device's stereo speakers (or headphones if the user prefers). This novel tactile technology was used to create an intramodal combination [81], i.e. combining feedback from both types of actuator, creating new types of tactile cues not possible before.

7.4 CrossTrainer Stimuli

CrossTrainer uses an audio and tactile feedback design based on crossmodal icons. For standard questions in CrossTrainer as seen in Figure 7-1 a and b, the following three parameters were chosen for the feedback design based on the results of Chapters 4 and 5: rhythm, roughness and location. The type of CrossTrainer widget is encoded in the rhythm (QWERTY button, number button, radio button, scroll bar and notification dialogue), the widget's location on the display is encoded in spatial location (if the buttons are on the left of the screen, the audio feedback will be panned to the left and the tactile feedback will be provided by the piezo actuator under the left-hand side of the screen) and urgency is encoded in texture (i.e. as every 10 seconds pass and the time for the task runs out, the feedback provided by the widgets increases in roughness and intensity). Therefore, 5 different rhythms and 4 different levels of texture produce a set of 20 crossmodal icons: 20 earcons and 20 tactons each capable of providing the same feedback at different spatial locations.

The crossmodal rhythms and spatial location are based exactly on parameters previously used in research on multi-dimensional icons in Chapter 5. However, one of the most novel aspects of the feedback design in CrossTrainer is the different audio and tactile textures used in the crossmodal feedback.

7.4.1 Texture

Two tactile textures were created using different waveforms established in Chapter 4 and investigations into the use of frequency and intramodal tactile textures led to the creation of two completely new textures.

Task urgency is encoded in the texture of each widget. For example, when pressing number keypad buttons in tactile mode, a 2-beat rhythm is used and it becomes increasingly rough as the current game question time limit approaches. This allows users to keep track of how much time is left before an answer must be submitted without having to switch their visual focus away from the task to look at a clock or other type of alert displayed visually on the screen.

Time (secs)	**40**	**30**	**20**	**10**
Texture	Smooth	Semi Rough	Rough	Very rough, high intensity
Tactile	Sine wave	Square wave	Random increasing frequencies	Intramodal combination (piezo and vibrotactile)
Audio	Piano	Tremolo Trumpet	Guiro	Saxophone and violin

Table 7-1: Urgency and texture mapping in CrossTrainer.

As shown in Table 7-1, with 40 seconds remaining for a game question, the tactile rhythm is presented using a smooth piezo-electric pulse like a sine wave, while a piano (General Midi Patch Number 01) plays the audio rhythm. With 30 seconds remaining, the same tactile rhythm occurs when a widget is touched but this time with a rougher texture shaped like a square wave from the piezo-electric actuators and the audio rhythm is played by a tremolo (softly vibrating) trumpet (General Midi Patch Number 57). Then, when there are 20 seconds to go, a much rougher version of the rhythm is presented. This is created using a piezo-electric pulse made up of random increasing frequencies ranging from 1 to 400Hz. The audio is a 10ms burst from a guiro (General Midi Patch Number 73, a percussion instrument played using a scraping motion).

7.4.2 Using an Intramodal Tactile Design

To create a very urgent sensation during the last 10 seconds of each task, a rough and intense (almost bouncy) stimulus has been created using a novel technique involving the use of intramodal combinations. Piezo-electric actuators can create short display-localised tactile bursts, by moving the touchscreen display module [79]. Piezo elements have also been used by Luk *et al.* [89] to create skin-stretch feedback. In this case, the piezo-electric actuators are used to generate short pulses resembling the tactile feedback in physical buttons while the conventional vibrotactile motor is optimised for longer vibrations, where the whole device mass shakes without any localisation. Both the vibrotactile and piezo-electric actuators are activated simultaneously which leads to a sharp piezo bump combined with long rough vibrations (Figure 7-2). The piezo-electric actuator maintains the spatial location parameter while extra strength is added through the vibrotactile actuator. This combination gives a very different feel compared to the standard vibration actuators commonly used in mobile devices.

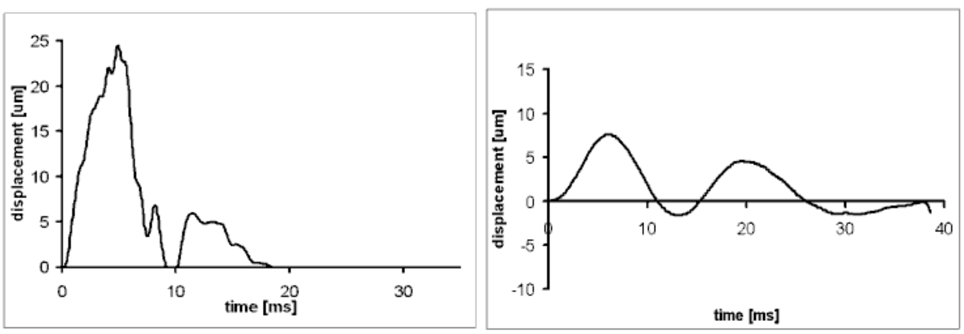

Figure 7-2: Example piezo-electric and vibrotactile output.

Combining two different types of tactile feedback is similar to the use of musical chords in the audio modality played by two different instruments. In this case the audio feedback consists of a chord played by a saxophone (General Midi Patch Number 66) and violin (General Midi Patch Number 41).

7.4.3 Crossmodal Vibrotactile and Audio Alerts in CrossTrainer

In addition to the tactile feedback described above for widget events, CrossTrainer includes crossmodal feedback for alerts such as ‘Urgent Voicemail Received’ as seen in tasks such as Figure 7-1(c). Whilst playing CrossTrainer, participants were presented with alerts randomly throughout each game and asked to identify them after minimal training in the lab. The reason these extra alerts were included was so that there was a mixture of basic and complex crossmodal icons and also to take the previous experiment in Chapter 6 one step further by establishing if it is possible for users to achieve 100% identification rates of more complex cues.

The piezo-electric actuator is capable of providing localised feedback to the fingertip but this means it is only initiated when the user actively touches it. In most mobile devices there are alerts when, for example, there is an incoming phone call. Most often devices use audio feedback for incoming calls and these ringtones are commonly accompanied by vibrotactile feedback from the built-in actuator. Piezo-electric actuators cannot provide these types of alert. So, an EAI C2 Tactor is ideal in this case as it shakes the whole device and can easily catch the attention of the user. The previous experiments in this research have also shown that 3-dimensional earcons played through device speakers match 3-dimensional tactons presented through the C2, if designed in a crossmodal manner. The alert feedback exemplifies the use of crossmodal icons where all three parameters are used – rhythm, roughness and spatial location. The parameter design was as follows:

- Rhythm: type of message as shown in Figure 7-3 (text, email or voicemail)
- Roughness: urgency of message (urgent, semi-urgent, not urgent)
- Spatial Location: message sender (personal, work, junk)

Figure 7-3: Rhythm 1, 2 and 3 used in the alerts.

For example, an urgent personal email would be represented by rhythm 2 with a very rough texture and would be presented on the left-hand side of the device.

7.5 Experiment 6: The Long-Term Use of CrossTrainer

A longitudinal study was conducted to test the cues described above. It used a within-subjects design where all participants completed the tasks under all conditions. A control session was conducted in the laboratory for one hour before participants took the devices home and completed the eight-day study. The lab-based control session was included because the environment can be controlled providing the opportunity to train all participants to use CrossTrainer and to extract measures of their initial performance on each condition for later comparison.

Nine participants took part in the study (3 female, 6 male, all right handed, members of staff or students at the University with an age range of 23 to 32) and all had experience of mobile devices; sending on average four text messages or emails per day on a mobile device. All participants were also somewhat familiar with touchscreen devices although none owned such a device.

There were three main conditions in this study:

- No crossmodal feedback (purely visual)
- Audio feedback
- Tactile feedback

In the first condition, the widgets only provided standard visual feedback during each CrossTrainer game. For the audio and tactile conditions, all widgets provided audio or tactile feedback through the crossmodal icons described above plus the standard visual feedback.

Participants were asked to manually tag their location each time CrossTrainer was played and were also encouraged to leave voicenotes for the experimenter detailing their experiences with CrossTrainer after each game. At the end of the study of CrossTrainer, participants were asked to complete a short post-study questionnaire on their experiences. As motivation to continue to perform well in each game of CrossTrainer, a monetary prize was given to the participant with the highest brain score over the 8-day study.

An additional option was given to participants in the final part of the study after having completed the experiment under all conditions mentioned above. For the final two days, participants could choose their preferred modality of feedback. This additional part of the study provided another method of measuring which of the modalities was most appropriate and most preferred in different situations. The experiment on the underground train discussed in Chapter 6 provided exact measurements of when each modality becomes ineffective. This experiment provides subjective information on user preference for the different modalities and shows if preference changes depending on the situation or location or if, despite the results in Chapter 6, participants choose different modalities to the ones that have been shown to be most effective.

Overall each participant spent 2 days playing the visual version of CrossTrainer, 2 days on the audio version, 2 days on the tactile version and then finally 2 days using the modality of their choice. Participants were asked to play CrossTrainer regularly as much as they liked throughout the 8-day period and were sent reminder emails if CrossTrainer had not been played in the last 24 hours.

The hypotheses in this experiment were as follows:

1. Widget feedback performance will depend on location, situation and modality;
2. CrossTrainer alert and IQ task scores will improve over time for all conditions;
3. 100% recognition rates for crossmodal audio and tactile alerts will be achieved;
4. Modality choice will depend on location, situation and environmental disturbance levels.

CrossTrainer logged the location of the user through manual tagging by participants, surrounding noise levels were measured through the built-in microphone, accelerometer data with a sensor pack attached to the back of the device beside the C2 vibrotactile actuator (detailed later), accuracy (for tasks and alert responses), the time taken to complete tasks and to respond to alerts, and all keystrokes. Participants were asked to enter answers as quickly and as accurately as possible.

7.5.1 Training

All participants attended a lab session during which they were introduced to concepts such as crossmodal feedback and were given the opportunity to use the mobile device so that they became accustomed to the different types of feedback provide. For training in the crossmodal alerts presented by CrossTrainer, the standard Absolute Identification (AI) paradigm was employed where participants receive feedback after each task. The set of stimuli used to train the participants was identical to the set on which they would be later tested. The participants had to identify the information in the cue he/she heard or felt and then choose the appropriate button on the display shown in Figure 7-1(c). Each stimulus alternative was applied twice during each training run, resulting in a total of 36 tasks per run. During training the participants were required to repeat 3 experimental runs (in audio and tactile) in the initial lab control session.

7.5.2 Results

On average participants played CrossTrainer 3 times a day with an average IQ task score of 68.2% on the first day and 73.6% on the last. Additional training data can be found in Appendix H.

7.5.2.1 Crossmodal Alerts

During the training and the experiment itself data were collected on the number of correct responses to the crossmodal alerts. The average learning curves for all participants and each stimulus set during training are shown in Figure 7-4.

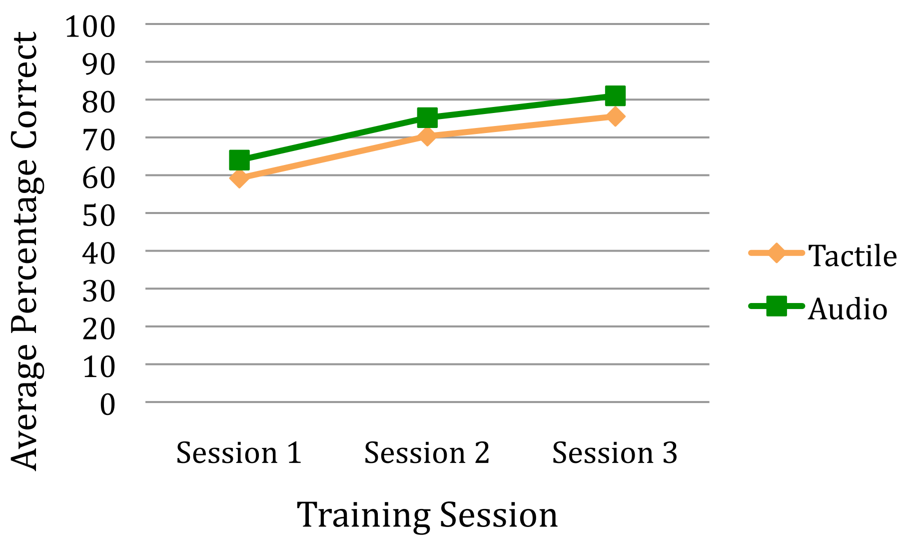

Figure 7-4: Average recognition rates over 3 training sessions.

The performance levels reached by each participant during the training time varied across participants. These results show that, on average, after 3 training games of CrossTrainer (each lasting 10 minutes), participants can identify earcons with recognition rates of 75% or higher (standard deviation = 2.7, 2.4 and 2.12% for each training session). They also show that, on average, it takes 2 training games of CrossTrainer for participants to identify tactons with recognition rates of 75% or above (standard deviation = 2.8, 1.9 and 2.6% for each training session).

Once the participants had completed the training, they were presented with the absolute identification tests randomly throughout the CrossTrainer games during the field study (each participant was exposed to the same number of earcon and tacton alerts). The results for overall recognition of earcon Alerts after the fourth game of CrossTrainer were 100% as can be seen in Figure 7-5 (given that each participant played a different number of times each day this result occurred between days 1 and 2). The alerts using rough textures and short rhythms achieved maximum recognition at the fastest rate while the alerts with medium rough textures and long rhythms resulted in the lowest recognition rate of 61% and only reached 100% during the 6th game of CrossTrainer.

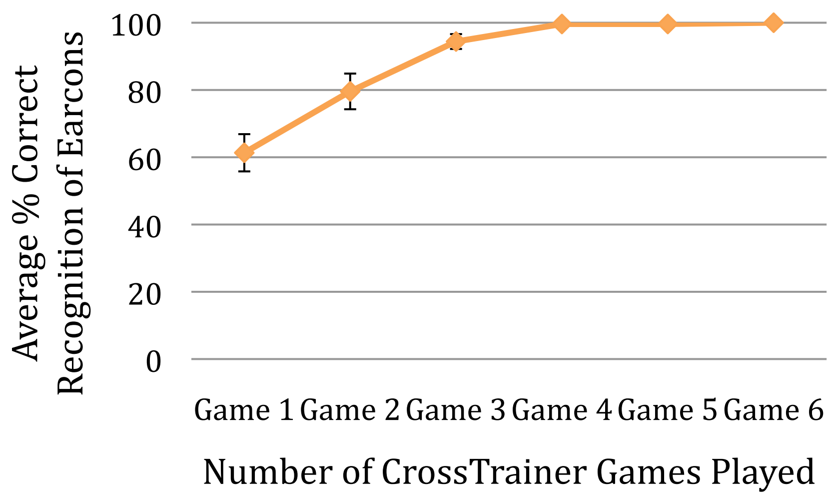

Figure 7-5: Average percentage correct for earcons in each CrossTrainer game (with standard deviations).

The results for overall tacton Alert recognition also showed an average recognition rate of 100% after the third game of CrossTrainer (Figure 7-6). As before, the alert using rough textures and short rhythms achieved the highest recognition rates the fastest and alerts using medium rough textures and short rhythms resulted in the lowest recognition rate of 58% reaching 100% during the last game of CrossTrainer.

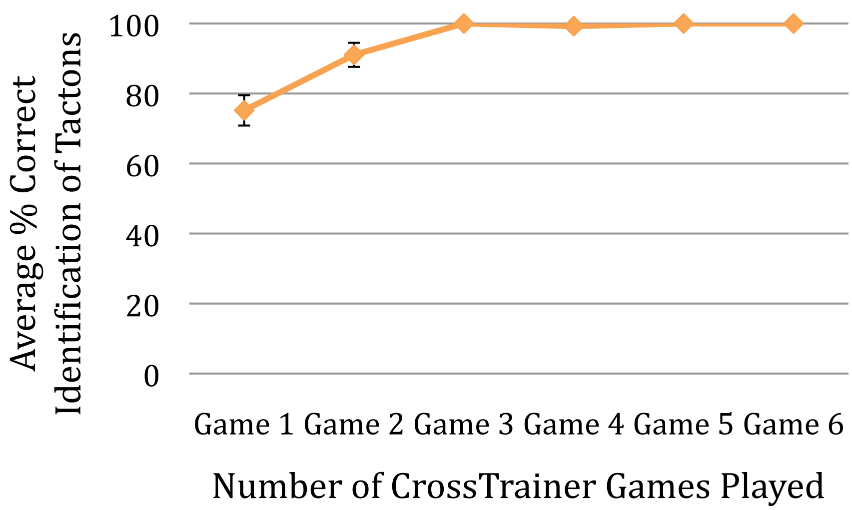

Figure 7-6: Average percentage correct for tactons in each CrossTrainer game (with standard deviations).

Overall, these results show that after 30 minutes of training with crossmodal alerts, participants could recognise the individual modality alerts 75% accuracy, with rates rising to 100% after 4 games of CrossTrainer or in other words, after 40 minutes of playing CrossTrainer.

In the post-study interview, all 9 participants stated they found the crossmodal alerts very easy to identify after the training sessions. As one participant commented "*I could recognise them without even thinking about it after a while*".

7.5.2.2 Performance Over Time: Typing Speeds

Figure 7-7 shows the average words per minute (WPM) for each feedback condition at the beginning and end of the two days spent using each feedback condition. Submitted answers were checked for typos and misspellings. In these cases, the calculation of WPM was the same. During the audio condition, participants typed with an average speed of between 15.2 and 18.6 WPM (words per minute) in their 1st and last games of CrossTrainer. In the tactile condition, participants achieved speeds of between 14.8 and 19 WPM (1st and last games of CrossTrainer) while during the visual condition, text entry took longer with rates of between 13.5 and 14.3 WPM. Raw data can be found in Appendix H.

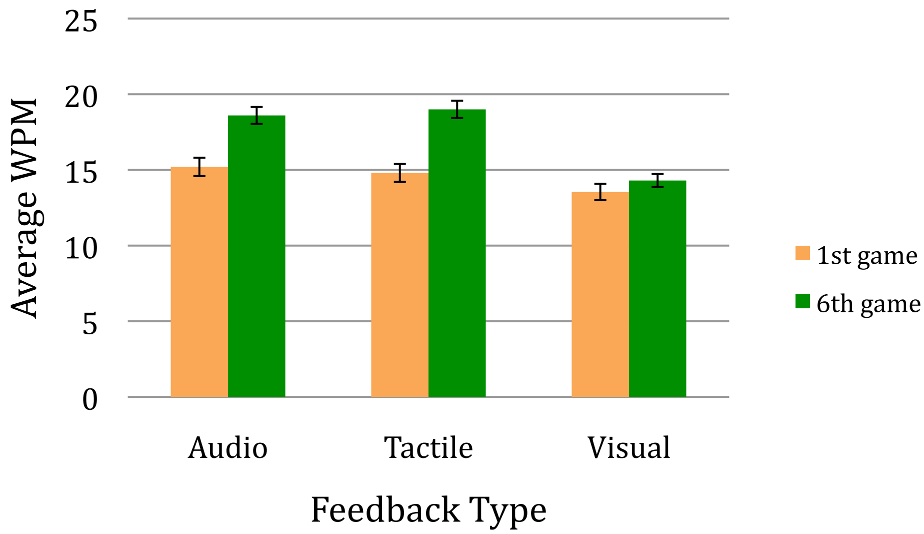

Figure 7-7: Average WPM for each feedback type at the beginning and end of each condition (with standard deviations).

A 2-factor repeated measures ANOVA (alpha level 0.05) on typing speeds for modality types on the 1st and last games of CrossTrainer showed a significant main effect for modality type ($F(2,16) = 14.29$, $p<0.01$). *Post hoc* Tukey HSD tests showed that typing speeds in the visual condition were significantly lower than the audio and tactile ones ($p = 0.05$).

Comments from participants in the voicenotes suggest that participants found it much easier to type on the audio and tactile versions. Six of the 9 participants said they found the tactile keyboard the most effective and 5 of the participants commented that there was no need to look at the screen whilst typing thus increasing the overall speed.

There was also a significant main effect for typing speeds at the start of the first game compared to those at the end of last game ($F(1,8) = 112.11$, $p<0.01$), with typing speeds significantly increasing over the course of each set of 2 days spent on each condition ($p = 0.05$).

Overall these results suggest that typing speeds increase after prolonged use of the application regardless of modality feedback. However, the rate of improvement on the audio and tactile versions is much better than the visual version. The typing speeds achieved on the tactile version of CrossTrainer are comparable to those found by MacKenzie *et al.* [91] for novices typing on touchscreens with a stylus. This first test of long-term use of tactile and audio feedback suggests that they add significant value to typing performance, extending over the longer term.

7.5.2.3 Performance Over Time: Keystrokes Per Character (KSPC)

KSPC were recorded for each game of CrossTrainer. Given that accuracy scores were based on whether or not the submitted answer was correct in terms of the IQ test not if the participants were able to easily and accurately type with the different touchscreen keyboards, KSPC was recorded to examine how many corrections users had to make before submitting an answer. The average number of KSPC for each condition is shown in Figure 7-8 and in Appendix H.

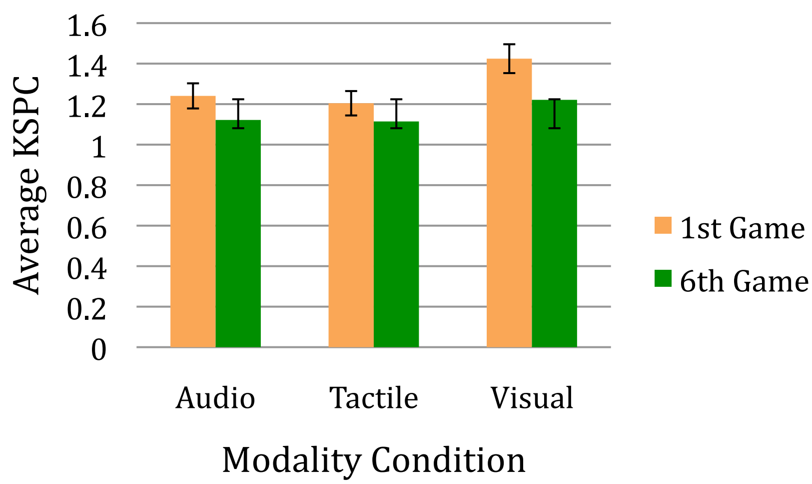

Figure 7-8: Average KSPC for each modality condition from first to last CrossTrainer games (with standard deviations).

A 2-factor repeated measures ANOVA (alpha level 0.05) was performed on the KSPC data comparing the effects of modality on performance during the first and last games of CrossTrainer. A significant main effect on KSPC for modality was found ($F(2,16) = 3.97$, $p<0.01$) over the first and last games of CrossTrainer. Tukey tests showed a significantly higher KSPC when typing on the visual version than on the tactile and audio versions ($p = 0.05$). There were also significant differences between the first and last games ($F(1,8) = 6.21$, $p<0.01$) with less KSPC on the last game than the first game ($p = 0.01$). There was no interaction between modality and number of games played ($F(2,16) = 0$, $p<0.01$). After the last game of CrossTrainer, the tactile version had a lower KSPC than the other modalities.

These results would suggest that by the end of the tactile condition, participants no longer needed to correct as many errors compared to the audio and visual versions. A high number of KSPC is not necessarily bad because this indicates that although participants make errors, they are aware of these errors and make an attempt to correct them. However, the ideal situation would be where there are no corrections required. As mentioned, typing speeds on the tactile version were higher than the audio and visual versions after the last game. This means that after prolonged use, the typing speeds and accuracy on the tactile version of CrossTrainer both improved significantly.

7.5.2.4 Location of Interaction

Table 7-2 shows the distribution of the self-reported locations associated with each game of CrossTrainer. It was found that the most popular location was "at home" with over 53.8% of CrossTrainer games completed there.

Location	**Number of Games Played**	**% of total games**
At home	29	53.8
At work	11	20.4
Commuting	8	14.8
Bar/Restaurant	3	5.5
Other (lecture, friend's house and lab)	3	5.5

Table 7-2: Number and percentage of games played at various locations.

When the location data associated with WPM was analysed, a number of trends were identified (see Figure 7-9). A 2-factor repeated measures ANOVA (alpha level 0.05) was performed on the WPM data for each modality (visual, audio, tactile) used at each of the five locations (home, work, commuting, bar/restaurant, other). The analysis showed a significant main effect for WPM at different locations ($F(4,32) = 11.26$, $p<0.01$). A Tukey test ($p = 0.01$) revealed that a significantly higher WPM occurred in the tactile modality when compared to visual at home and at a bar/restaurant. The analysis also shows that significantly higher WPM ($F(2,16) = 8.76$, $p<0.01$) were achieved in both the audio and tactile conditions compared to the visual when commuting (p = 0.01). There were no other significant differences.

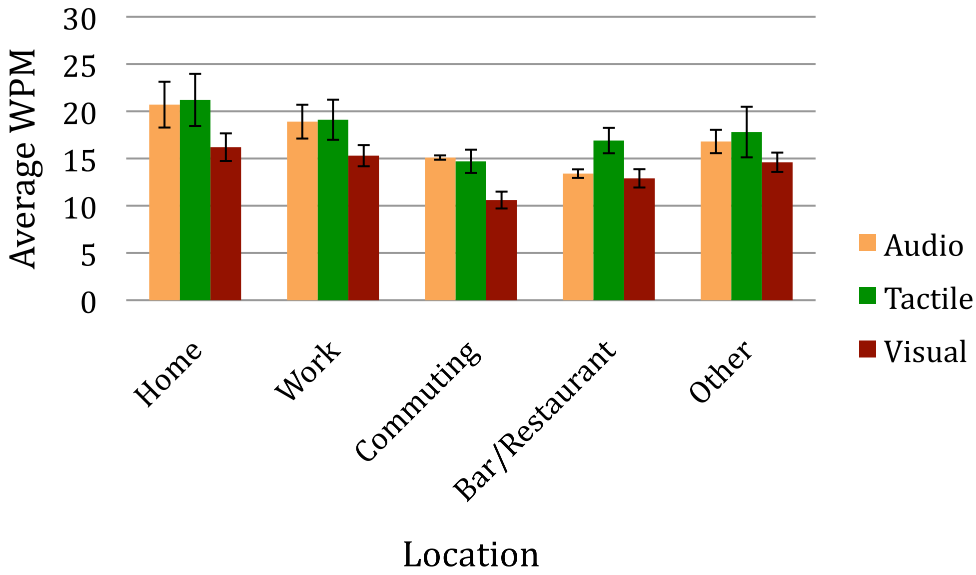

Figure 7-9: Average WPM for each modality per location (with standard deviations).

The average KSPC for each modality and location are shown in Figure 7-10. A repeated measures ANOVA (alpha level 0.05) was performed on the KSPC for each modality (visual, audio, tactile) used at each of the five locations (home, work, commuting, bar/restaurant, other). The analysis showed a significant main effect for KSPC at different locations ($F(4,32) = 9.87$, $p < 0.01$).

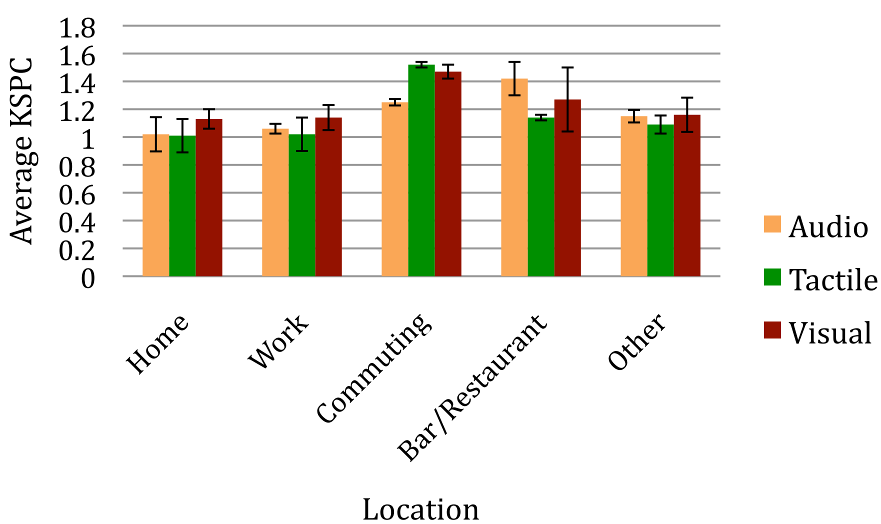

Figure 7-10: Average KSPC for each modality at each location (with standard deviations).

Tukey tests (p = 0.01) revealed that a significantly higher number of KSPC were generated in the tactile modality when compared to the audio version when commuting and a significantly higher number were generated in the audio modality compared to the tactile modality in bars/restaurants. There were no other significant differences.

When at home or at work, WPM in both the audio and tactile modalities improved but the visual version still produced lower typing speeds. In a bar/restaurant tactile performed better (perhaps because it is more socially appropriate than audio feedback). In terms of KSPC, when commuting participants generated a higher number of keystrokes in the visual and tactile modalities than the audio version. This could imply that the audio feedback was not noticeable enough in these locations for participants to recognise errors and correct them. These results are comparable to those discovered in Chapter 6. When at home and at work, both the audio and tactile modalities achieved KSPC levels close to 1.0 which is the ideal number of keystrokes per character. Regardless of location, the visual version resulted in a higher number of KSPC and lower WPM meaning that although participants typed slowly on the visual version, they still made high numbers of errors which required correction.

7.5.2.5 Modality Preference and Location of Interaction

As mentioned earlier, at the end of the CrossTrainer study participants were given two days during which they could choose their preferred modality. When given a choice, participants chose tactile for 82% of the time and audio 18% of the time. The visual only version was never chosen.

In terms of location, the average percentage of votes for each modality can be seen in Table 7-3. Analysis of the number of votes for each modality chosen for each location using Kruskal-Wallis tests showed a significant difference when participants were at home, work, and at a bar/ restaurant (H = 9.87, df = 4, p = 0.05). A Dunn's test revealed that the tactile modality was chosen significantly more often than the audio modality at these locations. There were no other significant differences. Commuting results are comparable in both modalities and in 'other' locations.

	Home	Work	Commuting	Bar/ Restaurant	Other
Audio	22	15.5	48.15	1.85	35.2
Tactile	**78**	**84.5**	51.85	**98.15**	64.8
Visual	0	0	0	0	0

Table 7-3: Average percentage of votes for each modality at each location.

7.5.2.6 Modality Preference and Environmental Levels

During each game of CrossTrainer, aspects of the surrounding environmental context were logged. The factors measured were the accelerations the device was subjected to and the noise level in the environment. To measure movements and disturbances affecting the device that the experiment ran on, a 3DOF linear accelerometer in a SHAKE sensor pack was [155] attached to the back of the device (the author would like to acknowledge Andrew Crossan who helped to develop the code for this part of the CrossTrainer system).

To analyse the effects of environmental disturbance on modality preference, the vibrations and noise were grouped into three blocks of increasing value with the preference data for each modality condition mapped to these blocks using the approach in Chapter 6 (Table 7-4 and Table 7-5).

	Vibration Level: 0 – 3.6 g/s	Vibration Level: 3.61–8.0 g/s	Vibration Level: 8.1 – 10.8 g/s
Audio	7.4%	18.5%	90.74%
Tactile	92.6%	81.5%	9.26%

Table 7-4: Summary of the vibration levels and modality preference.

	Sound Level: 0 – 70 dB	Sound Level: 71 – 90 dB	Sound Level: 91 – 110 dB
Audio	11.2%	42.6%	5.55%
Tactile	88.8%	57.4%	94.45%

Table 7-5: Summary of sound levels and modality preference.

The results suggest that audio feedback becomes the preferred feedback modality at vibration levels of 8.1 g/s and above. Tactile feedback is the preferred modality at vibration levels of 0 - 8 g/s. For noise levels, tactile feedback is the preferred modality for 0 – 70 dB and 91+ dB. Interestingly, when noise levels are between 71 and 90 dB it appears as though both audio and tactile feedback result in similar preference levels. These noise levels are comparable to the noise levels experienced when travelling inside a car.

7.5.2.7 Participant Preference

In the post-study questionnaire and voicenotes, participants explained their reasons for choosing a particular modality for each game of CrossTrainer. A common theme in their answers related to 'social acceptability'. Seven of the nine participants mentioned that they chose tactile over audio because it is less disturbing to other despite the fact that participants were permitted to wear headphones when using CrossTrainer. When commuting, five participants said that they chose audio over tactile because the surrounding vibration levels made it too bumpy for them to feel the tactile feedback. Three participants said that they chose the audio version as often as the tactile version because they found them equally good. Eight of the participants also stated they would like to use both audio and tactile at the same time on some occasions.

Participants also mentioned that, for certain tasks, audio would be better than tactile and *vice versa*. Six out of nine participants said they would prefer audio feedback for small widgets such as radio buttons and tactile feedback for larger ones such as progress bars. Eight participants stated that, for tasks requiring a large amount of interaction for example, typing a paragraph on a keyboard, they would choose to use audio feedback and seven participants stated that, for important tasks such as 'delete' or 'close', they would like the ability to choose to use combined audio and tactile feedback.

7.5.3 Discussion

The 8-day study of CrossTrainer generated many interesting results. As far as the author is aware, this is the first longer-term study of user preference and performance for crossmodal audio and tactile feedback on mobile touchscreens. Participants were allowed to play CrossTrainer whenever and wherever they wished providing 72 days worth of data from a wide range of different locations. Furthermore, the feedback design in CrossTrainer is also novel as it uses a combination of piezo-electric and vibrotactile feedback which has not been explored before.

Throughout the CrossTrainer study, three areas were explored:

- The effects of longer term use, location and modality on performance with CrossTrainer;
- Whether 100% recognition rates can be achieved for crossmodal audio and tactile icons;
- The effects of location, situational context and environmental levels on modality preference.

In terms of performance changes over the 8-day study, the results showed that typing speeds were significantly faster at the end of the study for both audio and tactile versions. Analysis also showed that less KSPC occurred for the audio and tactile versions in the last game of CrossTrainer. Given the results of previous research these outcomes are not entirely unexpected but the data show that although performance can improve with audio and tactile feedback, performance with visual feedback remained consistently lower even after 8 days of use. In the words of one participant, "*I could never get the hang of the visual CrossTrainer, I tried to type as fast as I could but I never noticed my mistakes until it was too late, it doesn't feel natural*".

Location also had an effect on typing speeds and KSPC for each feedback condition. As mentioned, the majority of previous research has been static, i.e. it was lab-based or took place in a single location. By conducting this research as part of the users' everyday lives, it has been possible to record users' WPM and KSPC at different locations and the results show that location can affect the performance in each modality. For example, when the majority of participants recorded their location as 'commuting', WPM in all modalities was considerably lower but still significantly faster than the visual version. Five of the participants commuted via bus or underground train and the other 4 classed walking as commuting.

Through the post-study questionnaires it became apparent that location affected tying speeds and KSPC for a number of reasons. Participants stated that using CrossTrainer while commuting was difficult because of the surrounding environmental sound and vibration levels whereas when using it at work or in a bar/restaurant surrounded by other people made it embarrassing to use the audio version for fear of disturbing others.

As predicted, recognition rates for crossmodal alerts did indeed reach 100%. The results for overall earcon recognition after the fourth game of CrossTrainer showed an average recognition rate of 100%. The results for overall tacton recognition showed an average recognition rate of 100% after the third game of CrossTrainer. This is the first study where such high performance levels have been recorded and shows the users can learn such tactile and audio cues.

Interestingly, there were many outcomes from the analysis of personal modality preference. The experiment on the underground train discussed in Chapter 6 provided exact measurements of when each modality became ineffective. The experiment described here provided subjective information on user preference for the different modalities and showed if personal preference changed depending on the situation or location at which participants played CrossTrainer. There is little point in providing an adaptable style of feedback that switches depending on surrounding noise and vibration levels if it switches to modalities that users do not want. When given a choice of modalities, participants chose tactile for 82% of the time and audio 18% of the time. The visual version of CrossTrainer received no votes. Environmental vibration and noise levels appear to have an effect on the choice of modality with audio feedback chosen when surrounded by high vibration levels and tactile feedback chosen when surrounded by both high and low noise levels.

In the post-study questionnaire and voicenotes, participants explained their reasons for choosing a particular modality for each game of CrossTrainer. A common theme in their answers related to ‘social acceptability’. In other words, when in the company of others it can be embarrassing to use audio feedback on a mobile device and it may be considered rude to wear headphones.

Lastly, when participants were asked about the complexity of the audio and tactile feedback in CrossTrainer, most of the comments from participants changed over the 8 days. At the beginning participants appreciated all of the crossmodal feedback but by the end, they said ‘less is more’. As the participants became more experienced less feedback was required. The CrossTrainer logs also indicate that participants often moved on to the next interaction before the previous feedback had completed. Therefore, the duration of feedback should also be reduced over time.

7.6 Conclusion

To conclude, this chapter has described a research prototype called CrossTrainer which makes use of novel crossmodal audio and tactile feedback on a mobile touchscreen device.

The following research questions were addressed:

RQ3: Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?

RQ4: Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

By applying all of the previous work on crossmodal icons this chapter has answered Research Question 3 by showing that crossmodal applications can be created where different modalities can provide the same interaction feedback, and are therefore interchangeable. An 8-day field study of CrossTrainer was carried out involving 9 participants focusing on elements such as the longitudinal effects on performance with audio and tactile feedback, the impact of context such as location and situation on performance and personal modality preference.

This research shows that the crossmodal feedback can aid users in entering answers quickly and accurately using a variety of different widgets. This study has shown that users can switch between modalities and reach 100% recognition rates after 2 days of regular use suggesting that crossmodal feedback is a viable option in touchscreen applications.

With respect to Research Question 4, the results suggest that, when choosing between audio and tactile feedback for a mobile touchscreen application, the following aspects should be taken into account:

- Environmental noise and vibration levels
- Preference
- Location
- Period of use

There are clearly times when audio is more appropriate than tactile and *vice versa*. For this reason devices should support crossmodal tactile and audio feedback to cover the widest range of environments, preference, locations and tasks.

Chapter 8 Discussion and Conclusions

This thesis has investigated crossmodal audio and tactile interaction with mobile touchscreen displays through the use of crossmodal icons. In Chapter 1, the thesis statement was as follows:

This thesis asserts that using crossmodal auditory and tactile interaction can aid mobile touchscreen users in accessing data non-visually and, by providing a choice of modalities, can help to overcome problems that occur in different mobile situations where one modality may be less suitable than another. By encoding data using the crossmodal parameters of audio and vibration, users can learn mappings and translate information between both modalities. Therefore, data may be presented to the most appropriate modality given the situation and surrounding environment.

The thesis statement and the following four research questions have been addressed throughout the thesis:

RQ1: What are the parameters of vibration and non-speech audio that can be manipulated to encode data in crossmodal icons?

RQ2: What levels of performance can be achieved when these parameters are combined to create multi-dimensional crossmodal icons?

RQ3: Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?

RQ4: Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

These four questions have been addressed through a review of related literature and a series of empirical studies evaluating individual parameters, multi-dimensional crossmodal icons and the application of crossmodal icons for everyday mobile touchscreen use.

This chapter summarises the work reported in this thesis and discusses how the findings answer the four research questions above. It then sets out a series of guidelines derived from this research, which could be employed by designers or researchers who wish to make use of crossmodal audio and tactile icons in mobile touchscreen applications. Then possibilities for future work in this research area are described. Finally, general conclusions are drawn from this research, with a focus on the main contributions of this thesis.

8.1 Thesis Summary

Chapter 2 reviewed related work on perception and the presentation of information through the tactile and audio modalities along with current research on the use of mobile touchscreen devices. Several parameters were identified through the review as potential crossmodal parameters including spatial location, roughness/texture, and rhythm. The review also identified previous studies that showed both audio and tactile feedback could be beneficial to touchscreen interaction in a unimodal capacity, meaning that crossmodal use of these modalities could be a potentially fruitful route of investigation.

Chapter 3 outlined the definition of crossmodal interaction as used in this research with a focus on initial perceptual studies in the field of psychology. The other main aspect of this chapter was the introduction of crossmodal icons and their parameters. The approach taken in this research was to base the design of crossmodal icons on the design principles employed in the creation of their auditory and tactile sub-parts. Although the research in earcons and tactons has some similarities, these icons have never been combined and their amodal attributes have never been exploited to aid in mobile touchscreen use.

Chapter 4 reported two experiments (with two follow up experiments) investigating the different possible parameters and mappings that can be used to facilitate crossmodal auditory/tactile feedback through rhythm, texture and spatial location. Previous research was shown to already establish rhythm as an effective parameter in both modalities and results are comparable for both the audio and tactile versions. The two experiments focusing on roughness and texture showed that crossmodal roughness in the auditory domain should be created using either amplitude modulation or differing timbres and that different waveforms (sine, square and sawtooth) can be used as the roughness parameter in tacton design. In terms of spatial location, it was shown that 3D audio positions can be mapped to tactile body positions on the waist and wrist most effectively and that there are significantly more errors made when using the ankle. In mobile situations, 3D audio positions can be mapped to tactile body positions on the waist most effectively. The spatial location experiments have shown that it is possible for users to match spatial locations in the auditory and tactile domains.

Chapter 5 discussed the development of a set of multi-dimensional crossmodal icons, and then reported an experiment investigating the learning of such icons and the extent to which this learning transfers between the two modalities. This research investigated whether, if trained to understand multidimensional audio alerts, a user can then also understand the corresponding tactile alerts with no additional training and *vice versa*. Results showed that an identification rate of 92% can be achieved for three-dimensional audio crossmodal icons when trained in the tactile equivalents, and identification rates of 89% can be achieved for tactile crossmodal icons (using tactile waveforms to create roughness) when trained in the audio equivalent. Users in a mobile environment can accurately recognise 78% of messages presented by earcons, if they have been trained to recognise the same alerts presented by tactons. Similarly, users in a mobile environment can accurately recognise 79% of messages presented by tactons, if they have been trained to recognise the same alerts presented by earcons. The results indicate that it may not be necessary to train users to understand icons in all the modalities a system might use. If crossmodal icons are used to present information, training is only required in one modality as results show that users can then understand the same messages in the other modality.

Chapter 6 examined the incorporation of multi-dimensional crossmodal icons in a mobile touchscreen application with an aim to find out if situationally impaired users can benefit from such crossmodal feedback. The QWERTY keyboard with crossmodal audio and tactile feedback demonstrated one application of crossmodal icons in a mobile touchscreen device for use within unpredictable and ever-changing environments. Overall, the first study showed that touchscreen keyboards with tactile feedback produce fewer errors and greater speeds of text entry compared to standard touchscreen keyboards without tactile feedback. An audio equivalent of the tactile text entry experiment was also conducted showing that audio feedback can significantly improve fingertip interaction and performance with soft keyboards on touchscreen mobile devices in lab settings reaching comparable levels to those reached on the equivalent or crossmodal tactile version.

The other study in this chapter focused on the effects of situational impairments on the performance of crossmodal feedback. The experiment examined how changing noise and disturbance in the environment affects user performance in a touchscreen typing task with the interface being presented through visual only, visual and tactile, or visual and audio feedback. The aim of the study was to show at what exact environmental levels audio or tactile feedback becomes ineffective. Overall, the data showed that while tactile and audio feedback both improved performance over a visual only interface, each modality performs differently when the levels of background noise or vibration vary. As expected, audio feedback was shown to become ineffective in noisy environments and tactile feedback become ineffective in bumpy environments. However, this study established the exact levels at which these performance decreases occur. The thesis declares that, as the context changes, so should the feedback modality. If the mobile device could automatically switch to the most effective type of feedback based on these experiment results, this could lead to greater usability, more socially appropriate interaction and less redundant feedback.

Chapter 7 involved a longitudinal summative evaluation of a touchscreen application, CrossTrainer, which makes use of novel crossmodal audio and tactile feedback on a mobile touchscreen device. The aim was to investigate the everyday use of crossmodal audio and tactile feedback and to study user performance and preference over time. An 8-day field study of CrossTrainer was carried out involving 9 participants focusing on elements such as the longitudinal effects on performance with audio and tactile feedback, the impact of context such as location and situation on performance and personal modality preference. This study showed that crossmodal feedback aids users in entering answers quickly and accurately using a variety of different widgets. Furthermore, the results demonstrate that users can switch between modalities and reach 100% recognition rates of multi-dimensional crossmodal alerts after 2 days of regular use suggesting that crossmodal feedback is a viable option in touchscreen applications. Overall, when choosing between audio and tactile feedback for a mobile touchscreen application, environmental noise and vibration levels, personal preference, location and period of use should be taken into account.

8.2 Research Question 1

What are the parameters of vibration and non-speech audio that can be manipulated to encode data in crossmodal icons?

Research Question 1 is answered in Chapters 2, 3 and 4. The review of audio and tactile perception in Chapter 2 provided some insight into the potential parameters of sound and vibration for use in crossmodal icons. The findings of the review indicated that the most successful parameters for encoding information in vibrotactile messages, such as tactons, are spatial location, roughness and rhythm. The review also indicated that the most successful parameters for encoding information in audio messages, such as earcons, are timbre, pitch, rhythm, duration and spatial location.

Chapter 3 introduced the concept of amodal attributes: the parameters available in both the senses of touch and hearing that can be used to represent the same information. These include intensity, spatial location, rate, texture and rhythmic structure [86]. Based on these findings, the parameters of vibration and audio that can be manipulated to encode data in crossmodal icons are a subset of the most successful parameters in earcon and tacton research that have also been identified as amodal attributes: rhythm, texture and spatial location.

To verify the parameters as suitable for crossmodal icons, two experiments were conducted in Chapter 4. These experiments indicate that rhythm can be used to encode data in the audio and tactile modalities, and that rhythms in both modalities can be perceived as equivalent. When crossmodal texture is created using audio timbre and tactile waveforms, users perceive a match between the information in both modalities at a rate of 94.2%. Lastly, spatial location can be used as a crossmodal parameter in both static and mobile settings. Experiments showed that spatial location can be perceived as equivalent when using a 3D audio soundscape and tactile locations on the waist.

This thesis concludes that rhythm, roughness and spatial location can be used to encode data in crossmodal icons. Identification rates of 52 - 94% have been achieved for these parameters in a lab-based setting. In addition, when in a mobile setting identification rates of 43 – 91% have been reached. Furthermore, the sub-study described in Chapter 5 showed that, by using tactile waveforms in the roughness parameter, identification rates rose from 52% to 72%.

8.3 Research Question 2

What levels of performance can be achieved when these parameters are used to create multi-dimensional crossmodal icons?

Research Question 2 is answered in Chapter 5 through an experiment evaluating three-dimensional crossmodal icons. The results of which provided identification rates for these crossmodal icons, and also the extent to which users' abilities to learn the meaning of crossmodal icons in one modality can be transferred to the other modality.

More specifically, the results from the experiment demonstrated that an identification rate of 92% can be achieved for three dimensional audio crossmodal icons when trained in their tactile equivalents, and identification rates of 89% can be achieved for tactile crossmodal icons (using tactile waveforms to create roughness) when trained in the audio equivalent. Users in a mobile environment can accurately recognise 78% of messages presented by earcons, if they have been trained to recognise the same alerts presented by tactons. Similarly, users in a mobile environment can accurately recognise 79% of messages presented by tactons, if they have been trained to recognise the same alerts presented by earcons.

The results of this research indicate that it may not be necessary to train users to understand icons in all the modalities a system might use. One concern with using lots of different modalities is the increase in complexity, however crossmodal interaction does not cause this. In fact, by eliminating the need for further user training with the addition of more modalities, crossmodal interaction can avoid such complexities. If crossmodal icons are used to present information, training is only required in one modality as results show that users will then be able to understand the same messages in the other modality. Using crossmodal icons to communicate information to mobile device users could therefore reduce the learning time for the user and also increase the number of modalities through which this information may be transmitted.

8.4 Research Question 3

Can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications?

Research Question 3 is answered in Chapter 6 where section 6.2 covers the development of crossmodal feedback for a touchscreen keyboard application through a lab-based experiment, followed by a mobile version of the study based on an underground train. The application and following related experiments established the effects of crossmodal feedback on user performance levels with a touchscreen QWERTY keyboard using a text entry task.

Overall, crossmodal feedback was found to be useful when used with a standard touchscreen QWERTY keyboard application. The study showed that touchscreen keyboards with audio or tactile feedback produce fewer errors and greater speeds of text entry compared to standard touchscreen keyboards without audio or tactile feedback.

Text entry on the tactile touchscreen only took 22% longer on average than the physical keyboard and the accuracy results between both keyboards are comparable. Furthermore, a comparison of two different types of tactile actuator showed that text entry can be further improved by using multiple, specialised actuators which can incorporate the spatial location parameter through localised feedback (the C2 Tactor) as opposed to a single standard actuator which vibrates the whole device.

The audio keyboard achieved accuracy scores in the lab setting that were comparable to the physical keyboard. However, in both mobile and lab settings, the average number of words per minute were significantly lower than on the physical keyboard. Unfortunately, in the mobile setting, audio feedback was not so effective with levels similar to those achieved on the standard touchscreen showing that audio feedback had no additional benefit. This is most likely because the mobile setting chosen was the underground subway train which is an extremely noisy environment at times.

With the addition of crossmodal audio or tactile feedback, typing accuracy on touchscreen keyboards can be brought close to the level of real, physical keyboards in both static and more realistic mobile environments. Tactile feedback can also significantly improve typing speeds in noisy mobile situations compared to keyboards with no feedback. Overall, the experiments showed that crossmodal applications can be created where different modalities can provide the same interaction feedback, and are therefore interchangeable.

In Chapter 7, a study of CrossTrainer (a mobile touchscreen game with crossmodal feedback) was detailed involving 9 participants focusing on elements such as the longitudinal effects on performance with audio and tactile feedback, the impact of context such as location and situation on performance and personal modality preference.

This research shows that the crossmodal feedback can aid users in entering answers quickly and accurately using a variety of different widgets. In terms of performance changes over the 8-day study, the results showed that typing speeds were significantly faster at the end of the study for both audio and tactile versions. Analysis also showed that less keystrokes per character (KSPC) occurred for the audio and tactile versions in the last game of CrossTrainer. Performance with visual feedback remained consistently lower even after 8 days of use.

This study also showed that users can switch between modalities and reach 100% recognition rates for multi-dimensional crossmodal icons after 2 days of regular use suggesting that crossmodal feedback is a viable option in touchscreen applications. This is the first study where such high performance levels have been recorded and shows the users can learn such tactile and audio cues.

8.5 Research Question 4

Given different contexts and situations, what type of feedback (audio or tactile) is most appropriate?

Research Question 4 is answered in Chapters 6 and 7. Another study was conducted on an underground train to establish how changing vibration and noise levels in the surrounding environment affect the perception and usefulness of crossmodal feedback. There have been no such experiments before. The aim was to investigate how the usability of crossmodal feedback alters as a user's surroundings alter. More specifically, the aim was to determine whether performance with one modality was better than the other at different levels of vibration and noise in the environment and at what levels these changes in performance occurred. Overall, the data showed that while tactile and audio feedback both improved performance over a unimodal visual interface, each modality performed differently when the levels of background noise or vibration varied. As expected, audio feedback was shown to become ineffective in very noisy environments and tactile feedback ineffective in very bumpy environments. The most important contribution of this study was the discovery of the exact levels at which these modalities become ineffective for this context of use (for this particular device using the crossmodal feedback design outlined earlier). The results of the study suggest that manufacturers may be able to use the data obtained from conventional sensors already present in mobile devices to determine the most appropriate feedback modality for users and allow devices to automatically switch between audio and tactile feedback.

In Chapter 7 user performance on CrossTrainer with different modalities and the users' choice of modalities was compared in different situations. As far as the author is aware, this is the first longer-term study of user preference and performance for crossmodal audio and tactile feedback on mobile touchscreens.

With respect to Research Question 4, the results suggest that, when choosing between audio and tactile feedback for a mobile touchscreen application, the following aspects should be taken into account: environmental noise and vibration levels, preference, location and period of use.

Location had an effect on typing speeds and KSPC for each feedback condition. By conducting this research as part of the users' everyday lives, it has been possible to record users' WPM and KSPC at different locations and the results show that location can affect the performance in each modality.

It was shown that typing speeds were faster when using the tactile version of CrossTrainer compared to the visual version when in a bar/restaurant (perhaps because it is more socially appropriate than audio feedback). Typing speeds in both the audio and tactile version were faster than in the visual version when at home, at work or commuting.

Regarding KSPC, there were higher levels of KSPC when using the tactile and visual versions than the audio when commuting. This could imply that the audio feedback was not noticeable enough in these locations for participants to recognise errors and correct them. In bars/restaurants there were higher levels of KSPC in the audio version compared to the tactile version. When at home and at work, both the audio and tactile modalities achieved KSPC levels close to 1.0 which is the ideal number of keystrokes per character.

A user's personal preference should also be taken into account when choosing the most appropriate modality for different locations. The CrossTrainer study showed that when participants were at home, work, and at a bar/ restaurant, the tactile modality was chosen significantly more often than the audio modality.

Participants stated that using CrossTrainer while commuting (on a bus, train or walking) was difficult because of the surrounding environmental sound and vibration levels whereas when using it at work or in a bar/restaurant surrounded by other people, it was socially inappropriate to use the audio version without headphones.

Environmental factors not only affect the perception of audio and tactile feedback but also user preference. The CrossTrainer results suggest that audio feedback becomes the preferred feedback modality at vibration levels of 8.1 g/s and above. Tactile feedback is the preferred modality at vibration levels of 0 - 8 g/s. For noise levels, tactile feedback is the preferred modality for 0 – 70 dB and 91+ dB. Interestingly, when noise levels are between 71 and 90 dB it appears as though both audio and tactile feedback result in similar preference levels.

Overall, when given a choice of modalities, participants chose tactile for 82% of the time and audio 18% of the time. The visual version received no votes. There are evidently times when audio is more appropriate than tactile and *vice versa*. As a result of this experiment, it is possible for devices to support crossmodal tactile and audio feedback to cover the widest range of environments, preferences, locations and tasks.

8.6 Guidelines

In addition to answering the four research questions posed in the introduction, another significant contribution of this thesis is the production of the first set of guidelines to aid designers who wish to use crossmodal audio and tactile icons in touchscreen interfaces. Guidelines have been extracted from the results of the experiments of every chapter and these are listed again below. The relevant chapters should be consulted for more detail on each of these guidelines.

Guidelines for Creating Parameters for Crossmodal Icons (from Chapter 3 and Chapter 4)

1. Rhythm, roughness and spatial location can be used to effectively encode data in crossmodal icons.
2. Rhythm: to make use of rhythm as a parameter in crossmodal icons, identical rhythms should be created in audio and tactile.
    a. Chords should not be used in the audio rhythms unless there are two or more tactile actuators available.
    b. If distinguishable levels are needed, use up to 4 different rhythms each with a different number of notes or pulses:
        i. The rhythms used in this research varied in the number of notes from 1 to 6 (with a maximum duration of 1 second and a minimum of 300 milliseconds).
3. Roughness: timbre in the audio modality should be mapped to waveform in the tactile modality.
    a. If distinguishable levels of texture are required, use up to three levels of roughness:
        i. Tactile: sine wave, sawtooth wave and square wave.
        ii. Audio: piano, tremolo trumpet and vibraphone.
4. Spatial Location: use a 3D audio soundscape with sounds placed at cardinal points (ideally presented through headphones) and use tactile actuators placed on the waist at cardinal points.
    a. If not mobile, stereo speakers may be used instead of headphones.
    b. If actuators are attached directly to the device, arrange them in cardinal positions or positions that can be easily recreated in a soundscape such as a 2.5D planar soundscape.

Guidelines on Training Users (from Chapter 5)

1. If a user is taught to understand alerts provided by crossmodal tactons, he/she could be expected to understand crossmodal earcons with no audio training.

2. If a user is taught to understand alerts provided by crossmodal earcons, he/she could be expected to understand crossmodal tactons with no tactile training.

3. On average, it takes 2 training sessions for participants to be able to identify three-dimensional earcons with recognition rates of 90% or higher.

4. On average, it takes 3 training sessions for participants to identify three-dimensional tactons with recognition rates of 90% or above.

Guidelines for When to Use Audio and When to Use Tactile Feedback (from Chapters 6 and 7)

Environmental Levels

1. Audio feedback should be used when there are environmental vibration levels of 9.1 g/s and above.
2. Tactile feedback should be used during noise levels of 94 dB and above

Modality Preference

1. In terms of user preference, the tactile modality was chosen 82% of the time during the CrossTrainer study. Thus, tactile feedback should be the default setting, as most people preferred it.

Social Situation

1. When in the company of others it can be embarrassing to use audio feedback on a mobile device and it may be considered rude to wear headphones so tactile feedback should be used instead.

Location

2. Bar/restaurant: both tactile and audio feedback can produce high levels of performance in this location. In terms of typing speeds, using tactile feedback results in higher typing speeds. However, audio produces higher KSPC meaning that users correct more errors with the audio modality. Regardless of this, the study of personal modality preference indicates that tactile is the preferred modality in this location.
3. Home: both audio and tactile feedback produce high typing speeds compared to typing on a keyboard with no crossmodal feedback. Furthermore, both modalities produce high accuracy rates and almost ideal KSPC rates. Most importantly, the study of personal modality preference indicates that tactile is the preferred modality in this location.
4. Work: both audio and tactile feedback produce high typing speeds compared to typing on a keyboard with no crossmodal feedback. Furthermore, both modalities produce

high accuracy rates and almost ideal KSPC rates. Again, the study of personal modality preference indicates that tactile is the preferred modality in this location.

5. Commuting: both audio and tactile feedback produce high typing speeds compared to typing on a keyboard with no crossmodal feedback. However, there are higher rates of KSPC in the tactile modality indicating that the audio feedback may go unnoticed and therefore errors are not corrected. Therefore, tactile feedback should be chosen in this location.

8.6.1 Guideline Limitations/Caveats

When applying the guidelines above, the following issues should be considered:

Attention Shifting

The time taken to shift attention from audio to tactile or *vice versa* should also be taken into account and perhaps, when applications automatically switch to the most appropriate modality, the swap should take place in between tasks or when there is enough time for the user's attention shift to occur or perhaps faded in as the user approaches a level where one modality will become ineffective.

Extreme Environmental Variables

It is often the case that in situations with high vibration levels, there will be high noise levels too. In these circumstances the effectiveness of both audio and tactile feedback will significantly decrease resulting in levels of performance similar to those achieved with visual feedback only.

Device Location versus User Location

Although the experiment results and guidelines show that the location of the user is important when choosing an appropriate modality, it must be noted in all cases that the location of the device is also important. If the device is not in contact with the user's skin, tactile feedback may not be detected.

8.7 Future Work

Although a substantial volume of work was completed during the course of this thesis, there are opportunities for further research in this field to overcome other problems and limitations in the future. These include:

Creating Crossmodal Icons With Different Hardware

New types of tactile technologies are becoming available all the time and it will be necessary to investigate whether the results achieved in this research can be recreated using different types of actuators.

Commercially available mobile devices are now equipped with stereo speakers. These improved capabilities are important because this may make it possible for 3D audio to be used without headphones on mobile devices. However, it may not be possible to deal with the crosstalk in a real world environment. Further studies could investigate the effectiveness of such feedback in mobile situations and the number of locations in a soundscape could be extended.

Longer-Term Studies

The study of CrossTrainer generated a rich set of data with 72 days of use overall. However, each participant only used each modality for a maximum of 4 days and a minimum of 2. It would be beneficial to conduct a longitudinal study over a much longer period to determine whether users' modality preferences change over a longer period of time and capture data in a wider variety of situations and locations.

Training and Learning

It would be beneficial to see how different types of training affect performance with crossmodal earcons and tactons. In the study of 3D crossmodal icons, users were trained for 10 minutes and given feedback on their progress. Further studies will look at the effectiveness of explicit versus implicit learning in crossmodal interaction to reduce the amount of training time needed.

Another interesting possibility would be to create a much larger set of crossmodal icons to investigate the thresholds of learning to find out the maximum number of icons that can be learned and the length of time this learning takes.

More Modalities and Parameters

Crossmodal icons make use of the amodal attributes available in the senses of touch and hearing. There are other amodal attributes that are present in the visual modality too, for example, texture. It could be beneficial to include visual crossmodal icons as well as audio and tactile to increase the number of modality choices.

There are also other amodal attributes available in audio and tactile that were not investigated in this thesis. These include duration and intensity. Further experiments like those in Chapter 4 could be conducted to establish whether these attributes can also be used as successful crossmodal parameters where users can match cues in both modalities when data is encoded in duration or intensity.

Other Situational Impairments

The results of Chapter 6 show that tactile feedback should be used when noise levels in the surrounding environment are >94 dB and that audio feedback should be used when vibration levels reach > 9.1 g/s. However, there may be times when the user is not subjected to any environmental vibrations therefore indicating that tactile feedback would be most appropriate but the device is actually placed in a bag or pocket and is not in direct contact with the skin. In these situations, audio feedback would be most appropriate. The next steps in further studies would be to examine these exceptions to the rules by making use of other device sensors such as GPS for location, and also take user preference into consideration.

Adaptive Crossmodal Touchscreen Interfaces

The most interesting potential future work that has emerged from this research is the development of an adaptive interface using crossmodal feedback. Using the results from the experiments in Chapters 6 and 7, a set of rules have been established as to when each modality should be used. Therefore a completely crossmodal interface can be created where the feedback modality automatically adapts given the user's situation, location and preference.

8.8 Conclusions

This thesis has investigated the use of crossmodal audio and tactile interaction with mobile touchscreens. This thesis has provided the first detailed experimental investigations into the design of crossmodal icons. Two icons may be considered to be crossmodal icons if and only if they provide a common representation of data, which is accessible interchangeably via different modalities. This is the first time that this approach has been applied to the design of audio and tactile icons. The results from this thesis research therefore provide a benchmark against which the results of future research on crossmodal audio and tactile icons, or icons using other modalities, can be measured.

While a range of studies on audio and tactile icons exists in the domain of human computer interaction, there has been little work on exploiting the similarities between the modalities. This thesis addresses the following question: how does one accomplish a translation across the modalities, namely acquire an audio understanding of something that has been touched and *vice versa*? The results of this research have shown that it is possible to use the amodal attributes available in both modalities to encode data in such a way that stimuli presented by each modality are perceived as synonymous. Furthermore, by designing multidimensional crossmodal icons using these parameters, users need only be trained to understand mappings in one modality because they can transfer their knowledge to the other. Thus reducing workload for users and allowing them to switch between modalities easily.

This carefully designed crossmodal feedback was added to a touchscreen QWERTY keyboard and to CrossTrainer, a mobile touchscreen IQ game with crossmodal audio and tactile feedback, to demonstrate that crossmodal icons can benefit the interactions with these interfaces. Studying the use of crossmodal icons in real-world applications enabled the measurement of surrounding environmental noise and vibration levels as well as personal modality preference. These results provide information as to when feedback should switch from audio to tactile and *vice versa*. This is the primary purpose of crossmodal icons. Furthermore, a set of guidelines has been produced from the results of the empirical experiments reported in this thesis to aid other researchers or interface designers to create crossmodal icons for their own use.

This thesis has successfully shown that using crossmodal auditory and tactile interaction can aid mobile touchscreen users in accessing information and feedback non-visually and, by providing a choice of modalities, can help to overcome problems that occur in different mobile situations where one modality may be more suitable than another. By encoding information using the crossmodal parameters of audio and vibration, users can learn mappings and translate information between both modalities. Therefore, information may be presented to the most appropriate modality given the situation, a significant improvement for future mobile interfaces.

Appendices

To save paper, the following appendices can be found online at http://www.dcs.gla.ac.uk/~eve/ThesisAppendices/

- Appendix A: Experiment 1a files
- Appendix B: Experiment 1b files
- Appendix C: Experiment 2 files
- Appendix D: Experiment 2b files
- Appendix E: Experiment 3 files
- Appendix F: Experiment 4 files
- Appendix G: Experiment 5 files
- Appendix H: Experiment 6 files
- Appendix I: Example crossmodal stimuli

References

[1] ISO: Ergonomics of Human-Computer Interaction — Part 910: Framework for Tactile and Haptic Interaction, (2009).

[2] "The Penguin Dictionary of Psychology," A. S. Reber, Ed.: Penguin.

[3] Akamatsu, M., Mackenzie, I. S., and Hasbrouq, T. A Comparison of Tactile, Auditory, and Visual Feedback in a Pointing Task Using a Mouse-Type Device. *Ergonomics* 38 (1995), 816 - 827.

[4] Association, A. S., "Acoustical Terminology," American Standards Association, 1960.

[5] Barnard, L., Yi, J. S., Jacko, J. A., and Sears, A. An Empirical Comparison of Use-in-Motion Evaluation Scenarios for Mobile Computing Devices. *International Journal of Human-Computer Studies* 62 (2005), 487 - 520.

[6] Begault, D. R. *3-D Sounds for Virtual Reality and Multimedia.* Academic Press, Boston, 1994.

[7] Bernsen, N. O. Modality Theory in Support of Multimodal Interface Design. *ERCIM Workshop on Multimodal Human-Computer Interaction* (1993), 27 - 44.

[8] Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. Earcons and Icons: Their Structure and Common Design Principles. *Human Computer Interaction* 4(1) (1989), 11 - 44.

[9] Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Chechosky, C. M. Four Channels Mediate the Mechanical Aspects of Touch. *Journal of the Acoustical Society of America* 84(5) (1988), 1680 - 1694.

[10] Brewster, S., Chohan, F., and Brown, L. Tactile Feedback for Mobile Interactions. In *Proc. CHI '07,* ACM Press (2007), 159 - 162.

[11] Brewster, S. A. Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing* 6 (2002), 188 - 205.

[12] Brewster, S. A. Sound in the Interface to a Mobile Computer. In *Proc. HCI International '99,* Lawrence Erlbaum Associates (1999), 43 - 47.

[13] Brewster, S. A. Using non-speech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction* 5(2) (1998), 224 - 259.

[14] Brewster, S. A. and Brown, L. M. Tactons: Structured Tactile Messages for Non-Visual Information Display. In *Proc. AUI Conference 2004,* ACS (2004), 15 - 23.

[15] Brewster, S. A., Wall, S., Brown, L. M., and Hoggan, E., "Tactile Displays," in *The Engineering Handbook on Smart Technology for Aging, Disability and Independence*. Eds. A. Helal, M. Mokhtari, and B. Abdulrazak. John Wiley & Sons, 2008.

[16] Brewster, S. A., Wright, P. C., Dix, A., and Edwards, A. D. N. The Sonic Enhancement of Graphical Buttons. In *Proc. Interact'95,* Chapman and Hall (1995), 43 - 48.

[17] Brewster, S. A., Wright, P. C., and Edwards, A. D. N. An evaluation of earcons for use in auditory human-computer interfaces. In *Proc. InterCHI 1993,* ACM Press (1993), 222 - 227.

[18] Brewster, S. A., Wright, P. C., and Edwards, A. D. N. Experimentally Derived Guidelines for the Creation of earcons. In *Proc. HCI '95,* Springer (1995), 155 - 159.

[19] Broadbent, D. E. *Perception and Communication.* Pergamon Press, New York, 1958.

[20] Brown, L. M. and Brewster, S. A. Multidimensional tactons for Non-Visual Information Display in Mobile Devices. In *Proc. MobileHCI '06,* ACM Press (2006), 231 - 238.

[21] Brown, L. M., Brewster, S. A., and Purchase, H. C. A First Investigation into the Effectiveness of tactons. In *Proc. WorldHaptics 2005,* IEEE (2005), 167 - 176.

[22] Brown, L. M., Brewster, S. A., and Purchase, H. C. Tactile crescendos and sforzandos: applying musical techniques to tactile icon design. In *Proc. Vol II CHI 2006,* ACM Press (2006), 610 - 615.

[23] Bruce, I. A. and McKennell, A. *Blind and Partially Sighted Adults in Britain: The RNIB Survey.* HMSO, London, 1991.

[24] Buttler, A. J. and Oravainen, K. S. The Effect of Rhythm on the Memorability of Auditive and Vibrotactile Patterns. *(unpublished) Perceptual and Cognitive Psychology: Course Assignment Report, University of Helsinki, Finland* (2002).

[25] Chan, A., MacLean, K., and McGrenere, J. Learning and Identifying Haptic Icons Under Workload. In *Proc. The First Joint Eurohaptics Conference and Symposium of Haptic Interfaces for Virtual Environment and Teleoperator Systems,* IEEE (2005), 432 - 439.

[26] Chandler, D. *Semiotics For Beginners.* Routledge Taylor and Francis Group, 1994.

[27] Chang, A., O'Modhrain, S., Jacob, R., Gunther, E., and Ishii, H. ComTouch: Design of a Vibrotactile Communication Device. In *Proc. Designing Interactive Systems,* ACM Press (2002), 312 - 320.

[28] Chang, A. and O'Sullivan, C. Audio-haptic feedback in mobile phones. In *Proc. CHI '05 extended abstracts on Human factors in computing systems,* ACM (2005).

[29] Cholewiak, R. W., Brill, J. C., and Schwab, A. Vibrotactile Localization on the Abdomen: Effects of Place and Space. *Perception and Psychophysics* 66 (2004), 970 - 987.

[30] Cholewiak, R. W. and Collins, A. A. Vibrotactile Localization on the arm: Effects of Place, Space and age. *Perception and Psychophysics* 65 (2003), 1058 - 1077.

[31] Cholewiak, R. W. and Collins, A. A. Vibrotactile Pattern Discrimination And Communality at Several Body Sites. *Perception and Psychophysics* 57(5) (1995), 724 - 737.

[32] Cholewiak, R. W., Collins, A. A., and Brill, J. C. Spatial Factors in Vibrotactile Pattern Perception In *Proc. EuroHaptics,* (2001).

[33] Cholewiak, R. W. and Craig, J. C. Vibrotactile Pattern Recognition and Discrimination at Several Body Sites. *Perception and Psychophysics* 35 (1984), 503 - 514.

[34] Cockburn, A. and Brewster, S. A. Multimodal feedback for the acquisition of small targets. *Ergonomics* 48(9) (2005), 1129 - 1150.

[35] Craig, J. C. and Johnson, K. O. The Two-Point Threshold: Not a Measure of Tactile Spatial Resolution. *Current Directions in Psychological Science* 9(1) (2000), 29 - 32.

[36] Craig, J. C. and Sherrick, C. E., "Dynamic Tactile Displays," in *Tactual Perception: A Sourcebook*. Eds. W. Schiff and E. Foulke. Cambridge: Cambridge University Press, 1982, 209 - 233.

[37] Crossan, A., Murray-Smith, R., Brewster, S. A., and Musizza, B., "Instrumented Usability Analysis for Mobile Devices," in *Handbook of Mobile HCI*. 2008.

[38] Delmont, J. The Tactile Acoustic Monitor as an Aid for Postlingually Totally Defeaned Adults: A Pilot Study. *British Journal of Audiology* 22 (1988), 313 - 316.

[39] Deutsch, D., "Grouping Mechanisms in Music," in *The Psychology of Music*. Ed. D. Deutsch. San Diego: Academic Press, 1999, 299 - 348.

[40] Driver, J. and Spence, C. Multisensory Perception: Beyond Modularity and Convergence. *Current Biology* 10 (2000), 731 - 735.

[41] Enriquez, M. and MacLean, K. The Role of Choice in Longitudinal Recall of Meaningful Tactile Signals. In *Proc. IEEE Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems,* IEEE (2008), 49 - 56.

[42] Enriquez, M., MacLean, K., and Chita, C. Haptic Phenomes: Basic Building Blocks of Haptic Communication. In *Proc. 8th International Conference on Multimodal Interfaces (ICMI'06),* IEEE (2006), 302 - 309.

[43] Fastl, H. Roughness and Temporal Masking Patterns of Sinusoidally Amplitude Modulated Broadband Noise. *Psychophysics and Physiology of Hearing* (1977), 403 - 414.

[44] Fukumoto, M. and Sugimura, T. Active Click: Tactile Feedback for Touch Panels. In *Proc. CHI' 01,* ACM Press (2001), 121 - 122.

[45] Gardner, W. G. (1999). *3D Audio and Acoustic Environment Modelling*. Report. Wave Arts Inc.

[46] Gaver, W. Auditory Icons: Using Sound in Computer Interfaces. *ACM SIGCHI Bulletin* 19(1) (1987), 74.

[47] Gaver, W. The SonicFinder: An Interface that Uses Auditory Icons *Human Computer Interaction* 4(1) (1989), 67 - 94.

[48] Geldard, F. A. Adventures in Tactile Literacy. *The American Psychologist* 12 (1957), 115 - 124.

[49] Geldard, F. A. Some Neglected Possibilities of Communication. *Science* 131(3413) (1960), 1583 - 1588.

[50] Gemperle, F., Ota, N., and Siewiorek, D. Design of a Wearable Tactile Display. In *Proc. International Symposium on Wearable Computers,* (2001).

[51] Gibson, E. J. Principles of Perceptual Learning and Development. *Annual Review Psychology* 14 (1969), 29 - 56.

[52] Gibson, J. J. *The Senses Considered as Perceptual Systems.* Houghton Mifflin, Boston, 1966.

[53] Goff, G. D. Differential Discrimination of Frequency of Cutaneous Mechanical Vibration. *Journal of Experimental Psychology* 74 (1967), 294 - 299.

[54] Gunther, E., Davenport, G., and O'Modhrain, S. Cutaneous Grooves: Composing for the Sense of Touch. In *Proc. New Interfaces for Musical Expression,* ACM Press (2002), 1 - 6.

[55] Hall, M., Hoggan, E., and Brewster, S. A. T-Bars: Towards Tactile User Interfaces for Touchscreen Mobiles. In *Proc. MobileHCI' 08,* ACM Press (2008).

[56] Hameed, S., Ferris, T., Jayaraman, S., and Sarter, N. B. Supporting Interruption Management Through Informative Tactile and Peripheral Visual Cues. In *Proc. Human Factors and Ergonomics Society 50th Annual Meeting,* Human Factors and Ergonomics Society (2006), 376 - 380.

[57] Hart, S. G. and Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload* (1988), 139 - 183.

[58] Heller, M. A., "Haptic Perception in Blind People" in *The Psychology of Touch*. Eds. M. A. Heller and W. Schiff. Hillsdale, NJ: Lawrence Erlbaum, 1991, 239 - 261.

[59] Hoggan, E., Brewster, S., and Anwar, S. Mobile Multi-Actuator Tactile Displays. In *Proc. 2nd Intl Workshop on Haptic and Audio Interaction Design,* Springer LNCS (2007), 22 - 33.

[60] Hoggan, E. and Brewster, S. A. Crossmodal Icons for Information Display. In *Proc. CHI '06 Extended Abstracts,* ACM Press (2006), 857 - 862.

[61] Hoggan, E. and Brewster, S. A. Crossmodal Spatial Location: Initial Experiments. In *Proc. NordiCHI '06,* ACM Press (2006), 469 - 472.

[62] Hoggan, E. and Brewster, S. A. CrossTrainer: Testing The Use of Multimodal Interfaces in Situ. In *Proc. CHI'10,* ACM Press (2010).

[63] Hoggan, E. and Brewster, S. A. Designing audio and tactile crossmodal icons for mobile devices. In *Proc. Proceedings of the 9th international conference on Multimodal interfaces,* ACM (2007), 162 - 169.

[64] Hoggan, E. and Brewster, S. A. New Parameters for tacton Design. In *Proc. Vol II CHI 2007,* ACM Press (2007), 2417 - 2422.

[65] Hoggan, E., Brewster, S. A., and Johnston, J. Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In *Proc. CHI' 08,* ACM Press (2008), 1573 - 1582.

[66] Hoggan, E., Crossan, A., Brewster, S. A., and Kaaresoja, T. Audio or Tactile Feedback: Which Modality When? In *Proc. CHI'09,* ACM Press (2009), 2253 - 2256.

[67] Hsu, W. Using Timbre in a Computer-Based Improvisation System. In *Proc. ICMC,* ICMA (2005).

[68] Jacko, J. A., Scott, I., Sainfort, F., Barnard, L., Edwards, P., Emery, V. K., Kongnakorn, T., Moloney, K., and Zorich, B. Older Adults and Visual Impairment: What Do Exposure Times and Accuracy Tell Us About Performance Gains Associated with Multimodal Feedback? . In *Proc. ACM CHI 2003,* ACM Press (2003).

[69] Johnson, K. O. and Phillips, J. R. Tactile Spatial Resolution: I. Two-point Discrimination, Gap Detection, Grating Resolution, and Letter Recognition. *Journal of Neurophysiology* 46 (1981), 1177 - 1191.

[70] Jokiniemi, M., Raisamo, R., Lylykangas, J., and Surakka, V. Crossmodal Rhythm Perception. In *Proc. 3rd International Workshop on Haptic and Audio Interaction Design* Springer (2008), 111 - 119.

[71] Jones, M., Jones, S., Bradley, G., and Holmes, G. Navigation by Music: an Initial Prototype and Evaluation. In *Proc. International Symposium on Intelligent Environments,* Microsoft Research (2006), 95 - 102.

[72] Kaaresoja, T., Brown, L. M., and Linjama, J. Snap-Crackle-Pop: Tactile Feedback for Mobile Touch Screens. In *Proc. Eurohaptics '06,* (2006), 565 - 566.

[73] Kaczmarek, K. A. and Bach-Y-Rita, P. *Tactile Displays.* Oxford University Press, New York, 1995.

[74] Kandel, E. R. and Jessell, T. M., "Touch," in *Principles of Neural Science*. Eds. E. R. Kandel, J. H. Schwartz, and T. M. Jessell. New York: Oxford University Press, 1991, Ch pp. 349 - 414.

[75] Karam, M., Russo, F., Branje, C., Price, E., and Fels, D. I. Towards a Model Human Cochlea: Sensory Substitution for Crossmodal Audio-Tactile Displays. In *Proc. Graphics Interface,* ACM Press (2008), 267 - 274.

[76] Kibens, V. Discrete Noise Spectrum Generated by Acoustically Excited Jet. *AIAA* 18(4) (1980), 434 - 441.

[77] Kjeldskov, J. and Stage, J. New Techniques for Usability Evaluation of Mobile Systems *International Journal of Human-Computer Studies* 60 (2004), 599 - 620.

[78] Koskinen, E., Kaaresoja, T., and Laitinen, P. Feel-Good Touch: Finding the Most Pleasant Tactile Feedback for a Mobile Touch Screen Button. In *Proc. ICMI 2008,* ACM Press (2008), 297 - 304.

[79] Laitinen, P. and Mäenpää, J. Enabling Mobile Haptic Design: Piezoelectric Actuator Technology Properties in Handheld Devices. In *Proc. HAVE'06,* (2006).

[80] Lederman, S. J. and Abbott, S. G. Texture Perception: Studies of Intersensory Organization Using a Discrepancy Paradigm, and Visual Versus Tactual Psychophysics. *Journal of Experimental Psychology* 7(4) (1981), 902 - 915.

[81] Lederman, S. J. and Klatzky, R. L. New Directions in Touch. *Canadian Journal of Experimental Psychology* 61 (2007), 169 - 170.

[82] Lee, J. C., Dietz, P. H., Leigh, D., Yerazunis, W. S., and Hudson, S. E. Haptic pen: a tactile feedback stylus for touch screens In *Proc. 17th annual ACM symposium on User interface software and technology* ACM Press (2004), 291-294.

[83] Lee, S. and Zhai, S. The Performance of Touch Screen Soft Buttons. In *Proc. CHI'09,* ACM Press (2009), 309 - 318.

[84] Lenay, C., Canu, S., and Villon, P. Technology and Perception: The Contribution of Sensory Substitution Systems. In *Proc. ICCT,* IEEE (1997).

[85] Leplatre, G. and Brewster, S. A. Designing non-speech sounds to support navigation in mobile phone menus. In *Proc. ICAD 2001,* (2001), 190 - 199.

[86] Lewkowicz, D. J., "Development of Intersensory Perception in Human Infants," in *The Development of Intersensory Perception: Comparative Perspectives*. Eds. D. J. Lewkowicz and R. Lickliter. Hillsdale, N.J.: Erlbaum, 1994, 165 - 203.

[87] Lewkowicz, D. J. The Development of Intersensory Temporal Perception: An Epigenetic Systems/Limitations View. *Psychological Bulletin* 126 (2000), 281 - 308.

[88] Loomis, J. M. and Lederman, S. J., "Tactual Perception," in *Handbook of Perception and Human Performance*. vol. 2. Eds. K. Boff, L. Kaufman, and J. P. Thomas. Oxford, England: John Wiley and Sons, 1986, 1143.

[89] Luk, J., Pasquero, J., Little, S., MacLean, K., Levesque, V., and Hayward, V. A role for haptics in mobile interaction: initial design using a handheld tactile display prototype In *Proc. CHI '06,* ACM Press (2006), 171-180.

[90] MacKenzie, I. S. and Soukoreff, R. W. Phrase Sets for Evaluating Text Entry Techniques. In *Proc. CHI '03,* ACM (2003), 754 - 755.

[91] Mackenzie, I. S., Zhang, S., and Soukoreff, R. W. Text Entry Using Soft Keyboards. *Behaviour and Information Technology* 18 (1999), 235 - 244.

[92] MacLean, K. and Enriquez, M. Perceptual Design of Haptic Icons. In *Proc. Eurohaptics,* (2003), 351 - 363.

[93] Marks, L. E. *The Unity of Senses. Interrelations Among The Modalities.* Academic Press, New York, 1978.

[94] McGookin, D. The Presentation of Multiple earcons in a Spatialised Audio Space. In *Proc. British HCI 2002,* BCS (2002), 228 - 229.

[95] McGookin, D. and Brewster, S. A. Space, The Final Frontearcon: The Identification of Concurrently Presented earcons in a Synthetic Spatialised Auditory Environment. In *Proc. ICAD 2004,* ICAD (2004), 1 - 8.

[96] McGookin, D. and Brewster, S. A. Understanding Concurrent earcons: Applying Auditory Scene Analysis Principles to Concurrent earcon Recognition. *ACM Transactions on Applied Perception* 1 (2004), 130 - 155.

[97] Mendelson, M. J. Acoustic-Optical Correspondences and Auditory-Visual Coordination in Infancy. *Canadian Journal of Experimental Psychology* 33 (1979), 334 - 346.

[98] Mereu, S. and Kazman, R. Audio Enhanced 3D Interfaces for Visually Impaired Users. In *Proc. CHI 1996,* ACM Press (1996), 72 - 78.

[99] Miller, G. A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review* 63 (1956), 81 - 97.

[100] Moore, B. C. J. *An Introduction to the Psychology of Hearing.* Academic Press, Boston, 2004.

[101] Mursell, J. L. *The Psychology of Music.* W.W. Norton and Company, Inc., New York, 1937.

[102] Namba, S. and Kuwano, S., "The Loudness of Non-Steady State Sounds: Is a Ratio Scale Applicable?," in *Ratio Scaling of Psychological Magnitude*. Eds. S. J. Bolanowski, G. A. Gescheider, and S. S. Stevens. Lawrence Erlbaum, 1991, 229 - 245.

[103] Nashel, A. and Razzaque, S. Tactile Virtual Buttons for Mobile Devices. In *Proc. CHI '03 extended abstracts,* ACM Press (2003), 854 - 855.

[104] Neisser, U. Gibson and the Pyshcology of Perception: Two Reviews: Gibson's Revolution. *PsycCRITIQUES* 35(8) (1990), 749 - 750.

[105] Parhi, P., Karlson, A. K., and Bederson, B. B. Target Size Study for One-Handed Thumb Use on Small Touchscreen Devices. In *Proc. Mobile HCI '06,* ACM Press (2006), 203 - 210.

[106] Parncutt, R. Harmony: A Psychoacoustical Approach. *Springer Series in Information Sciences* 19 (1989).

[107] Phillips, J. R. and Johnson, K. O. Neural Mechanisms of Scanned and Stationary Touch. *Journal of the Acoustical Society of America* 77 (1985), 220 - 224.

[108] Pierce, A. D. *Acoustics: An Introduction to Its Principles and Applications.* Acoustical Soc of America, 1989.

[109] Pirhonen, A., Brewster, S., and Holguin, C. Gestural and Audio Metaphors as a Means of Control for Mobile Devices. In *Proc. CHI '02,* ACM Press (2002), 291 - 298.

[110] Plomp, R., "Timbre as a Multidimensional Attribute of Complex Tones," in *Frequency Analysis and Periodicity Detection in Hearing*. Eds. R. Plomp and G. F. Smoorenburg. Sijthoff: Leiden, 1970.

[111] Poupyrev, I. and Maruyama, S. Tactile Interfaces for Small Touch Screens. In *Proc. 16th Annual ACM Symposium on User Interface Software and Technology,* ACM Press (2003), 217 - 220.

[112] Poupyrev, I., Maruyama, S., and Rekimoto, J. Ambient touch: designing tactile interfaces for handheld devices. In *Proc. 15th annual ACM symposium on User interface software and technology,* ACM Press (2002), 51 - 60.

[113] Poupyrev, I., Rekimoto, J., and Maruyama, S. TouchEngine: A Tactile Display for Handheld Devices. In *Proc. ACM CHI 2002,* ACM Press (2002), 644 - 645.

[114] Rabinowitz, W. M., Houtsma, A. J. M., Durlach, N. I., and Delhorne, L. A. Multidimensional Tactile Displays: Identification of Vibratory Intensity, Frequency, and Contactor Area. *Journal of the Acoustical Society of America* 82(4) (1987), 1243 - 1252.

[115] Raento, M., Oulasvirta, A., Petit, R., and Toivonen, H. ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. *IEEE Pervasive Computing* 4(2) (2005), 51-59.

[116] Reales, J. M. and Ballesteros, S. Implicit and Explicit Memory for Visual and Haptic Objects: Cross-modal Priming Depends on Structural Descriptions. *Journal of Experimental Psychology* 25 (1999), 644 - 663.

[117] Reeves, L. M., KLai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Kraal, B., Martin, J. C., McTear, M., Raman, T. V., Stanney, K. M., Su, H., and Wang, Q. Y. Guidelines for Multimodal User Interface Design. *Commun. ACM* 47(1) (2004), 57 - 59.

[118] Rigas, D. I., "Guidelines for Auditory Interface Design: An Empirical Investigation." vol. PhD: Loughborough University, 1996.

[119] Rossing, T. D., Moore, F. R., and Wheeler, P. A. *The Science of Sound.* Addison wesley, San Francisco, 2002.

[120] Rovan, B. and Hayward, V., "Typology of Tactile Sounds and Their Synthesis in Gesture Driven Computer Music Performance," in *Trends in Gestural Control of Music*. Eds. M. Wanderley and M. Battier. Paris: IRCAM, 2000, 297 - 320.

[121] Rovers, A. F. and van Essen, H. A. Using Haptic Feedback in Everyday Products. In *Proc. EuroHaptics,* Springer-Verlag (2006), 447 - 453.

[122] Saito, D. N., Okada, T., Morita, Y., Yonekura, Y., and Sadato, N. Tactile–visual cross-modal shape matching: a functional MRI study. *Cognitive Brain Research* 17 (2003), 14 - 25.

[123] Sawhney, N. and Schmandt, C. Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. In *Proc. ACM Transactions on Computer Human Interaction,* ACM Press (2000), 353 - 383.

[124] Schiff, W. and Foulke, E. *Tactual Perception: A Sourcebook.* Cambridge University Press, Cambridge, 1982.

[125] Sharmin, S., Evreinov, G., and Raisamo, R. Non-Visual Feedback Cues for Pen Comuputing. In *Proc. Eurohaptics,* IEEE Computer Society (2005), 625 - 628.

[126] Sherrick, C. E. A Scale for Rate of Tactual Vibration. *Journal of the Acoustical Society of America* 78(1) (1985), 78 - 83.

[127] Sherrick, C. E. and Cholewiak, R. W., "Cutaneous Sensitivity," in *Handbook of Perception and Human Performance*. Eds. K. Boff, L. Kaufman, and J. L. Thomas. New York: Wiley and Sons, 1986, Ch 12 pp. 1 - 58.

[128] Sherrick, C. E., Cholewiak, R. W., and Collins, A. A. The Localization of Low- and High-Frequency Vibrotactile Stimuli. *Journal of the Acoustical Society of America* 88(1) (1990), 169 - 179.

[129] Shneiderman, B. *Designing the User Interface.* Addison-Wesley, Reading, MA, 1998.

[130] Shneiderman, B. Touch Screens Now Offer Compelling Uses. *IEEE Software* 8 (1991), 93 - 107.

[131] Soukoreff, R. W. and MacKenzie, I. S. Metrics for Text Entry Research: an Evaluation of MSD and KSPC, and a new Unified Error Metric. In *Proc. CHI '03,* ACM (2003), 113 - 120.

[132] Spence, C. Multisensory Attention and Tactile Information-Processing. *Behavioural Brain Research* 135(1) (2002), 57 - 64.

[133] Sumikawa, D. A., Blattner, M. M., Joy, K., and Greenberg, R. M. (1986). *Guidelines for the syntactic design of audio cues in computer interfaces*. (Technical Report No. UCRL 92925).Technical Report. Lawrence Livermore National Laboratory.

[134] Summers, I. R., "Single-Channel Information Transfer Through the Skin: Limitations and Possibilities," in *International Sensory Aids Conference*. Eds. I. R. Summers and J. Whybrow. Exeter, UK, 2000.

[135] Summers, I. R. *Tactile Aids for the Hearing Impaired.* Whurr Publishers Ltd., London, UK, 1992.

[136] Summers, I. R., Peake, M. A., and Martin, M. C. Field Trials of a Tactile Acoustic Monitor for the Profoundly Deaf. *British Journal of Audiology* 15(3) (1981), 195 - 199.

[137] Tamminen, S., Oulasvirta, A., Toiskallio, K., and Kankainen, A. Understanding Mobile Contexts. *Personal and Ubiquitous Computing* 8(2) (2004), 135 - 143.

[138] Tan, H. Z., Durlach, N. I., Reed, C. M., and Rabinowitz, W. M. Information Transmission with a Multifinger Tactual Display. *Perception and Psychophysics* 61 (1999), 993 - 1008.

[139] Terhardt, E. On the Perception of Periodic Sound Fluctuations (Roughness). *Acustica* 30 (1974), 201 - 213.

[140] Ternes, D. and MacLean, K. Designing Large Sets of Haptic Icons with Rhythm. In *Proc. EuroHaptics,* (2008), 199 - 208.

[141] van Erp, J. B. Guidelines for the use of Vibro-Tactile Displays in Human Computer Interaction. In *Proc. EuroHaptics,* Springer-Verlag (2002), 18 - 22.

[142] van Erp, J. B., "Tactile Displays For Navigation and Orientation: Perception and Behaviour," Computing Science PhD: Utrecht University, 2007.

[143] van Erp, J. B. Tactile Navigation Display. In *Proc. First international Workshop on Haptic Human - Computer Interaction,* LNCS (2001), 165 - 173.

[144] van Erp, J. B. and Spape, M. M. A. Distilling the Underlying Dimensions of Tactile Melodies. In *Proc. Eurohaptics,* (2003), 111 - 120.

[145] van Erp, J. B. and van den Dobbelsteen, J. J. (1998). *On The Design of Tactile Displays*. Report. TNO Human Factors Research Institute.

[146] van Erp, J. B., van Veen, H. A. H. C., Jansen, C., and Dobbins, T. Waypoint Navigation with a Vibrotactile Waist Belt. *ACM Transactions on Applied Perception* 2(2) (2005), 106 - 117.

[147] van Erp, J. B. and Verschoor, M. H. Cross-modal visual and vibrotactile tracking. *Applied Ergonomics* 35(2) (2004), 105 - 112.

[148] Verrillo, R. T. Vibrotactile Thresholds for Hairy Skin. *Journal of Experimental Psychology* 72(1) (1966), 47 - 50.

[149] Verrillo, R. T. and Gescheider, G. A. Perception Via The Sense of Touch. *I.R. Summers (Ed), Tactile Aids for the Hearing Impaired* (1992), 1 - 36.

[150] Vitense, H. S., Jacko, J. A., and Emery, V. K. Multimodal feedback: an assessment of performance and mental workload. *Ergonomics* 46(1) (2003), 68 - 87.

[151] Walker, B. N. and Kramer, G., "Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making," in *Ecological Psychoacousics*. Ed. J. G. Neuhoff. New York: Academic Press, 2004, 150 - 175.

[152] Walker-Andrews, A., "Taxonomy for Intermodal Relations," in *The Development of Intersensory Perception: Comparative Perspectives*. Eds. D. J. Lewkowicz and R. Lickliter. Hillsdale, N.J.: Erlbaum, 1994, 39 - 56.

[153] Wenzel, E., "Spatial Sound and Sonification," in *Auditory Display: Sonification, Audification and Auditory Interfaces*. Ed. G. Kramer. Addison-Wesley, 1994, 127 - 150.

[154] Williamson, J., Murray-Smith, R., and Hughes, S. Shoogle: excitatory multimodal interaction on mobile devices. In *Proc. CHI' 07,* ACM Press (2007), 121 - 124.