# CrossTrainer:
# Testing the Use of Multimodal Interfaces in Situ

**Eve Hoggan and Stephen A. Brewster**
Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
{eve, stephen}@dcs.gla.ac.uk www.tactons.org

## ABSTRACT

We report the results of an exploratory 8-day field study of CrossTrainer: a mobile game with crossmodal audio and tactile feedback. Our research focuses on 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. The results of this study indicate that crossmodal feedback can aid users in entering answers quickly and accurately using a variety of different widgets. Our study shows that there are times when audio is more appropriate than tactile and *vice versa* and for this reason devices should support both tactile and audio feedback to cover the widest range of environments, user preference, locations and tasks.

### Author Keywords

Tactile, audio, multimodal interaction, touchscreens, mobile interaction, crossmodal interaction.

### ACM Classification Keywords

H5.m. Information interfaces and presentation: User Interfaces: *Auditory (non-speech) feedback, Haptic I/O*.

### General Terms

Design, Experimentation, Human Factors.

## INTRODUCTION

Mobile devices, including those with touchscreens, are becoming evermore popular and are designed with the intention of everyday use. Audio and tactile feedback are becoming prevalent features in mobile touchscreen devices and recent studies [2] [9] [11] [20] [15] 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 [7] and Tactons [Brown, 2006 #5] and of the long-term use of such feedback in mobile applications. The research described in this paper 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. It has been stated that as the user's context changes so should the feedback modality [8]. For example, on a building site with high noise levels, tactile feedback may be more appropriate, whereas on a bumpy train ride, audio may be more suitable. The experiments in previous research have involved situations such as the laboratory, walking on a treadmill and travelling on an underground train, 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 in related research often test the identification and use of crossmodal icons after very short training periods commonly around ten minutes [8]. Although some longer term 2-week studies have taken place [6], 100% performance rates have never been achieved. Our 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 the following questions: how can crossmodal icons be incorporated into the design of real-world mobile touchscreen applications and improve the usability of such applications? In different real-world situations, what modality is most appropriate?

*CHI 2010*, April 10–15, 2010, Atlanta, Georgia, USA.

## BACKGROUND WORK

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 results of experiments using audio or tactile feedback [2] [9] [3] [20] [17] showing that high recognition rates can be achieved with a small amount of training. Alongside this research, there have been several studies exploring the effects on user performance and satisfaction of adding audio and tactile feedback to mobile applications [12]. However, much of the research does not give the user a choice of modalities but simply provides one modality. The majority of commercially available mobile devices have the capacity to provide both audio through speakers or headphones and tactile feedback through small built-in actuators. Given that both audio and tactile feedback appear to produce better results than visual feedback alone in terms of performance, the question is which modality should be used: audio or tactile?

The research discussed in this paper focuses on the choice between audio and tactile feedback for mobile touchscreens. This research is related to Bernsen's concept of Modality Theory [1] which addresses the mapping of information to different modalities. The outcomes of Bernsen's research include a methodology for information mapping which focuses on establishing the most appropriate modality given the task. The research in this paper 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.

Existing research has already been conducted to investigate the most appropriate modality to use for feedback when surrounded by different environmental disturbances on a subway train [10]. 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 this research focus on the effects of environmental disturbances on performance not on user preference. In our paper, 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.

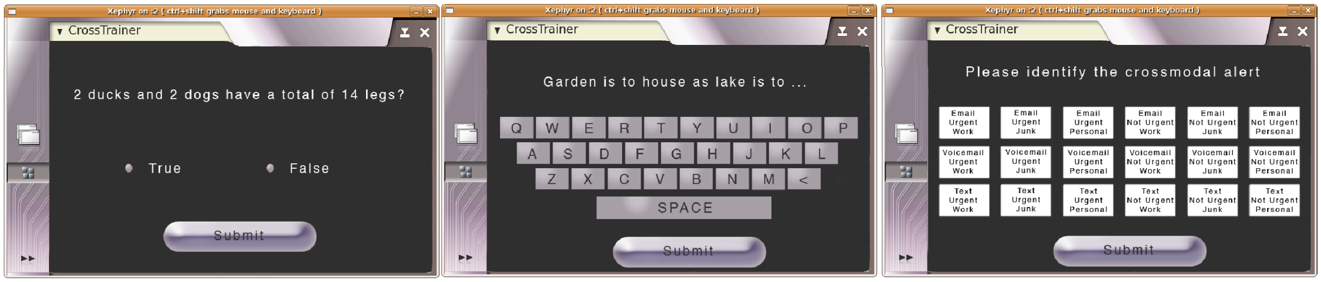
The approach used in our research involves crossmodal audio and tactile feedback using crossmodal icons [8] which can be instantiated as either an Earcon or a Tacton. Both types of icons have been subject to a great deal of research and can provide an alternative to visual icons. Earcons [4] are structured, abstract non-speech audio messages which use musical, rather than natural, sounds and use an abstract mapping that must be learned. Tactons [5] are used as the vibrotactile counterparts of Earcons in the design of crossmodal icons. Unlike multimodal, crossmodal interaction uses the different senses to provide the same information. This is much like sensory substitution where one sensory modality is used to supply information normally gathered by another. Both modalities share temporal and spatial properties so the potential shared parameters are intensity, rate, texture, rhythmic structure, duration and spatial location. These parameters are *amodal* i.e. they can specify similar information across modalities [16]. By making information available to both the auditory and tactile senses, users can receive the information in the most appropriate modality given the context.

## CROSSTRAINER

Current 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 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 (e.g. buttons, radio buttons). Upon completion, users are informed of their CrossTrainer IQ score in terms of brain age (similar to many commercial IQ games).



2 ducks and 2 dogs have a total of 14 legs?

Garden is to house as lake is to ...

Please identify the crossmodal alert

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

### CrossTrainer Hardware

CrossTrainer has been implemented on the Nokia 770 Internet Tablet, a commercial device which has been augmented with novel piezo-electric actuators [13] (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 wave files played through the device's stereo speakers (or headphones if the user prefers). We exploit this novel tactile technology by using an intramodal combination [14], i.e. combining feedback from both types of actuator, creating new types of tactile cues not possible before.

### Stimuli

CrossTrainer uses an audio and tactile feedback design based on crossmodal icons. For standard questions in CrossTrainer as seen in Figure 1 a and b, the following three parameters have been chosen for the feedback design based on previous research [8]: rhythm, texture 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 multi-dimensional icons research in Hoggan *et al.* [9]. One of the most novel feedback design aspects in CrossTrainer is the different audio and tactile textures used in the crossmodal feedback.

### Texture

Two tactile textures were created using different waveforms established in [10] 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.

**Table 1: Urgency and Texture Mapping in CrossTrainer**

| 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 | Vibraphone | Saxophone and violin |

As shown in Table 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 flute 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 (soft vibrating) horn. 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 (a percussion instrument played using a scraping motion).

### 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 [13]. Piezo elements have also been used by Luk *et al.* [17] 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 optimized 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 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.

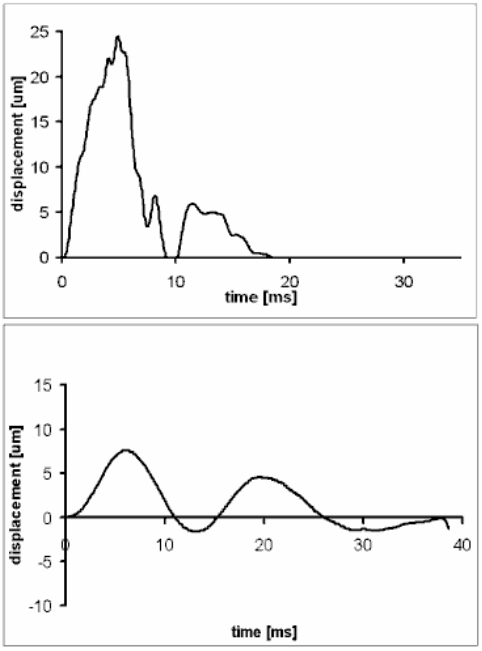


**Figure 2: Example piezo-electric and vibrotactile output.**

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 (a sharp mid-range note) and violin (a long tremolo note).

### Crossmodal Vibrotactile 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 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 previous experiments 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 [19] is ideal in this case as it shakes the whole device and can easily catch the attention of the user. The alert feedback exemplifies the use of transformational crossmodal icons where all three parameters are used – rhythm, texture and spatial location. The parameter design is based on [8] as follows:

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

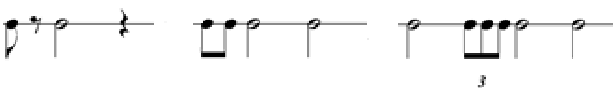


**Figure 3: Rhythm 1, 2 and 3 used in the alerts.**

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.

## EVALUATION

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 they played CrossTrainer and were also encouraged to leave voicenotes for the experimenters detailing their experiences with CrossTrainer after each game. All conditions and tasks were counterbalanced and 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.

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 they had not played CrossTrainer 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 vibration and noise levels.

CrossTrainer logged the location of the user through manual tagging by participants, surrounding noise levels 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.

### 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 they heard or felt and then choose the appropriate button on the display shown in Figure 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.

## 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.

### 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 4.

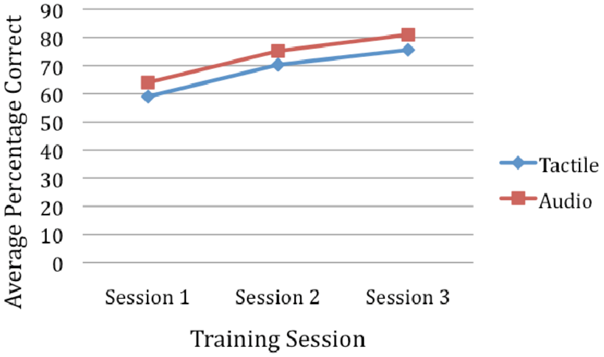


**Figure 4: Average recognition rates over 3 training sessions.**

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 and Tactons with recognition rates of 75% or higher. 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. There has been little research into multimodal training which makes these new findings beneficial to designers of such systems.

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 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.

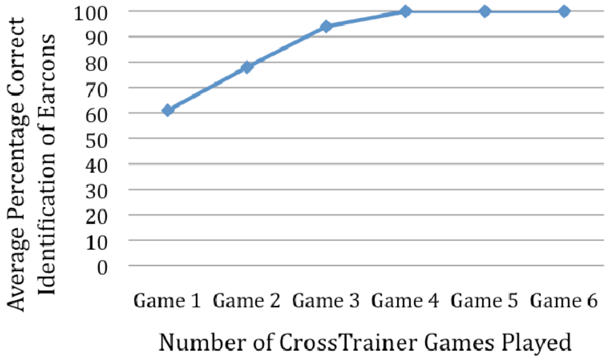


**Figure 5: Average percentage correct for Earcons in each CrossTrainer game**

The results for overall Tacton Alert recognition also showed an average recognition rate of 100% after the third game of CrossTrainer (Figure 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.

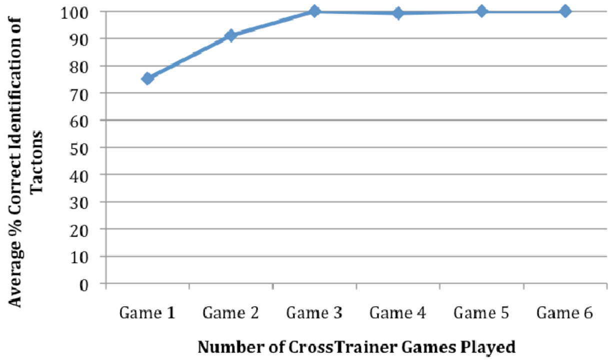


**Figure 6: Average percentage correct for Tactons in each CrossTrainer game**

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.

## Performance Over Time

### *Typing speeds:*

Figure 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.

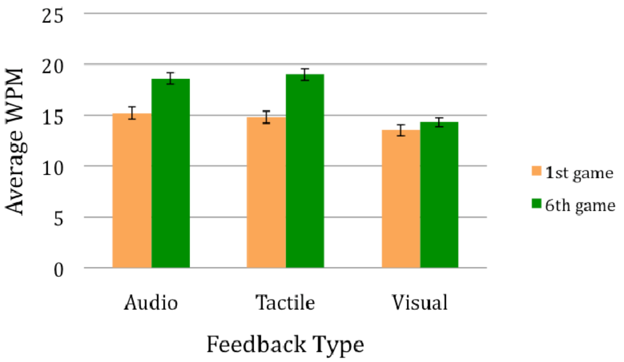


**Figure 7: Average WPM for each feedback type at the beginning and end of each condition (with standard deviations).**

A 2-factor ANOVA on typing speeds for modality types on the 1$^{st}$ 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 ($p = 0.05$).

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 using fingertips achieved on the tactile version of CrossTrainer are comparable to those found by MacKenzie *et al.* [18] 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.

### *Keystrokes Per Character (KSPC)*

KSPC was recorded for each game of CrossTrainer. 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 [21]. 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 8.

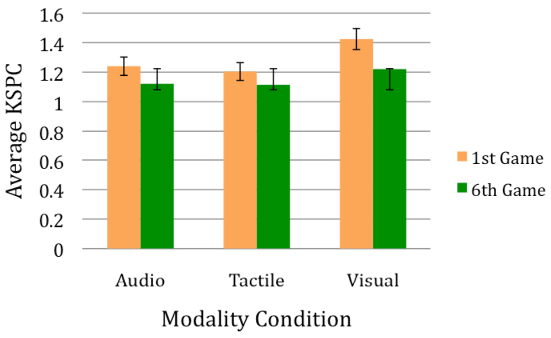


**Figure 8: Average KSPC for each modality condition from first to last CrossTrainer games (with standard deviations).**

A 2-factor ANOVA 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.

### Location of Interaction

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

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

| 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 | 3 | 5.5 |

When we analysed the location data associated with WPM we identified a number of trends (See Figure 9). A 2-factor ANOVA 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.

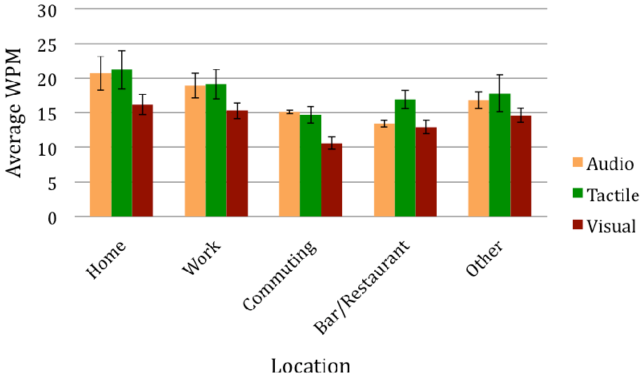


**Figure 9: Average WPM for each modality per location.**

The average KSPC for each modality and location are shown in Figure 10. An ANOVA 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, $p<0.01$).

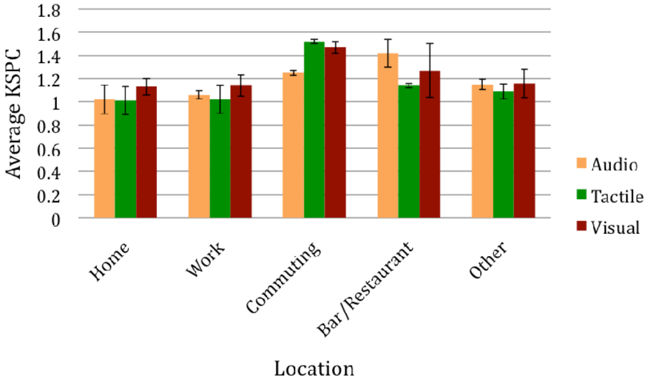


**Figure 10: Average KSPC for each modality at each location.**

Tukey tests ($p=0.01$) revealed that a significantly higher number of KSPC were generated in the tactile modality when compared to the audio 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 and correct errors. These results are comparable to those discovered in [9]. 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.

### Modality Preference

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.

#### *Modality Preference and Location of Interaction*

In terms of location, the average percentage of votes for each modality can be seen in Table 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.

**Table 3: Percentage of votes for modalities at each location.**

| | % at Home | % at Work | % Commuting | % at 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 |

#### *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, we used the 3DOF linear accelerometer in a SHAKE sensor pack [22] attached to the back of the device.

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 of Hoggan *et al.* [10] (Tables 4 and 5).

**Table 4: Vibration levels and modality preference.**

| | 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 5: Noise 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% |

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.

#### *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 they chose 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 e.g. 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.

### DISCUSSION AND CONCLUSIONS

The 8-day study of CrossTrainer has generated many interesting results. As far as we are aware, this is the first longer-term study of user preference and performance for audio and tactile feedback on mobile touchscreens. Participants were allowed to play CrossTrainer whenever and wherever they wished giving us 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 we were interested in exploring 3 areas:

- 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 2 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 our 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 walked.

Through post-study questionnaires it became apparent that location affected performance for a number of reasons. Participants stated that using CrossTrainer while commuting was difficult because of surrounding environmental sound and vibration levels whereas when at work or in a bar/restaurant surrounded by people, it was 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. Our results are very promising and if users can learn the meanings of audio and tactile icons so quickly and persistently, these icons could be a realistic way of presenting information such as device state or function/mode in everyday contexts or provide feedback for UI widgets (commercial devices already exist with basic audio/tactile feedback but there have been no studies investigating how long users take to adapt to such features to reach 100% recognition in the wild or whether users use the feedback at all on a long-term basis).

Interestingly, there were many outcomes from the analysis of modality preference. The experiment discussed in [10] 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 no point 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 82% of the time and audio for 18%. The visual version received no votes. Environmental vibration and noise levels appear to have an effect on modality choice 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 they appreciated all of the crossmodal feedback but by the end, they said 'less is more'. As they became more experienced less feedback was required (this could perhaps be addressed by reducing the duration of feedback or removing feedback from frequently used interactions). 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.

To conclude, in this paper we have described a research prototype called CrossTrainer which makes use of novel crossmodal audio and tactile feedback on a mobile touchscreen device. By applying previous work on crossmodal icons we have shown that crossmodal applications can be created where different modalities can provide the same interaction feedback, making them interchangeable. We carried out an 8-day field study of CrossTrainer with 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.

Our 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

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

## ACKNOWLEDGMENTS

Thanks to Topi Kaaresoja for his advice and NRC, Helsinki for the augmented N770. Thanks also to Andy Crossan for helping with the instrumented usability. This work was supported by EPSRC grant EP/E042740 and EP/F023405/1.

## REFERENCES

1. Bernsen, N. O. Modality Theory in Support of Multimodal Interface Design. *ERCIM Workshop on Multimodal HCI* (1993), 27 - 44.
2. Brewster, S. A. Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing* 6 (2002), 188 - 205.
3. Brewster, S. A. Sound in the Interface to a Mobile Computer. In *Proc. HCI International '99,* Lawrence Erlbaum Associates (1999), 43 - 47.
4. 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.
5. 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.
6. 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.
7. Garzonis, S. *et al.* Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. In *Proc. CHI '09,* ACM Press (2009), 1513 - 1522.
8. Hoggan, E. and Brewster, S. A. Designing audio and tactile crossmodal icons for mobile devices. In *Proc. ICMI '07,* ACM Press (2007), 162 - 169.
9. 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.
10. 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.
11. Kaaresoja, T., Brown, L. M., and Linjama, J. Snap-Crackle-Pop: Tactile Feedback for Mobile Touch Screens. In *Proc. Eurohaptics,* (2006), 565 - 566.
12. 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 '08,* ACM Press (2008), 297 - 304.
13. Laitinen, P. and Mäenpää, J. Enabling Mobile Haptic Design: Piezoelectric Actuator Technology Properties in Handheld Devices. In *Proc. HAVE '06,* (2006).
14. Lederman, S. J. and Klatzky, R. L. New Directions in Touch. *Canadian Journal of Experimental Psychology* 61 (2007), 169 - 170.
15. Lee, S. and Zhai, S. The Performance of Touch Screen Soft Buttons. In *Proc. CHI '09,* ACM Press (2009), 309 - 318.
16. Lewkowicz, D. J. The Development of Intersensory Temporal Perception: An Epigenetic Systems/Limitations View. *Psychological Bulletin* 126 (2000), 281 - 308.
17. Luk, J. *et al.* A role for haptics in mobile interaction: initial design using a handheld tactile display prototype In *Proc. CHI '06,* ACM Press (2006), 171-180.
18. Mackenzie, I. S., Zhang, S., and Soukoreff, R. W. Text Entry Using Soft Keyboards. *Behaviour and Information Technology* 18 (1999), 235 - 244.
19. Mortimer, B., Zets, G., and Cholewiak, R. W. Vibrotactile Transduction and Transducers. *The Journal of the Acoustic Society of America* 121 (2007), 2970 - 2977.
20. 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.
21. 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 Press (2003), 113 - 120.
22. Williamson, J., Murray-Smith, R., and Hughes, S. Shoogle: excitatory multimodal interaction on mobile devices. In *Proc. CHI' 07,* ACM Press (2007), 121 - 124.