

Audio or Tactile Feedback: Which Modality When?

Eve Hoggan¹, Andrew Crossan¹, Stephen Brewster¹ and Topi Kaaresoja²

Glasgow Interactive Systems Group
University of Glasgow, Glasgow, G12 8QQ, UK
{eve,ac,stephen}@dcs.gla.ac.uk

Nokia Research Center
Helsinki, Finland
topi.kaaresoja@nokia.com

ABSTRACT

When designing interfaces for mobile devices it is important to take into account the variety of contexts of use. We present a study that 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. The aim of the study is to show at what exact environmental levels audio or tactile feedback become 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. These findings will allow designers to take advantage of sensor enabled mobile devices to adapt the provided feedback to the user's current context.

Author Keywords

Mobile interaction, crossmodal, touchscreen, audio, tactile.

ACM Classification Keywords

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

INTRODUCTION

As touchscreen mobile devices become more powerful and equipped with more sensors, they bring new interaction advantages and disadvantages. On the one hand, wherever the user goes, the mobile device goes too allowing interaction to occur in a variety of different situations and locations. On the other, the very fact that interaction can occur in a variety of situations leads to new challenges for designers who must make applications usable at all times.

Despite all of the advancements in mobile touchscreens, one key issue has not changed: information is mainly displayed visually, placing a higher load on the visual sense when interacting with the device. Given the small size of

mobile devices, the widgets and text displayed on the screen can be small making them difficult to see and operate. Situations, locations and even screen configurations can change yet the form of output has remained the same. Mobile usage scenarios afford many completely different interactions, so displays should be designed accordingly.

In response to these issues, there has been much research in the areas of tactile and audio feedback with results showing that they can improve performance over purely visual displays [1, 6]. However, like visual feedback, there are drawbacks as, if the environment is too loud or there is too much vibration, audio or tactile feedback may be ineffective. Users must be able to switch effortlessly between modalities to select the most appropriate feedback for their task and environment. As stated by Hoggan *et al.* [3], as the context changes, so should the feedback modality. If the mobile device could automatically switch to the most effective type of feedback this would lead to greater usability, more socially appropriate interaction and less redundant feedback.

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 [7]. We exploit this by using the sensors to establish if it is too noisy for audio or too bumpy for tactile feedback and then switch to the more appropriate type.

The experiment described here investigated fingertip text entry performance using a QWERTY keyboard displayed on a touchscreen mobile device (with visual, audio or tactile feedback) in an everyday situation (an underground train). We measured vibration and noise levels to see if performance on one modality was better than the others at different levels of environmental disturbance. Our research questions were: at what vibration level does tactile feedback become ineffective? At what noise levels does audio feedback become ineffective? When is audio feedback more appropriate than tactile feedback and vice versa? The overall aim was to define the levels at which audio or tactile feedback in a real-world setting is no longer valuable.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2009, April 4–9, 2009, Boston, MA, USA.

Copyright 2009 ACM 978-1-60558-246-7/08/04...\$5.00

Crossmodal Feedback

The approach used in this research involves crossmodal audio and tactile feedback using crossmodal icons [2] which can be instantiated as either an Earcon or a Tacton. 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 [4]. By making information available to both the auditory and tactile senses, users can receive the information in the most appropriate modality given the context.

EXPERIMENT

In order to measure extreme vibrations and sounds we needed a controlled environment where high levels of noise and vibration occur naturally. We chose to investigate the interaction on the Glasgow underground. 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. Hoggan *et al.* [3] used this for a previous study and showed that it was an effective test environment for mobile text entry.

Crossmodal Stimuli and Hardware

A simple set of crossmodal Tactons and Earcons were created to represent the different keyboard events and keys that exist on a touchscreen keyboard. This stimuli set was based on the design by Hoggan *et al.* [3] in a tactile text entry study with an equivalent audio version included. A fingertip-over event (when the fingertip has touched a button) 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 (when the fingertip moved over the edge of a button) 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.

The C2 Tactor from EAI (www.eaiinfo.com) was used for this study to present the tactile feedback [3]. Audio feedback was created using standard wave files designed in an audio synthesis application. The feedback was presented through an earpiece. 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 us to calibrate the noise levels and estimate the sound levels heard by users through the earpiece before the train journey.

Instrumenting the Usability Evaluation

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 [9] attached to the back of each participant's hand holding the device (Figure 1).



Figure 1: Experiment set-up with PDA, C2 Tactor, bandage securing the SHAKE to hand and text entry GUI screenshot.

All SHAKES logged through Bluetooth to a 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 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. On average, all tactile stimuli on the touchscreen keyboard was 250Hz.

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. We used a between-subjects design where the conditions were a touchscreen keyboard with audio, tactile and visual feedback. For each journey, three participants each performed a different condition (Figure 2). Overall there were 4 journeys on the subway.



Figure 2: Experiment set-up on underground train.

The methodology and experimental application were based on a previously successful study which measured the effects of tactile feedback on touchscreen text entry (see [3] 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 they were all on the same journey they all experienced the same vibration and noise levels at the same time. Therefore we could compare speed and accuracy of text entry 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 1, right). Each phrase (from a set by MacKenzie [5]) 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.

RESULTS

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 4 in parallel with the vibration and noise levels for each modality. KSPC is the number of keystrokes required, on average, to generate a character for a given text entry technique with the ideal being 1 per character [8]. KSPC were recorded because accuracy scores (Figure 3) were based on whether submitted phrases matched the given phrase and did not include corrections as errors.

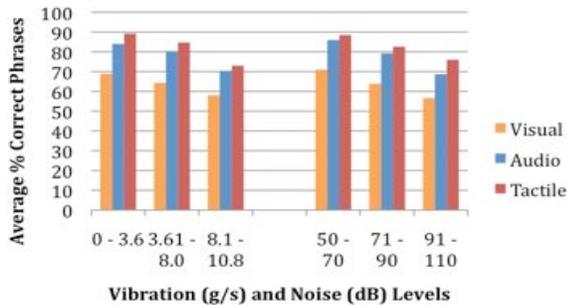


Figure 3: Average percentage of phrases entered correctly.

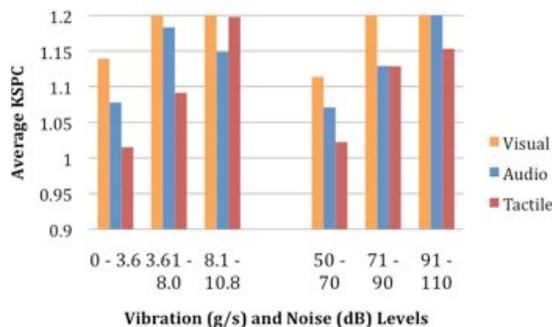


Figure 4: Average KSPC for vibration and noise level sets.

A two-factor 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 Tables 1 and 2.

| 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 1: 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 2: Summary of the KSPC and noise results.

Text Entry Rate (Words Per Minute)

The mean words per minute (WPM) in parallel with vibration and noise levels are shown in Figure 5.

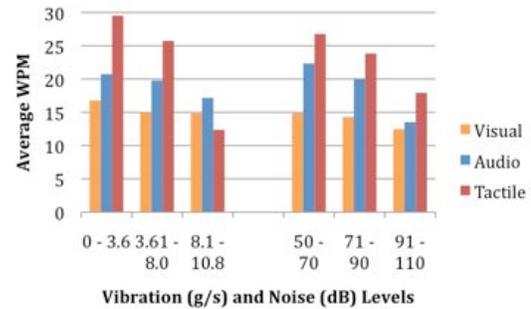


Figure 5: Mean WPM for each set of vibration and noise levels.

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 3: Summary of the 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 4: Summary of the 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 two-way ANOVAs 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 in the tactile condition compared to audio ($F = 34$, $F_{crit} = 3.06$, $p < 0.001$). In terms of words per minute, there are 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 ($F = 23.1$, $F_{crit} = 3.06$, $p < 0.001$).

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 ($F = 4.79$, $F_{crit} = 3.06$, $p < 0.001$). 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 ($F = 11.43$, $F_{crit} = 3.06$, $p < 0.001$).

DISCUSSION AND CONCLUSIONS

The results show that while 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 keystrokes per character 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 our 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.

In conclusion, this paper presented an experiment investigating text entry performance on mobile touchscreen devices (one with visual feedback only, one with audio and one with tactile) on an underground train. 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. As expected, audio feedback was shown to become ineffective in noisy environments and tactile feedback become ineffective in bumpy environments. However, this study has established the exact levels at which these modalities become ineffective. The study reported here focused on an extreme situation (the underground train). Future studies will involve other situations such as walking or music concerts to confirm that our findings apply whenever these vibration and noise levels occur. Our results 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 modalities.

ACKNOWLEDGEMENTS

This work was supported by EPSRC grant EP/E042740.

REFERENCES

1. Chang, A. and O'Sullivan, C. Audio-Haptic Feedback in Mobile Phones. In *Proc CHI '05*, ACM (2005), pp. 1264 - 1267.
2. Hoggan, E. and Brewster, S. A. Designing Audio and Tactile Crossmodal Icons for Mobile Devices. In *Proc ICMI '07*, ACM (2007), pp. 162 - 169.
3. Hoggan, E., Brewster, S., and Johnston, J. Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In *Proc CHI '08*, ACM (2008), pp. 1573-1582.
4. Lenay, C., Canu, S., and Villon, P. Technology and Perception: The Contribution of Sensory Substitution Systems. In *Proc ICCT*, IEEE (1997), pp.44.
5. MacKenzie, I. S. and Soukoreff, R. W. Phrase Sets for Evaluating Text Entry Techniques. In *Proc CHI '03*, ACM (2003), pp. 754 - 755.
6. Poupyrev, I. and Maruyama, S. Tactile Interfaces for Small Touch Screens. In *Proc User Interface Software and Technology*, ACM (2003), pp. 217 - 220.
7. Raento, M., Oulasvirta, A., Petit, R., and Toivonen, H., "ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications," *IEEE Pervasive Computing*, vol. 4(2), pp. 51-59, 2005.
8. 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), pp. 113 - 120.
9. Williamson, J., Murray-Smith, R., and Hughes, S. Shoogle: Excitatory Multimodal Interaction on Mobile Devices. In *Proc CHI '07*, ACM (2007), pp. 121 - 124.