Evaluating the Use of Sound as a Navigational Aid on a Mobile Device

Crease, Murray; Lau, T.Y.

Publisher's version / Version de l'éditeur:

Workshop on Mobile HCI and Sound (Hosted by Mobile HCI 2004) [Proceedings], 2004

National Research Council Canada
Institute for Information Technology

Conseil national de recherches Canada
Institut de technologie de l'information

NRC · CNRC

*Evaluating the Use of Sound as a Navigational Aid on a Mobile Device – Discussion **

Crease, M., and Lau, T.Y.
September 2004

* published at the Workshop on Mobile HCI and Sound (Hosted by Mobile HCI 2004). Glasgow, United Kingdom. September 19, 2004. pp. 19-24. NRC 48064.

Canada

Evaluating the Use of Sound as a Navigational Aid on a Mobile Device - Discussion

Murray Crease[1] and Thenk Yet Lau[2]

[1] National Research Council - Institute for Information Technology,
46 Dineen Drive, Fredericton, NB, Canada, E3B 9W4.
murray.crease@nrc-cnrc.gc.ca
http://iit-iti.nrc-cnrc.gc.ca

[2] Faculty of Computer Science, University of New Brunswick,
P.O. Box 4400, 540 Windsor Street, Gillin Hall, E126,
Fredericton, NB, Canada, E3B 5A3.
h78jc@unb.ca

Abstract. The use of sound has obvious advantages when considering interaction with a mobile device and yet it remains under utilised. One reason may be that the sounds commonly evaluated are not aesthetically pleasing [6]. This paper discusses a second possibility – that users may prefer not to use sound regardless of its effectiveness – and describes a proposed experiment to determine whether this is the case. The two stage experiment will initially investigate whether users are able to navigate at least as effectively using audio cues as they are with visual cues. The participants will then be presented with both audio and visual cues to see if they choose to use the visual cues regardless of how well they performed using the audio cues only.

1 Introduction

There are many research examples of the use of sound at the human-computer interface on both desktop and mobile devices (e.g.[2, 7]). The reality is, however, that interfaces use little - if any - more audio feedback now than ten years ago. One possible reason for this is that the majority of audio feedback that has been evaluated by the research community has been designed with the emphasis on functionality rather than aesthetics. One notable example where the aesthetics was given close examination was the “Out to Lunch” system [3]. Cohen found that users were far more willing to use a second iteration of the system which had professionally designed sounds as opposed to the initial version which used less aesthetic sounds. Leplâtre *et al.* successfully showed that music could be used as an aesthetic alternative to standard earcons [5]. More recently, the relationship between the aesthetics and functional properties of audio feedback has been investigated [6].

A second possibility why audio feedback has not been widely adopted at the human-computer interface is that, regardless of usability, users are simply not prepared to use the sounds. This may be due to the cognitive effort required to map the sounds to actions or events. This would seem unlikely, however, given that many of the evaluations undertaken have used some form of workload analysis which in some cases has indicated that users find the workload reduced when using sound (e.g. [2]). It is more likely to be due to users not expecting to hear sound in the human-computer interface for anything other an alert. This would be surprising given that people gain information from sound on an everyday basis, for example using the whirring of a hard-drive to determine whether processing is occurring. It may be that users find this sub-conscious use of sound is in someway more acceptable than the conscious use of designed audio feedback from a computer. This paper discusses the design of an experiment which aims to answer two main questions: (1) can participants use simple audio cues to navigate as effectively as when using visual cues; and (2) when given the choice, will the participants continue to use the audio cues (for information rather than alerts) or will they rely on the visual cues?

2 Related Work

There are several systems which have used audio feedback to provide users with GPS information. The MOBIC system is an example of a system designed to allow blind users navigate [10]. MOBIC uses speech to provide users with assistance for macro-navigation – the navigation through the distant environment – which it typically done using visual cues such as church steeples. One of the main findings of the user analysis undertaken was that the users did not want to wear headphones as it was felt that this would block out useful, environmental sounds which are especially important to visually-impaired users. Audio GPS was designed to allow sighted users to receive navigational information using spatialised non-speech audio [4]. Non-speech sounds were chosen to minimise any interference with the users' conversations. The system presented the users with two pieces of information: distance to their destination (or intermediate waypoint) and its direction. Initial trials indicated that the sounds were effective in allowing users to discern the direction of the destination but that the implementation of the system meant that it was slow to respond to a user's change in direction.

3 Experimental Design

An experiment has been designed to investigate the questions outlined at the end of Section 1. The experimental task is in the form of a game, with two participants – in separate, partitioned areas of a room - 'competing' against each other to try and navigate through a virtual maze as quickly as possible. The participants cannot see the the maze but only the numbered grid in which the maze is located Figure 1(b & c). An experimental server provides the users with navigational cues which guide them through the maze. The experimenter has an interface which shows the status of both areas and the location of the two participants. This enables the experimenter to easily manage the running of the experiment. Figure 1(a) shows the maze as it is presented to the experimenter. The grey cells indicate the safe route from the start point (25) to the end point (27 – shown in yellow). The participant's current location is shown in blue (76) and the next location proposed by the experimental server is shown in green (66). Figure 1(b) shows the physical maze the participants must navigate through. The maze is a virtual maze with the only physical manifestation being the numbered grid on the floor of the lab. In this case, the participant is standing in cell 76. Figure 1(c) shows the maze as it is presented to the user on a handheld device. Again, the participant can only see the grid in which the maze is located. The participant's location is shown in green and the red square (containing the word 'help') at the bottom right of the interface indicates a visual cue is available.

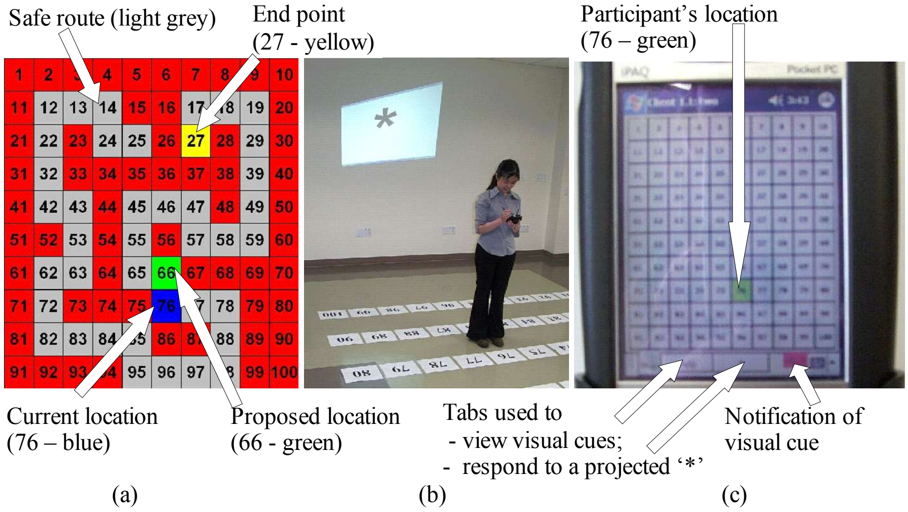


Figure 1 – three views of the same maze and/or the numbered grid it is in: (a) as it is presented to the experimenter; (b) as represented by a grid of cells on the floor which the participants must navigate through; and (c) as it is presented to the participant on the handheld device.

The experimental participants are required to navigate through the maze. In Figure 1(b) the participant is standing in cell 76 and has been sent the instruction to move forward (to cell 66). Once the participant has moved she can enter her new location on the handheld device by tapping the appropriate cell. This sequence of events will continue until she has completed the maze (i.e. reached cell 27). If the participant should make a mistake (i.e. not go to the specified cell) she will be instructed to go back 5 steps in the maze (in this case to cell 97) before she can continue.

To simulate a realistic mobile environment, the participants are also required to monitor their surroundings and react accordingly. Projectors display symbols at random intervals in front of and behind the participants (Figure 1(b) shows a '*' being projected behind the participant[1]). Six symbols were used - 'v', 'w', 'x', 'y', 'z' and '*' – with the users having to respond to the projection of a '*' by pressing a tab at the bottom of the interface on the handheld. In this way, we are mimicking a real mobile environment where users are required to be aware of their surroundings.

3.1 Navigational Cues

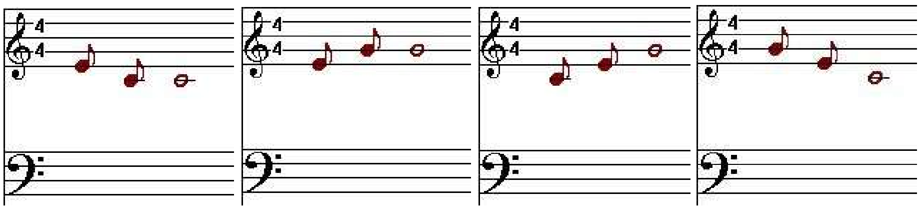
The navigational cues are simple cues which represent the directions left, right, forward, and backwards. The cues can be presented visually and/or sonically. The visual cues take the form of an arrow pointing in the appropriate direction. These cues are

1 Figure 1(b) shows a temporary lab where the projection is done from the side. When the experiment is run the projection will be done from ceiling mounted projectors.

accessed by pressing the “Help” tab at the bottom of the screen Figure 1(c). The arrows are visible for 350ms[2] before the display returns to the view of the map. The arrow can only be viewed once for each step.

The audio cues are earcons[1] which all share the same basic structure: two notes of duration 80ms followed by a third note of duration 480ms. All the notes are played in a piano timbre. The directions are differentiated using the notes’ pitch:

- Left – E3, followed by C3, followed by C3[3] (Figure 2(a).)
- Right – E3, followed by G3, followed by G3 (Figure 2(b).)
- Forward – C3, followed by E3, followed by G3 (Figure 2(c).)
- Backward – G3, followed by E3, followed by C3 (Figure 2(d).)



(a) (b) (c) (d)

Figure 2 -The four sounds: (a) left; (b) right; (c) forward; and (d) backward

By using pitch as the only parameter by which the sounds can be distinguished, the possibility for encoding more complex navigational cues which could inform the user about a more detailed direction and/or the distance to be travelled is left open. By relying only on relative pitch (i.e. the way the pitch changes within the earcons) as opposed to relying on absolute pitch (i.e. the way the pitch changes between earcons) these sounds follows the guidelines outlined by Lumsden *et al.*[8]. Because the earcons are not spatialised they can be easily generated by a handheld device and the user is not required to wear headphones which may block out other sounds [10].

3.2 Experimental Procedure

The experiment consists of three conditions: visual cues only; audio cues only; and both audio and visual cues. The first two conditions will be counter-balanced with the audio-visual condition always presented last. Each condition will consist of a short training session where the users will be able to familiarise themselves with the navigational cues for that condition followed by a short training maze consisting of 16 steps. The participants then have to navigate a 40 step maze for the actual condition.

4 mazes are used with each designed so that they have an equal number of forward and backward steps (10 each for the actual mazes and 4 each for training). This has been done to ensure the mazes are all of a similar difficulty as pilot studies showed it to be harder to move backwards than forwards. Furthermore, the order of presentation for the mazes is such that each maze is used an equal number of times for each condition after 2 sessions, where a session consists of 2 participants working in parallel

---

2 350ms was chosen as a suitable length of time to display the visual cue as Öquist *et al.* report that the average time required to fixate – or parse – visual information is ~330 ms [9].

3 C3 is middle C (261.63 Hz).

(Table 1). The mazes are not presented in the same order to the two participants in a particular session because the areas are only separated by a partition. If the two participants were navigating the same maze it may be possible for one user to successfully navigate using the second user's audio cues. To keep the mazes as similar as possible though, the mazes A and C share the same layout as mazes B and D respectively but with the start and end points reversed.

Session 1			**Session 2**		
Participant 1	**Condition**	**Participant 2**	**Participant 1**	**Condition**	**Participant 2**
Maze A	audio	Maze B	Maze A	visual	Maze B
Maze C	visual	Maze D	Maze C	audio	Maze D
Maze B	a/v	Maze C	Maze D	a/v	Maze A

Table 1 – The order the conditions and mazes are presented in.

Because the first two conditions will always be audio only or visual only these can be considered in isolation. This will allow the effectiveness of the audio cues versus the visual cues to be determined, answering the first research question defined at the end of Section 1. The effectiveness will be measured in terms of time taken to complete the mazes, percentage of '*'s missed and number of incorrect steps taken. The second question will be answered by seeing how often the participants choose to use the visual cues in the third condition. This will be measured by recording how often the participants click on the 'help' tab to view the visual navigational cues.

4 Conclusions

This paper has described has described the design of an experiment which will hopefully demonstrate that participants are able to use simple audio cues to navigate at least as effectively as with visual cues. The experiment will also investigate whether the participants continue to use the audio cues as navigational cues (as opposed to simple alerts) when they have the option of using the visual cues. If the participants are able to use the audio cues as effectively as the visual cues and yet still rely on the visual cues when given the option this will require further investigation to determine why. If, on the other hand, the participants continue to use the audio cues when given the choice then that is further evidence that users are happy to use sound at the human-computer interface.

The design of the experiment tries to mimic a real mobile environment by introducing visual distractions which the participants are required to attend to. Other than the audio cues being presented to a participant's 'opponent', however, there are no audio distractions which would normally be found in such an environment. Future experiments will have to include such distractions not only to make the experiment more realistic but also to determine whether users are able to use audio feedback effectively in such an environment.

Acknowledgments
We would like to thank Danny D'Amours for his technical assistance and Jo Lumsden for helping with some of the finer points of the experimental design. The software was written in the EWE programming language (www.ewesoft.com). The sounds were de-

signed using MIDI Studio V4.20 (www.sonicspot.com/midistudio/midistudio.html) and converted to wave files using Midi2Wave Recorder V3.5 (www.midi2wav.com).

References

[1] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction*, vol. 4 (1), pp. 11-44, 1989.

[2] S. Brewster, "Overcoming the Lack of Screen Space on Mobile Computers," *Personal and Ubiquitous Computing*, vol. 6(3), pp. 188.205, 2002.

[3] J. Cohen, "Out to Lunch: Further Adventures Monitoring Background Activities," in *Proceedings of the Second International Conference on Auditory Display (ICAD'94)*. Santa Fe, NM: Addison-Wesley, 1994.

[4] S. Holland, D. R. Morse, and H. Gedenryd, "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface," *Personal and Ubiquitous Computing*, vol. 6 (4), pp. 253-259, 2002.

[5] G. Leplâtre and S. Brewster, "An Investigation of Using Music to Provide Navigational Cues," in *Proceedings of ICAD 98*, A. Edwards and S. Brewster, Eds. Glasgow, Scotland: International Community for Auditory Display, 1998.

[6] G. Leplâtre and I. McGregor, "How to Tackle Auditory Aesthetics? Discussion and Case Study," in *Proceedings of ICAD 04 - Tenth Meeting of the International Conference on Auditory Display*, S. Barrass and P. Vickers, Eds. Sydney, Australia: International Community for Auditory Display, 2004.

[7] G. Lorho, J. Hiipakka, and J. Marila, "Structured Menu Presentation Using Spatial Sound Separation," in *Mobile Human-Computer Interaction : 4th International Symposium, Mobile HCI 2002*. Pisa, Italy: Springer-Verlag Heidelberg, 2002, pp. 419-424.

[8] J. Lumsden, S. Brewster, M. Crease, and P. D. Gray, "Guidelines for Audio-Enhancement of Graphical User Interface Widgets," in *Proceedings of HCI 2002*, vol. II, F. Détienne, X. Faulkner, and J. Finlay, Eds. London,UK: Springer-Verlag, 2002, pp. 6-9.

[9] G. Öquist and M. Goldstein, "Towards an Improved Readability on Mobile Devices: Evaluating Adaptive Rapid Serial Visual Presentation," in *Proceedings of Mobile HCI 2002*, F. Paternò, Ed. Pisa, Italy: Springer-Verlag, Berlin Heidelberg, 2001, pp. 225-240.

[10] H. Petrie, V. Johnson, T. Strothotte, A. Raab, S. Fritz, and R. Michel, "MOBIC: Designing a Travel Aid for Blind and Elderly People," *Journal of Navigation*, vol. 49(1), pp. 45-53, 1996.