Non-Visual Interfaces for Wearable Computers

Stephen Brewster and Ashley Walker
Department of Computing Science,
University of Glasgow
Glasgow, G12 8QQ, UK
Email: stephen@dcs.gla.ac.uk Web: www.dcs.gla.ac.uk/~stephen

INTRODUCTION

As computing devices drop in size and rise in power/bandwidth, novel interface techniques are needed to keep pace. Interfaces which rely on screens and keyboards/handwriting are not always effective in mobile computing scenarios. Visual displays are wasted on eyes engaged in navigational tasks such as walking or driving. Similarly, keyboards and handwriting-based interfaces are difficult to use whilst the arms and hands are involved in real-world tasks (including posture control). Furthermore, these input techniques typically require a dedicated visual resource in the interaction loop.

There are two aspects to our work. The first is looking how we can overcome some of the problems in current wearable or mobile displays by the addition of sound. The second aspect looks at doing away with the visual display altogether. The technique consists of an egocentric three-dimensional (3D) auditory display space in which natural head gestures control interaction. It overcomes the presentation bandwidth limitation inherent in audio displays by concurrently playing multiple spatialised sound streams (speech and non-speech) around a user's head. In turn, it overcomes the possibility of acoustic clutter in the resulting 'soundscape' by giving users powerful and natural mechanisms for swapping their attention between streams. By detecting the focus of a listener's auditory attention via his/her natural acoustic orienting gestures (i.e., which streams she orients her ears toward), the system can artificially raise and lower the volume of streams to enhance selective listening.

OUR APPROACH

Our approach to solving the problems with interfaces to wearables and mobiles is in two stages. In the first stage we have looked at how current interfaces to handheld computers can be improved by the use of audio, in the second stage we have developed purely auditory interfaces with gestural control to avoid the need for visual attention.

Stage one: augmenting the display of current handheld computers

As discussed above, there are many problems with the current displays on handheld and wearable computers. In this stage of our work we wanted to see if we could reduce the dominance of visual presentations on mobiles by the addition of sound to existing interface widgets. We conducted a series of experiments to investigate whether the size of on-screen widgets could be reduced by the addition of sound (for full details of the research see [6]). This would allow screen clutter to be reduced or to allow more information to be displayed. However, if components are reduced in size it is important to make sure that they remain usable; otherwise the size reduction is pointless.

We added sound to improve feedback from buttons. Buttons in mobile displays can be hard to use as they are small and feedback from them is limited and can be obscured by the stylus or finger used to press them [6]. We added audio feedback similar to that which was previously shown to be beneficial in desktop interactions [8]. Simple structured non-speech sounds, *Earcons* [4, 9], were used for the audio feedback (the sounds were simple as the Palm III device used only provides very basic audio capabilities). Sound gave feedback on when users had pressed down on a button, when it was released correctly and when it was released incorrectly (and the button not activated).

The task the users had to perform was entering numeric codes as fast as possible via a calculator style keypad. This maximised the number of button presses our users had to make. The experiment was run on a 3Com Palm III handheld computer with input via a stylus. The 16 users used interfaces with two different sized buttons with and without sound (see Table 1) in a within-groups design. To make sure that the task was realistic, participants had to walk 10m laps along a path by the side of one of the university buildings whilst performing the tasks. To measure usability workload (using the standard 6 NASA TLX factors plus annoyance and user preference) and total amount of data entered were collected along with the distance walked by each participant in each condition.

Condition: Button size / Treatment: Sound type		
Standard (16 x 16)	Silent	Sound
Small (8 x 8 pixels)	Silent	Sound

Table 1: Format of the experiment. Conditions were fully counterbalanced.

Results

For the standard sized buttons sound significantly reduced workload in terms of mental demand, physical demand, effort expended, annoyance experienced and performance level achieved. Participants significantly preferred the buttons with sound to those without. For the small buttons sound again had a positive effect: it significantly reduced workload for mental demand, physical demand and effort expended. Participants significantly preferred the buttons with sound to those without.

Results for the numbers of codes entered with the different buttons sizes showed significantly more were entered with the sonically-enhanced buttons for both button sizes. In this experiment there was no significant difference between the small buttons with sound and the standard silent buttons.

Significantly more laps were walked using the standard sized buttons with sound than without. There was no significant difference in the number of laps walked between the small buttons with sound and those without. There was no difference shown between the small buttons with sound and the large buttons without.

Discussion

The qualitative results show that sound had a big effect on workload for both button sizes. In almost all of the categories workload was significantly reduced when sound was present. This is important as workload reductions mean that users do not have to devote so many of their cognitive resources to perform the task on the mobile or wearable computer. This means that they have more resources available to deal with walking, driving etc. when using a mobile device.

The quantitative results also back up the hypothesis that sounds improve usability, as participants were able to enter significantly more codes when sounds were present for both button sizes. The sounds helped users target the buttons better, know when they had been pressed correctly and when they had been mis-hit. This made them significantly easier to operate.

Participants also walked further when sounds were present. We suggest that this is because they did not have to devote so much of their visual attention to the device (as they could get feedback through their ears) and so could look where they were going. This is very important to consider when developing usable mobile computer systems.

These results show that sound can have a big impact on the display of a wearable or mobile computer. However, this stage just considered interfaces of current devices. The next section describes new interaction techniques that are completely non-visual.

Stage two: audio and gestural interfaces

The next part of our design focussed on developing new forms of interaction for mobiles that do not depend on visual techniques. A key feature of what we are proposing involves a profound increase in mobile device display space. Wearables typically have a single loudspeaker or, at best, playback stereo audio. This represents a display bottleneck, as streams of audio messages queue up for sequential presentation from a relatively small soundscape. By contrast, our approach is to use 3D sound technology [2] to increase the size of the display. This allows information to be displayed from anywhere in a head-centred spherical display space.

The egocentric nature of the display space facilitates navigability. Sound streams, although they have no visual correlate, remain in a fixed position relative to a user's head. To achieve this, a head tracker is used to continuously re-spatialise sound streams in response to a user's head movements.

This architecture provides a rich space in which to display different kinds of information [3, 7] simultaneously. For example, users might have an MP3 file playing in the space behind them, some sonified stock market data on the right (perhaps the user is listening to tell when the values of shares are changing rapidly to know when to buy or sell), there might be a news article about the stocks on the left of the audio space. A phone call might then be received which is presented from the front of the display space. Because each sound source is played from an independent spatial position, the seperability of each stream is enhanced [5]]. Three-dimensional sound has been shown to have many benefits for presenting multiple streams of sound for teleconferencing etc. [3].

Once we have a display space then we need a way of controlling it. Our aim is to use methods that do not require visual attention as the primary input. We are investigating the use of three categories of gestures: head, hand and device gestures.

Our aim is to make the gestures a simple, intuitive and natural as possible. It is important that users do not feel self-conscious when making them, for example. We conduct usability tests of our gestures to ensure this is the case.

Head gestures

As we are using a head tracker to re-spatialise the sounds, we have a device that we can use for gesture recognition. When focusing on a spatialised stream of sound (real or virtual), listeners reorient their heads and upper body in characteristic ways. The goal of these postural changes is to position the sound source of interest in the region of the frontal sound field where hearing is most acute. In this region, a sound stream can be heard louder and clearer (i.e., there is less filtering of extreme frequencies). The exact orientation of a listener's head depends upon a number of factors, including the nature of the auditory environment (how densely packed it is with competing sounds and where they are) and the anatomy of the listener (the shape of the head and pinnae, the position of the ears, etc) as well as the mobility of the neck and back. Therefore, the detection algorithm must employ a number of redundant and fuzzy metrics in order to determine in which sound stream a listener is interested. These detection rules operate on timing, orientation and rate of movement data.

In order to enhance selective stream listening in a dense soundscape, sound sources selected in this way must be made more audible (louder). However, sound level preference varies from person to person. Therefore, the automatic volume control mechanisms will artificially increase the volume on selected streams (i.e., increase volume above the amount which naturally occurs as a result of the repositioning) until the listener's head returns to the upright/neutral position. Detecting a listener's sensitivity to a sound in this way allows the system to alter the volume in a user-sensitive fashion. Once the listener's head is back in the neutral position, the volume on the other streams is adjusted to take into account the user's current attentional state. This will be attempted via a weighting factor, which takes into account the listening history associated with each stream.

Hand and device gestures

The same tracking device employed to monitor head movement, can also be employed can also be mounted on a mobile device or hand/finger to allow explicit manipulation of sound sources. The most obvious utility of such a tool is to support the selection, de-selection and movement of sources by a series of grab and move operations. Here there is a direct analogy with mouse/cursor kinds of control except that the gestures we propose employ the same physical actions and ranges of motion as one might employ to move a physical device such as a loudspeaker. Moreover, these interactions do not require the user to look at a display; rather, they provide real-time audio feedback to help the user's hand navigate through a head-centred display space. To this end, we are developing an audio cursor whose frequency characteristics are systematically altered based on the cursors proximity to a sound stream.

We have experimented with other forms of gesture-based content entry. For example, one could be used for writing simple Graffiti style characters (Grafitti is the handwriting recognition system used on the 3Com Palm series of computers [1]). This would allow access to simple information without the need for opening a device, turning it on, getting the stylus out and then writing. Using our method would not be a comfortable way to write a novel, but it could be used for simple input (e.g., telephone numbers, appointments, etc.) and controlling functions.

3D audio widgets to use in the audio space

We have developed a set of widgets that can be used in the audio space. One of the most sophisticated is the 3D audio progress bar [10, 11]. This uses the position of a sound in space around the listener's head to indicate the amount downloaded and movement around the head to indicate rate (see Figure 1). Comparison of this audio progress bar with a conventional visual progress bar shows that it can be used to track download progress more

Figure 1: A 3D audio progress bar.

effectively and, moreover, facilitates visually demanding foreground tasks [10, 11]. Along similar lines we have developed a diary widget that presents time of appointment on a clock face surrounding a user. A meeting at 3:00 would be by the user's right ear, a meeting a 9:00 by the left ear. An informal experiment to assess the effectiveness of such a display showed that it enhances recall of events as compared with the standard list of a normal diary application.

CONCLUSIONS

The aim of our research is to use audio, and especially non-speech audio, to present information in handheld and wearable computers. Mobile users need their eyes to look where they are going so we aim to present as much information as possible in sound. The first stage of our work has been to improve current displays of handheld and wearable computers with the addition of sound. The results of this work have shown that sounds can significantly improve usability and make such computers more effective in mobile environments.

The second stage of our work has been to develop new interaction techniques that do not rely on visual presentation. We have created a 3D auditory and gestural environment that allows multiple streams of data to be presented in sound and allows the user to control them by making head, hand or device gestures. The use of spatial audio profoundly increases the display space because windows of audio information can be presented from anywhere within a sphere centred on the user's head. In a mobile context, virtual sound can be rendered for a pair of headphones or shoulder-mounted speakers. These sound display options are becoming integrated with handheld computers (for example, the new version of Microsoft PocketPC for palm-top computers incorporates an MP3 audio player) so users are likely to have easy access to headphones. Headphones are also becoming wireless. One of the first Bluetooth devices from Erickson is likely to be a wireless hands-free kit for a mobile phone.

ACKNOWLEDGEMENTS

This work is part funded by EPSRC grant GR/L79212. Initial research was funded by a Microsoft New Discoveries grant.

REFERENCES

1. 3Com. 3Com Palm http://www.palm.com/. 2000
2. Begault, D.R. 3-D sound for virtual reality and multimedia. Academic Press, Cambridge, MA, 1994.
3. Billinghurst, M., Bowskill, J., Jessop, M. and Morphett, J. Spatial conferencing using a wearable computer. Personal Technologies 3, 1 & 2 (1999), 72-80.
4. Blattner, M., Sumikawa, D. and Greenberg, R. Earcons and icons: Their structure and common design principles. Human Computer Interaction 4, 1 (1989), 11-44.
5. Bregman, A.S. Auditory Scene Analysis. MIT Press, Cambridge, Massachusetts, 1990.
6. Brewster, S.A. Overcoming the Lack of Screen Space on Mobile Computers. Accepted for publication in Personal and Ubiquitous Technologies 5 (2001).
7. Brewster, S.A. and Murray, R. Presenting dynamic information on mobile computers. Personal Technologies 4, 4 (2000), 209-212.
8. Brewster, S.A., Wright, P.C., Dix, A.J. and Edwards, A.D.N. The sonic enhancement of graphical buttons. In Proceedings of IFIP Interact'95 (Lillehammer, Norway) Chapman & Hall, 1995, pp. 43-48.
9. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. An evaluation of earcons for use in auditory human-computer interfaces. In Proceedings of ACM/IFIP INTERCHI'93 (Amsterdam, Holland) ACM Press, Addison-Wesley, 1993, pp. 222-227.
10. Walker, V.A. and Brewster, S.A. Trading Space for Time in Interface Design. In Volume II of the Proceedings of INTERACT '99 (Edinburgh, UK) British Computer Society, 1999, pp. 67-68.
11. Walker, V.A. and Brewster, S.A. Spatial audio in small screen device displays. Personal Technologies 4, 2 (2000), 144-154.