Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices

Stephen Brewster, †Joanna Lumsden, Marek Bell, Malcolm Hall and Stuart Tasker

Glasgow Interactive Systems Group,
Department of Computing Science
University of Glasgow, G12 8QQ, UK
+44 (0)141 330 4966
{stephen, marek, hall}@dcs.gla.ac.uk
www.dcs.gla.ac.uk/~stephen

†National Research Council
Institute for Information Technology - e-Business
46 Dineen Drive
Fredericton, New Brunswick
Canada E3B 9W4
jo.lumsden@nrc.gc.ca

ABSTRACT

Mobile and wearable computers present input/output problems due to limited screen space and interaction techniques. When mobile, users typically focus their visual attention on navigating their environment - making visually demanding interface designs hard to operate. This paper presents two multimodal interaction techniques designed to overcome these problems and allow truly mobile, 'eyes-free' device use. The first is a 3D audio radial pie menu that uses head gestures for selecting items. An evaluation of a range of different audio designs showed that egocentric sounds reduced task completion time, perceived annoyance, and allowed users to walk closer to their preferred walking speed. The second is a sonically enhanced 2D gesture recognition system for use on a belt-mounted PDA. An evaluation of the system with and without audio feedback showed users' gestures were more accurate when dynamically guided by audio-feedback. These novel interaction techniques demonstrate effective alternatives to visual-centric interface designs on mobile devices.

Keywords: Gestural interaction, wearable computing

INTRODUCTION

Mobile and wearable computers have been one of the major growth areas of computing in recent years. Compared to desktop systems these devices have restricted input and output capabilities that typically reduces their usability. With often very limited amounts of screen space, their visual displays can easily become cluttered with information and widgets. Input is limited, with small keyboards or simple handwriting recognition the norm. Speech-recognition is not always an ideal option, even if recognition rates in noisy environments can be further improved. With the imminent dramatic increase in network bandwidth available to mobile and wearable devices, and the consequent rise in the number of possible services, new interaction techniques are needed to access services whilst on the move.

CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.

The interface designs of most mobile and wearable computers are based heavily on those of desktop GUIs. These were originally designed for users sitting at a computer to which they could give their full visual attention. Users of current mobile and wearable devices are often in motion when they use their devices (e.g. making/receiving calls, reading/sending text messages, etc.). If they are interacting whilst walking, running or driving, they cannot easily devote all of their visual attention to the interface [3]; it must remain with the main task for safety. It is hard to design visual interfaces that work well under these circumstances. Much of the interface work on wearable computers tends to focus on visual displays, often presented through head-mounted graphical displays [1]. These can be obtrusive and hard to use in bright daylight, plus they occupy the users' visual attention [7].

Our aim is to create a system that uses as little of our users' visual attention as possible and to assess the effectiveness of such a system. Initial work has shown non-speech audio to be very effective at improving interaction on mobile devices [13, 15]. It allows users to keep their visual attention on navigating the world around them and allows information to be presented to their ears. The research described here builds on this to investigate multi-dimensional auditory and gestural techniques that would enable richer and more complex interactions with devices and services when mobile.

POTENTIAL SOLUTIONS

The solutions we are investigating use simulated 3D sound and multi-dimensional gestures. 3D sound allows a sound source to appear as if it is coming from anywhere in space around a listener [2]. We use standard head-related transfer function (HRTF) filtering (see [2] for details) implemented in many PC soundcards with head tracking to improve quality of localisation.

† This work was undertaken whilst Joanna Lumsden was employed by the University of Glasgow.

One of the key pieces of work our research is based on is Cohen and Ludwig's *Audio Windows* [5]. In this system, users wore a headphone-based 3D audio display with different areas in space around them mapped to different items. This is a powerful technique as it allows a rich, complex audio environment to be created. Users could point at items with a data glove to make selections. This is potentially very important for mobile interactions as no screen is required. No evaluation of this work was presented so it is not known how successful it was with users in real use. Savidis *et al.* [14] also used a non-visual 3D audio environment to allow blind users to interact with standard GUIs. Different menu items were mapped to different places around the user's head. In this case, users were seated and could point at audio menu items to make selections. Again, no evaluation of the system was presented. Neither of these examples was designed to be used when mobile but they have many potential advantages for mobile interactions.

Schmandt and colleagues at MIT have done work on 3D audio in a range of different applications. One, *Nomadic Radio,* used 3D sound on a mobile device [15]. It was a wearable audio personal messaging system that used non-speech and speech sounds to deliver information and messages to users on the move. Users wore a microphone and shoulder-mounted loudspeakers that provided a planar 3D audio environment. The advantage of the 3D audio presentation was that it allowed users to listen to multiple sound streams at the same time and still be able to distinguish and separate each one (the 'Cocktail party' effect). The spatial position of the sounds around the head also gave information about time of occurrence. We wanted to build on this to create a wider range of interaction techniques for a wider range of 3D audio applications.

Non-speech audio has been shown as effective in improving interaction and presenting information non-visually on mobiles. For example, Brewster [3] ran a series of experiments which showed that, with the addition of earcons, graphical buttons on the Palm III interface could be reduced in size but remain as usable as larger buttons when the device was used whilst walking. The sounds allowed users to keep their visual attention on navigating the world around them.

Our solution to input focuses on multi-dimensional gestural interaction. Input is difficult on mobiles as there is no space for a full keyboard and mouse. Many handheld devices require users to use a stylus to write characters on a touch screen. When mobile, this can be problematic; since both the device and the stylus are moving, the accurate positioning required can prove extremely difficult. It also demands the use of both hands. The 'Twiddler' [1], a small one-handed chord keyboard, is often used on wearables but it can be hard to use and requires learning of the chords.

There has been little use of physical hand and body gestures for input on the move. Such gestures are advantageous because users do not need to look at a display to interact with it (as they must when clicking a button on a screen). Although Harrison *et al.* [8] showed that simple, natural gestures can be used for input in a range of different situations

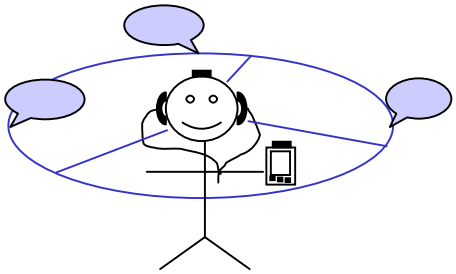

Figure 1: An illustration of the gesture-driven 3D audio wearable computer.

on mobile devices, they never tested the use of gestural input on the move.

Pirhonen *et al.* [13] looked at the effect of using non-speech audio feedback and gestures in an MP3 player on a Compaq iPAQ. Centred on the functions of the music player - such as play/stop, previous/next track - they designed a simple set of gestures that people could perform whilst in motion. Users generated the gestures by dragging their finger across the touch screen and received audio feedback upon completion of each gesture. Users did not need to look at the display of the player to be able use it. An experiment showed that the audio/gestural interface was significantly better than the standard, graphically based, media player on the iPAQ. They found that the audio feedback upon completion of the gestures was very important so that users knew what was going on; without such feedback, users performed gestures worse than when good audio feedback was provided.

Friedlander *et al.*[6] developed non-visual 'Bullseye' menus where the menu items ringed the user's cursor in a set of concentric circles divided into quadrants. Non-speech audio cues (played without spatialisation) indicated when the user moved across a menu item (using a simple beep). A static evaluation of Bullseye menus showed them to be an effective non-visual interaction technique; users were able to select items just using the sounds. One use the authors suggest for their menus is in mobile devices with limited screen space, which makes them very useful for the problems we are trying to solve. The two interaction techniques we propose in this paper draw on elements of their design for non-visual interaction when on the move.

We are developing a wearable device based around audio for output and gestures for input. An illustration of our system can be seen in Figure 1, with some of the actual components in Figure 2. The user wears a pair of lightweight head-

Figure 2: The wearable system used in the experiments.

headphones to hear the audio output (without obscuring real world sounds), which is spatialised in a plane around the user's head at the level of the ears (to achieve the best spatialisation with the largest group of listeners). Our 3D sounds are rendered by Microsoft's DirectX 8 API. An InterSense InterTrax II tracker is placed on the headphones to detect head orientation. This can then be used for the re-spatialisation of the sounds. It also allows us to use head gestures as an interaction technique: head movements such as nods or shakes can be used to make selections in the audio space. Head pointing is more common for desktop users with physical disabilities [11], but has many advantages for all users, as head movements are very expressive.

The wearable device itself (a Xybernaut MA V running Windows XP) sits on the user's belt. In Figure 1 the user is holding a PDA (a Compaq iPAQ which can also be belt mounted with a belt clip). As needed, it can be kept on the belt or removed and held, and is connected to the wearable via a cable or wireless network connection. Where information must be displayed visually, this serves as the screen of the wearable. Using a finger on the screen of the iPAQ users can make 2D gestures. A tracker can also be mounted on the PDA so that it too can be used for pointing or gesturing.

Our interface is based around Cohen's audio windows – each audio source has its own position in space around the user, in which interactions with the window occur. It might be used thus: (a) whilst walking, a user might nod towards an earcon which represents a phone call and which would appear in the front of the audio space (as dealing with this is likely to be a main activity); (b) if not considered important, the user might 'grab' the audio window of an MP3 file and 'drag' it to the rear of the audio space by pointing at it with the PDA or head; and (c) on the right hand side of the space, a sonification of some stock market data might play whilst on the left, a news feed from the stock market might be spoken via a speech synthesiser. The user can then attend to whichever of these is of interest via the Cocktail Party Effect. Such simple and natural gestures allow users to control the audio space without being overloaded with sound.

This paper describes two experiments performed to answer questions that have arisen when designing the system described. The first looks at using head movements as a selection mechanism for audio items presented in a 3D audio space, the second looks at audio feedback on 2D gestures made with a finger on a screen of a PDA.

GENERAL EXPERIMENTAL SET UP

Both experiments used a similar set up. Users had to walk 20m laps around obstacles set up in a room in the University - the aim being to test our system whilst users were mobile in a fairly realistic environment but maintain sufficient control so that measures could be taken to assess usability.

During the experiments, a full range of measures was taken to assess the usability of the interfaces tested. We measured time to complete tasks, error rates and subjective workload (using the NASA TLX [9] scales). Workload is important in a mobile context: since users must monitor and navigate their surroundings, fewer attentional resources can be devoted to the computer. An interface that reduces workload is therefore likely to be successful in a real mobile setting. We added an extra factor to the standard TLX test: annoyance. This was to allow us to test any potential annoyance caused by using sound in the interface.

To assess the impact of our device on the participants, we also recorded percentage preferred walking speed (PPWS) [12]: the further below their normal walking speed that users walked the more negative the effect of the device. Pirhonen *et al.* [13] found this to be a sensitive measure of the usability of a mobile MP3 player, with an audio/gestural interface affecting walking speed less than the standard graphical one. Prior to the start of each experiment, participants walked a set number of laps of the room; their lap times were recorded and averaged so that we could calculate their standard PWS when not interacting with our device.

The final measure taken was comfort. This was based around a new scale developed by Knight *et al.* [10] called the Comfort Rating Scale (CRS) which assesses various aspects to do with the comfort of a wearable device. For a device to be accepted and used it needs to be comfortable and people need to be happy to wear it. Using a range of 20-point rating scales similar to NASA TLX, CRS breaks comfort into 6 categories: emotion, attachment, harm, perceived change, movement and anxiety. Knight *et al.* have used it to assess the comfort of two wearable devices they are building in their research group. Using this will allow us to find out more about the actual acceptability our systems.

HEAD GESTURE EXPERIMENT

To enable users to select, control and configure applications, our 3D audio wearable device needed some mechanism for choosing items from menus or lists. We have developed 3D auditory radial pie menus to allow this. The user's head is in the middle of the pie (or Bullseye) with sounds or speech for the menu items presented in a plane around the head (see Figure 1). Nod gestures in the directions of the sounds allow the items corresponding to the sounds to be chosen (in a similar way to Cohen's audio windows). An experiment was needed to find out if nodding was an effective interaction technique when on the move and what design of sounds would be most beneficial. To do this we generated 3D sounds from the MA V and used the tracker on the headphones to generate the angles for recognising nods.

Head gesture recognition

A simple 'nod' recogniser was built to allow us to recognise selections. Since the recogniser has to be robust to noise in the data coming from the movements of the user walking, much trial, error and iterative testing was used to generate the actual values used in our algorithms. The recogniser works as follows for forward nods.

The main loop for detection runs every 200ms. If there is a pitch change > 7° then this signifies the head is moving forward (this avoids small movements of the head which are not nods). For example, if the head started at 5° and then moved to 15° then a nod has started. Allowing for differ-

ences in users' posture, the algorithm needed to be flexible about its start point and so this allows the nod to start wherever the user wants. If the user then moves his/her head back >= 7^{o} within 600ms a nod is detected; outside this timeframe, the nod times out (the person may just have his/her head down looking at the ground and not be nodding. It also gives users a chance to 'back out' if they decide they want to choose nothing). The same method works for all directions, but using roll for left and right nods. This method is very simple but is fairly robust to the noise of most small, normal head movements, movements due to walking and gross individual differences in nodding.

Soundscape Design

For our experimental application we needed a simple audio environment for users to work with so that we could test the interaction techniques. We chose current affairs information - four menu items were presented: Weather, News, Sport and Traffic. The scenario was that a user wearing the device might want information about one or more of these when out and in motion. Simple Auditory Icons were used for each of the items:

- *Weather:* A mix of various rain, lightening and bird samples;
- *News:* A clip taken from the UK Channel 5 TV News theme tune;
- *Sport:* A clip taken from the UK TV "Question of Sport" theme tune;
- *Traffic:* A mix of various busy street samples, including cars, trucks, engines, horns and skids.

Three soundscapes were designed, giving three conditions to the experiment. These looked at different placements of the sounds in the audio space and whether the space was ego- or exocentric. The design of these soundscapes was based on some initial pilot studies, which showed that egocentric was the most effective, but users complained of neck strain when nodding backwards. They were:

1. Egocentric: Sounds are placed at the four cardinal points (every 90° from the user's nose). The sounds are egocentric, so when turning the sounds remain fixed with respect to the head. The sound items play for two seconds each and play in order rotating clockwise around the head. This is a simple design but involves many backwards nods that are hard on the neck muscles. With this method it is also hard to have more than 4 items in the soundscape as nodding accurately at 45° in the rear hemisphere is difficult.

2. Exocentric, constant: This interface has the four sounds arranged in a line in front of the head. The user can select any one of the items by rotating the head slightly until directly facing the desired sound, and then nodding. All nods are basically forward nods, which are much easier to do, can be done more accurately and are the most natural for pointing at, or selecting items. Clicks are played as the head rotates through the sound segments (each segment is 40°) and a thump is played when the last segment on each side is passed (to let the user know that the last sound has been reached). All sounds are played simultaneously; the sound currently in front of the head is, however, played slightly louder than the rest to indicate it is in focus. If the user physically turns then the sounds will no longer be in front, but can be reset to the front again by nodding backwards. This is a more complex design than (1) but involves much less backward nodding. The sounds get their information across more quickly (as they are all playing simultaneously) but the soundscape may become overloaded.

3. Exocentric, periodic: This interface is exactly the same as (2), except sounds are played one after another in a fixed order from left to right, similar to (1). This means there are fewer sounds playing so that the soundscape is less crowded but it may take longer to select items because the user may have to wait for a sound to play to know where to nod.

Experimental design and procedure

An experiment was conducted to assess whether 3D audio menus would be a usable method of selection in a wearable computer when the user was in motion, and to investigate what arrangement of sounds would be the most successful.

A fully counterbalanced, within-groups design was used with each participant using the three interface designs whilst walking. Brief training was given before each condition. Forty menu selections were required in each condition, 10 for each menu item. Synthetic speech was used to tell the user the next selection to be made - for example "Now choose weather" – and the required selections were presented in a random order. No feedback was given on the correctness of the gestures made. Eighteen people participated: 13 males and 5 females with ages ranging from 18-55. In addition to the measures described above, we also collected the number of incorrect selections made and distance walked.

The main hypothesis was that nodding would be an effective interaction technique when on the move. The second hypothesis was that soundscape design would have a significant effect on usability. Egocentric selection of items should be faster than exocentric. With egocentric presentation the user needs just to nod at the chosen object, with exocentric the user must first turn to the sound and then nod.

Condition	*Average Overall Time (secs.)*
Egocentric	127.7
Exocentric, constant	270.8
Exocentric, periodic	337.5

Table 1: Mean time taken per condition when using the audio pie menus.

Results

A single factor ANOVA showed total time taken was significantly affected by soundscape type ($F_{2,51}$=14.24, $p<0.001$), see Table 1. *Post hoc* Tukey HSD tests showed that Egocentric was significantly faster than both of the other conditions ($p<0.05$), but there were no significant differences between the two Exocentric conditions. Soundscape type also affected the total distance walked; people walked significantly fewer laps in the Egocentric condition ($F_{2,51}$=5.23, $p=0.008$) because they completed the selections

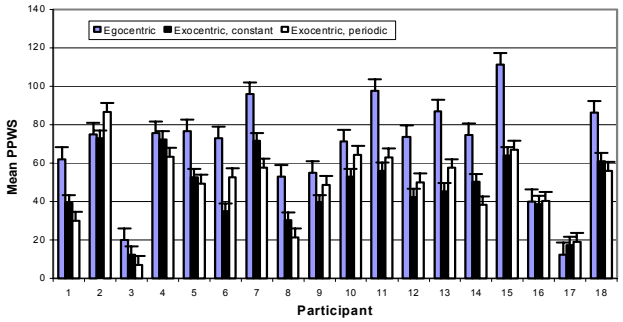

Figure 3: Mean Percentage Preferred Walking Speed results. Standard error bars are shown.

more quickly. Distances walked ranged from 50m in the Egocentric condition to 90m in the Exocentric periodic.

There were no significant differences in overall workload. Only annoyance was significantly effected ($F_{2,51}$=3.29, $p<0.05$). Tukey HSD tests showed Exocentric periodic was significantly worse than Egocentric ($p<0.05$) but no other differences were significant.

An analysis of PPWS (Figure 3) showed significant results ($F_{2,51}$=5.88, $p=0.005$). Tukey HSD tests showed that the Egocentric interface affected walking speed significantly less than either of the two exocentric ones ($p<0.05$), but there were no significant differences between the two Exocentrics. The mean score in the egocentric condition was 69.0% of PPWS, with 47.5% and 48.5% for exocentric constant and periodic respectively.

There were no significant differences in the number of incorrect nods in each condition (with a range of 8.1 – 8.3 input errors out of 40 per condition).

Discussion

The results showed that nod gestures were effective on the move and the Egocentric display was generally the best with participants walking at around 70% of their normal speed (which would be expected to improve with practice) plus significant reductions in time taken for the experiment and annoyance. There were few differences between the Exocentric conditions.

There was a lot of variance in PPWS. Some users found the wearable easy to use, whilst others slowed down dramatically. As can be seen from Figure 3, participant 15 walked faster than his normal walking speed with the egocentric display. Participants 3 and 17 had problems and walked much slower than their normal speed in all conditions. Participant 3 found the distance needed for the experiment hard work and slowed down even after the initial assessment of PWS. It is important that the device is kept light for such a user. Participant 17 stopped a lot whilst selecting items, finding it hard to walk and select simultaneously. We will investigate the issues these users had with our system in the next stage of our work as we are keen that our device is usable by as many people as possible.

As mentioned previously, the design of the Egocentric display encounters problems if more than four items are needed in a menu. A further experiment is needed to assess how many items a user could deal with in a soundscape. It may be that four is a maximum given that the user has to handle the complexities of navigating the environment and listening to sounds from it. Savidis *et al.* [14] suggest that, in informal studies with seated users, 6 items placed around the user caused problems. If more than four items is possible for a user then the exocentric interface designs become more useful. It is likely that any more than 8 items in the plane around a user's head would be very difficult to deal with because of the non-individualised HRTFs we are using; users would have problems accurately locating the sounds in space in order to be able to nod in the correct direction.

The results suggest that the sounds should be played simultaneously for faster performance, although this may not be true with a larger number of items in the soundscape. Further study is needed to investigate this issue.

The simple nod recogniser returned an error rate of around 20%. Some errors occurred because the recogniser mistook a nod, others were not really errors, e.g. a user simply nodded at the wrong item. Our recogniser was very simple and we are currently working on a more sophisticated one that will be more robust and handle a wider range of gestures.

The design of our menus could be extended to allow hierarchical menu structures. If, as suggested above, it is difficult to have many menu items at one time, hierarchical menus will be needed (similar to hierarchical pie menus). A nod at one item could take the user into a submenu and a backward nod could be used to return to the previous level. To ensure users knew where they were in such a structure (as there is no visual display) hierarchical earcons could be used to indicate position [4]. Care must be taken when designing such earcons so that they do no conflict with the sounds for the menu items. A mix of auditory icons for menu items and earcons for navigation would help this separation.

HAND GESTURE EXPERIMENT

Pirhonen *et al.* [13] looked at the use of metaphorical gestures to control an MP3 player. For example, a 'next track' gesture was a sweep of a finger across the screen left to right and a 'volume up' gesture was a sweep up the screen, bottom to top. Results showed that these were an effective way of interacting and more usable than the standard interface to an MP3 player. Pirhonen *et al.* demonstrated increased usability when gestures were supported by end-of-gesture audio feedback; we have taken this further to investigate the use of dynamic auditory feedback during the progress of the gestures in order to assess its affect on the *accuracy* of gestures. Like Pirhonen *et al.,* it was not our intention to develop a handwriting recognition system (as it is very hard to handwrite when on the move) and so we also concentrated on metaphorical gestures that could be used for a range of generic operations on a wearable device. For the purpose of our investigation, we focussed on a combination of 12 single- and multiple- stroke alphanumeric and geometric gestures (/, \, -, |, N, circle, +, S, ↑, ↓, X and Z, encompassing those used by Pirhonen) which might potentially be used to control mobile applications.

Hand Gesture Recognition

A hand gesture recogniser has been developed to allow a user to draw, simply using his/her finger, 2D gestures on the screen of an iPAQ without any need to look at the iPAQ's display. The recogniser is generic in that it can be used to recognise any gesture which is predefined as valid.

The recogniser is based around a conceptual 3 x 3 grid (see Figure 4a) overlaid upon the touch-screen of the iPAQ (we used a square layout rather than the circles of Friedlander's Bullseye system as it was a better fit with the shape of the iPAQ screen); derived from a publicly available algorithm [16], the co-ordinate pairs that are traversed during a given gesture are condensed into a path comprising the equivalent sequence of grid square ('bin') numbers. This resolution strikes a balance between that required for most application gestures and our desire for genericity and simplicity.

To accommodate gestures comprising two or more discrete strokes, the recogniser pauses for 0.5sec between finger-up and finger-down actions before recording a gesture. If, during this time, the user begins to draw again, the current stroke is appended to the previous stroke(s) to form a compound gesture; outside this timeframe, the completed gesture is recorded as such and a system-level beep is played to inform the user that the gesture has been registered and that the system is ready to accept further gestures. At any time, a user can abort a gesture by double-tapping the iPAQ screen.

Sound design

Sounds were designed to represent the 3x3 matrix. Our sounds were used to dynamically guide users correctly through gestures, rather than Friedlander *et al.'s* where a single beep represented all menu items so navigation was based on counting. Our design was based on the C-major chord; the sounds used are shown in Figure 4b. Hence, the sounds increase in pitch in accordance with the notes in the C-major chord from left to right across each row and increase by an octave from bottom to top across the bins in each column. A sweep left to right across a row would therefore generate the notes C_x E_x G_x (where x corresponds to the octave for the selected row). On the basis of the above design and the assumption that, in order to be differentiable, no two gestures can be defined by the same bin-path, each gesture has a distinct audio signature. It was anticipated that users would learn or become familiar with these audio signatures to the extent that they would recognised them when heard. Two implementations of this design were developed:

1	2	3
4	5	6
7	8	9

C_6	E_6	G_6
C_5	E_5	G_5
C_4	E_4	G_4

Figure 4: (a) The 3 x 3 grid used, (b) Sounds used in gesture recogniser.

1. Simple Audio: This implementation simply plays the note corresponding to the bin in which the user's finger is currently located. For example, if the user's finger is currently within the boundaries of Bin 1, then C_6 will be played. This note will sound continuously until the user moves his/her finger into another bin (in which case the note played will change to that corresponding to the new bin location) or until the user lifts the finger from the iPAQ screen.

2. Complex Audio: This implementation extends the Simple Audio design by providing users with pre-emptive information about the direction of movement of their finger in terms of the bin(s) they are approaching and into which they might move. For example, if the user is drawing towards the bottom edge of Bin 1, he/she will simultaneously hear C_6 corresponding to that bin and, at a lesser intensity, C_5 corresponding to Bin 4. Similarly, if the user draws further towards the bottom right-hand corner of the same bin, he/she will additionally hear E_5 and E_6 reflecting the multiple options for bin change currently available. It was hoped that, by confirming location and direction of movement, this information would allow users to pre-emptively avoid unintentionally slipping into incorrect bins for any given gesture and thus improve accuracy.

Experimental Design and Procedure

An experiment was conducted to see whether presenting dynamic auditory feedback for gestures as they progressed would, in particular for use in motion, improve users' gesturing accuracy (and thereby the usability and effectiveness of the recogniser) and to compare the two sound designs.

The experiment used the same basic setup as before. This time, however, a Compaq iPAQ was used as the input device with participants drawing gestures on the screen with a finger. The iPAQ was mounted on the user's waist on the belt containing the MA V wearable and was used to control the MA V using the Pebbles software from CMU. The sounds were not presented in 3D in this case.

A fully counterbalanced, between-groups design was adopted with each participant using – whilst walking (as described) – the recogniser minus all audio feedback (excepting the system level beep) and one of the audio designs. Participants were allowed to familiarise themselves with the recogniser for use under each condition but no training was provided. They were required to complete 4 gestures per lap and to complete 30 laps in total under each condition (hence 120 gestures – 10 each of 12 gesture types – were generated per participant per condition). Gestures were presented to users on a flip chart located adjacent to the circuit they were navigating. Participants were not required to complete a gesture correctly before moving onto the next gesture since we wanted to assess participants' awareness of the correctness of their gestures. Twenty people participated (10 per experimental group): 13 males and 7 females all of whom were right-handed and none participated in Experiment 1. In addition to the measures previously discussed, we also collected information on the paths drawn by each participant and the number of gestures they voluntarily aborted.

The main hypotheses were that users would generate more accurate gestures under the audio conditions and, as a result of better awareness of the progression of their gestures, would abort more incorrect gestures. As a consequence of the increased cognitive load, it was also hypothesised that the audio conditions would have a greater detrimental affect

on participants' PWS than the non-audio condition. Since both audio designs were previously untried, we made no hypothesis as to which would return better results.

Results and Discussion

A two factor ANOVA showed that the accuracy of gestures was significantly affected by audio condition ($F_{1,36}$=17.93, $p<0.05$). Tukey HSD tests showed that participants within the simple audio group generated significantly more accurate gestures under the audio condition than under the non-audio condition (p=0.012) and that participants within the complex audio group generated significantly more accurate gestures under the audio condition than under the non-audio condition (p=0.046). There were no significant differences between the results for the two audio designs. Figure 5 shows the average accuracy rates achieved per condition according to experimental group.

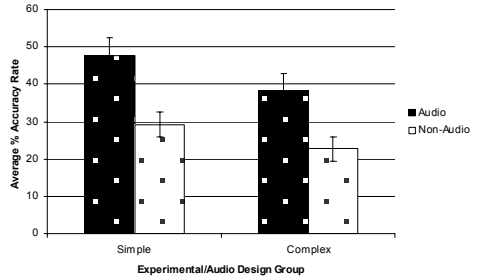

Figure 5: Mean percentage accuracy rates achieved for the hand gestures.

A two factor ANOVA showed that the number of gestures aborted by participants was significantly affected by audio condition ($F_{1,36}$=3.97, p=0.05). Tukey HSD tests revealed that participants in the complex audio group aborted significantly more gestures when under the audio condition than under the non-audio condition (p=0.04) and that there were significantly more aborted gestures from the participants in this group under the audio condition than from the participants in the simple audio group (p=0.05). Figure 6 shows the average number of aborted gestures according to experimental group and condition.

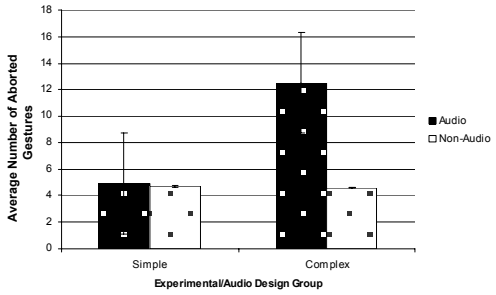

Figure 6: Mean number of aborted hand gestures.

The first of these results confirms the initial part of our main hypothesis: that audio-enhanced gesturing increases the accuracy of gestures when used 'eyes-free' and in motion. It is, however, more difficult to interpret the latter results. Although the complex audio condition returned a significantly higher number of aborted gestures, this was not reflected in a significantly higher accuracy rate for this condition compared to the simple audio condition. It is, therefore, unlikely that the participants under this condition were aborting more gestures as a result of a heightened awareness of mistakes they were making whilst gesturing. Instead, although only at the level of conjecture, it is more likely that the complex audio design confused participants. Further evaluation would be required to confirm or counter this observation.

Participants reported no significant differences in the overall workload experienced under any of the conditions, nor was any condition significantly more popular than the others.

We had hypothesised that, as a result of increased levels of feedback, the audio designs would increase participants' cognitive load to the extent that it would be reflected in significantly slower walking speeds under the two audio conditions. This was not found to be the case. Although under all conditions participants' walking speeds were slower when performing the experimental tasks (speeds ranged from 94.7% to 32.8% of PPWS), a two factor ANOVA showed no significant affect of audio condition on PPWS.

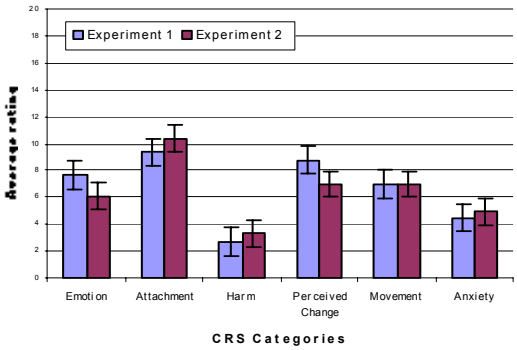

Figure 7: Comfort ratings for the two experiments.

GENERAL DISCUSSION

Overall the two experiments have shown that gestures and sound can improve the usability of wearable devices.

The comfort results for both experiments are shown in Figure 7. Despite the difference in focus of physical movement between the two experiments, the comfort ratings returned for each experiment were not significantly different. Like the NASA TLX, low ratings are desirable; of the six categories, the 'Attachment' of our device is shown to be the biggest obstacle to comfort (ratings of 9.3 and 10.4 for Experiments 1 and 2 respectively). This category is concerned with subjective awareness of the device when attached to the body. The MA V is relatively bulky (455g) and, since it is worn on a belt, users can feel its weight in a localised manner. In Experiment 2 participants also had an iPAQ attached to the belt, making extra weight. The pressure of the headphones against the user's head further add to the feeling of attachment. It is interesting to note that, despite wearing the device (with the added weight) for longer in Experiment 2 than in Experiment 1 (in the former each participant walked over 1.3km in total), participants did not appear to be significantly more aware of the device and its associated weight and fit during the course of Experiment 2.

It is interesting to note that walking speed was slower with head than with hand gestures (which had no significant effect on walking speed). Perhaps this is unsurprising as nodding may make it harder for users to look where they are going. Our more sophisticated recogniser will allow us to recognise smaller head gestures more reliably which may reduce this problem and its effects on walking speed.

Areas to investigate to try and lessen users' awareness of the device would include the style of headphones used, the manner and location in which the device is physically attached to the body and activity-specific requirements. One advantage of our device over ones using visual head mounted displays is that many people currently wear headphones (for music players, radios or mobile phones) making ours stand out less, lowering our Anxiety scores. A further, longer-term study is needed to see if people would use our interaction techniques in real situations. Even though the CRS ratings are good, nodding might well be unacceptable in public unless we can make the nods required very small. This will be a focus of a future investigation.

CONCLUSIONS

This paper has shown that novel interaction techniques based on sound and gesture can significantly improve the usability of a wearable device, in particular under 'eyes-free', mobile conditions. When rated by participants, our wearable device was considered comfortable and is likely to be acceptable to users.

Head gestures have been shown to be a successful interaction technique with egocentric sounds the most effective. This design had significantly less impact on walking speed than the others tried.

The accuracy of 'eyes-free' hand gestures has been shown to be significantly improved with the introduction of dynamic audio feedback; initial results would suggest that the simpler the audio design for this feedback the better, to avoid overloading the users' auditory and cognitive capacity. This improvement in accuracy is not at the expense of walking speed and results would suggest that there is potential for substantial recognition and recall of the audio signatures for gestures.

We have shown that non-visual interaction techniques can be used effectively on wearable computers in mobile contexts. These techniques wholly avoid visual displays, which can be hard to use when mobile due to the requirements of the environment through which the user is moving. Designers of mobile and wearable devices now have two new techniques available to them to make their devices more effective for their users when they are on the move.

ACKNOWLEDGMENTS

This work was funded by EPSRC grant GR/R98105 and ONCE, Spain. This paper is in memory of Stuart Tasker who very sadly died before it was published.

REFERENCES

1. Barfield, W. and Caudell, T. (eds.). Fundamentals of wearable computers and augmented reality. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2001.
2. Begault, D.R. 3-D sound for virtual reality and multimedia. Academic Press, Cambridge, MA, 1994.
3. Brewster, S.A. Overcoming the Lack of Screen Space on Mobile Computers. Personal and Ubiquitous Computing, 6 (3). 188-205.
4. Brewster, S.A. Using Non-Speech Sounds to Provide Navigation Cues. ACM Transactions on Computer-Human Interaction, 5 (3). 224-259.
5. Cohen, M. and Ludwig, L.F. Multidimensional audio window management. International Journal of Man-Machine Studies, 34. 319-336.
6. Fiedlander, N., Schlueter, K. and Mantei, M., Bullseye! When Fitt's law doesn't fit. in Proceedings of ACM CHI'98, (Los Angeles, CA, 1998), ACM Press Addison-Wesley, 257-264.
7. Geelhoed, E., Falahee, M. and Latham, K. Safety and comfort of eyeglass displays. in Thomas, P. and Gellersen, H.W. eds. Handheld and Ubiquitous Computing, Springer, Berlin, 2000, 236-247.
8. Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C. and Want, R., Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. in Proceedings of ACM CHI'98, (Los Angeles, CA, 1998), ACM Press Addison-Wesley, 17-24.
9. Hart, S.G. and Wickens, C. Workload assessment and prediction. in Booher, H.R. ed. MANPRINT, an approach to systems integration, Van Nostrand Reinhold, New York, 1990, 257-296.
10. Knight, J. and Baber, C., Physical load and wearable computers. in Proceedings of the 5th International Symposium Wearable Computers, (Atlanta, GA, 2000), IEEE Computer Society.
11. Malkewitz, R., Head pointing and speech control as a hands-free interface to desktop computing. in Proceedings of ACM ASSETS 98, (Marina del Rey, CA, 1998), ACM Press, 182-188.
12. Petrie, H., Furner, S. and Strothotte, T., Design Lifecycles and Wearable Computers for Users with Disabilities. in First workshop on human-computer interaction with mobile devices, (Glasgow, UK, 1998), Glasgow University.
13. Pirhonen, A., Brewster, S.A. and Holguin, C., Gestural and Audio Metaphors as a Means of Control for Mobile Devices. in Proceedings of ACM CHI 2002, (Minneapolis, MN, 2002), ACM Press, 291-298.
14. Savidis, A., Stephanidis, C., Korte, A., Crispien, K. and Fellbaum, C., A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction. in Proceedings of ACM ASSETS'96, (Vancouver, Canada, 1996), ACM Press, 117-123.
15. Sawhney, N. and Schmandt, C. Nomadic Radio: speech and audio interaction for contextual messaging in nomadic environments. ACM Transactions on Human-Computer Interaction, 7 (3). 353-383.