A Paradigm Shift: Alternative Interaction Techniques for Use with Mobile & Wearable Devices

Joanna Lumsden
NRC - IIT e-Business
46 Dineen Drive, Fredericton
Canada E3B 9W4
e-mail: jo.lumsden@nrc.gc.ca

Stephen Brewster
Department of Computing Science
University of Glasgow, Glasgow
U.K., G12 8RZ
e-mail: stephen@dcs.gla.ac.uk
web: www.dcs.gla.ac.uk/~stephen

Abstract

Desktop user interface design originates from the fact that users are stationary and can devote all of their visual resource to the application with which they are interacting. In contrast, users of mobile and wearable devices are typically in motion whilst using their device which means that they cannot devote all or any of their visual resource to interaction with the mobile application – it must remain with the primary task, often for safety reasons. Additionally, such devices have limited screen real estate and traditional input and output capabilities are generally restricted. Consequently, if we are to develop effective applications for use on mobile or wearable technology, we must embrace a paradigm shift with respect to the interaction techniques we employ for communication with such devices.

This paper discusses why it is necessary to embrace a paradigm shift in terms of interaction techniques for mobile technology and presents two novel multimodal interaction techniques which are effective alternatives to traditional, visual-centric interface designs on mobile devices as empirical examples of the potential to achieve this shift.

1 Introduction

Desktop user interface design has evolved on the basis that users are stationary – that is, sitting at a desk – and can normally devote all (or most) of their visual resource to the application with which they are interacting. The interfaces to desktop-based applications are typically very graphical, often extremely detailed, and utilise the standard mouse and keyboard as interaction mechanisms.

Contrast this with mobile and wearable devices. Users of these technologies are typically in motion whilst using their device. This means that they cannot devote all or any of their visual resource to interacting with the mobile device – it must remain with the primary task (e.g. walking or navigating the environment), often for safety reasons [6]. Additionally, in comparison to desktop systems, mobile and wearable devices have limited screen real estate, and traditional input and output capabilities are generally restricted – keyboards or simple handwriting recognition is the norm.

It is hard to design purely graphical or visual interfaces that work well under these mobile circumstances. Despite this, however, the interfaces and associated interaction techniques of most mobile and wearable computers are based on those of desktop GUIs. Consequently, much of the interface work on wearable computers tends to focus on visual displays, often presented through head-mounted graphical displays [2]. These can be obtrusive and hard to use in bright daylight, plus they occupy the users' visual attention [14].

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 effectively and safely access services whilst on the move. That is, we need to embrace a paradigm shift in terms of the interaction techniques harnessed to enable interaction with mobile and wearable devices. No longer can we, nor should we, rely on the mouse and keyboard as mechanisms of interaction.

1.1 Contextual Concerns

Unlike the design of interaction techniques for standard desktop applications, the design of interaction techniques for use with mobile and wearable systems has to address complex contextual concerns: failure to acknowledge and adequately respond to these concerns is likely to render the techniques inappropriate and/or useless. So what contextual factors are of concern ?

The constituent factors that together form the context of use for mobile and wearable applications is a matter of current debate, as indeed is the notion of context-awareness (e.g. [12, 20, 21]). It is not, however, the intention of this paper to examine the current arguments presented in the research field of context-aware computing. Instead, its aim is to briefly highlight the general areas of concern that impinge upon the design of appropriate interaction techniques for use with mobile and wearable devices – that is, to demonstrate the factors that underlie the need for a paradigm shift in the design of such interaction techniques.

In the first instance, the interaction design must cater to the user's need to be able to safely navigate through his/her environment whilst interacting with the mobile application. This is likely to necessitate interaction techniques that are 'eyes-free' or even 'hands-free'. Such interaction techniques need to be sufficiently robust as to accommodate the imprecision inherent in performing a task whilst walking, for example, and/or to provide appropriate feedback as to alert users to the progress of their interaction in order that they can explicitly adjust their actions to compensate.

More so than for desktop applications, the design of interaction techniques for use with mobile technology has to take into consideration the social context in which the techniques are to be employed. For instance, what gestural interaction is socially acceptable ? To what extent is speech-based interaction appropriate ?

Since mobile applications are typically designed to be used in motion, the physical context in which they are being employed is constantly changing. This includes changes in ambient temperatures, noise levels, lighting levels, and privacy implications to name but a few. Such environmental dynamism is a primary concern for context-aware computing, but equally, these factors impinge upon the applicability of design decisions when generating alternative techniques for mobile interaction and should therefore be a seminal factor in the design process.

Finally, users' interaction needs relative to mobile technology will differ greatly depending on the task context – that is, any given task might require different interaction techniques depending on the context in which the task is being performed. The real power of the next generation – or new paradigm – of interaction techniques will only be fully harnessed when the above contextual factors are taken into consideration and interaction techniques are designed to combine appropriate human senses (e.g. hearing, sight, touch etc).

The remainder of this paper focuses on two multimodal interaction techniques we designed (as part of an 'eyes-free' wearable system [9] and associated ongoing investigation into non-traditional interaction techniques for mobile technology) to overcome both the limitations placed on input and output with mobile and wearable devices and the current dependency on visual display (inherited from the desktop paradigm) that is prevalent amongst applications on such devices. The results of evaluating these techniques serve as empirical evidence of the potential for new paradigms to successfully address interaction issues with mobile technology; in particular, truly mobile 'eyes-free' device use. They also highlight areas on which to focus for future development of alternative interaction techniques. The first is a 3D audio radial pie menu that uses head gestures for selecting items. The second is a sonically enhanced 2D gesture recogniser for use on a belt mounted PDA. It should be noted, however, that these are only two *examples* of what could be achieved if we embrace a new interaction paradigm more suited to mobile and wearable device use.

2 Background

Our aim is to investigate interaction techniques which allow a user to communicate with mobile technology using as little visual attention as possible and to assess the effectiveness of such paradigms. Non-speech audio has proven to be very effective at improving interaction on mobile devices [23, 25]; by presenting information to their ears, it allows users to maintain their visual focus on navigating the world around them. The research described in the remainder of this paper builds on this to investigate the potential of multi-dimensional auditory and gestural techniques as alternative interaction paradigms able to support effective and accurate interaction with devices and services whilst mobile.

The solutions we are investigating use a combination of 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 [3]. We use standard head-related transfer function (HRTF) filtering (see [3] for details) implemented in many PC soundcards with head tracking to improve quality of localisation.

One of the seminal pieces of work upon which our research is based is Cohen and Ludwigs' *Audio Windows* [11]. In this system, users wear a headphone-based 3D audio display with different areas in space mapped to different items. This technique is powerful as it allows a rich, complex audio environment to be established; wearing a data glove, users can point at items to make selections. This is potentially very important for mobile interactions since no visual display is required. Unfortunately, no evaluation of this work has been presented so its success with users in real use is not known. For blind users, Savidis *et al* [24] also used a non-visual 3D audio environment to facilitate interaction with standard GUIs. In this case, different menu items are mapped to different locations in the space around the user's head; users are seated and can point to audio menu items to make selections. As with the Audio Windows, no evaluation of this work has been presented. Although neither of these examples was designed to be used when mobile, they have many potential advantages for mobile interactions.

Schmandt and colleagues at MIT have investigated 3D audio use in a range of applications.

Nomadic Radio, one such application, uses 3D sound on a mobile device [25]. This is a wearable personal messaging system that, via speech and non-speech sounds, delivers information and messages to users on the move. Users wear a microphone and shoulder-mounted loud speakers that provide a planar 3D audio environment. In accordance with the 'Cocktail Party Effect' [1], the 3D audio presentation allows users to listen to multiple sound streams consecutively whilst still being able to distinguish and separate each one. The spatial position of the sounds around the head also gives information about the time of occurrence. We wanted to build on this to extend the paradigm of mobile interaction by creating a wider range of interaction techniques for a wider range of 3D audio applications.

Non-speech audio has been shown to be effective in improving interaction and presenting information non-visually on mobile devices [5, 7, 8, 10, 18]. For example, Brewster [6] 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 large buttons when the device was used whilst walking; the sounds allowed users to keep their visual attention on navigating the world around them.

In terms of input, we focus on multi-dimensional gestural interaction. The design of input for mobile devices, perhaps even more so than output, requires a substantial paradigm shift given the contextually-dependent potential inappropriateness of 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 stylus are moving, the accurate positioning required can prove extremely difficult. Such interaction also demands the use of both hands which is not always possible or appropriate. The 'Twiddler' [2], a small one-handed chord keyboard, is often used on wearables but it can be hard to use and requires learning of the chords.

Little use has thus far been made of physical hand and body gestures for input on the move. Such gestures are advantageous because users do not need to look at the display to interact with it (as they are required to do when clicking a button on a screen for example). Although Harrison *et al.* [15] showed that simple, natural gestures can be used for input in a range of different situations on mobile devices, they did not test the use of gestural input on the move.

Pirhonen *et al.* [23] investigated the combined use of non-speech audio feedback and gestures for controlling an MP3 player on a Compaq iPAQ. Centred on the primary functions of the player – such as play/stop, previous/next track etc – they designed a simple set of gestures that people could perform whilst walking. To generate the gestures, users drag their finger across the touch screen of the iPAQ and, upon completion of each gesture, receive audio feedback. Users do not need to look at the display of the player to be able to use it. An experimental study of the use of the player showed that the audio/gestural interface is significantly better than the standard, graphically based, media player on the iPAQ. They found that the audio feedback on completion of each gesture is a very important factor in users' cognition of what is going on; without such feedback, users perform gestures worse than when good audio feedback is provided.

Friedlander *et al.* [13] developed non-visual 'Bullseye' menus where the menu items ring the user's cursor in a set of concentric circles divided into quadrants. Using a simple beep – played without spatialisation – non-speech audio cues are used to indicate when the user moves across a menu item. When statically evaluated, Bullseye menus were shown to be an effective non-visual interaction technique; users were able to select items using just the sounds. The authors suggest that their menus could be used in mobile devices with limited screen real estate, making them really useful for the problems we are trying to solve. The two interaction techniques we highlight in this paper draw on elements of their design for non-visual, mobile interaction.

3 Investigative Method

As previously mentioned, our aim is to investigate interaction techniques which allow a user to communicate, whilst in motion, with mobile technology using as little visual attention as possible and to assess the effectiveness of such paradigms. In particular, our investigation focuses on the ability of new interaction paradigms based around multidimensional audio for output and multidimensional gestures for input to support effective communication with mobile devices.

This paper describes two experiments performed as part of our investigation: the first looks at 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 the screen of a PDA.

An illustration of the hardware set up we used is shown in Figure 1. The user wears a pair of lightweight headphones to hear the audio output (without obscuring real world sounds). An InterSense InterTrax II tracker is placed on the headphones to detect head orientation. This can then be used for the re-spatialisation of 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 relative to the audio space. Head pointing is more common for desktop users with physical disabilities [19], but has many potential advantages for all users, as head gestures are naturally very expressive.

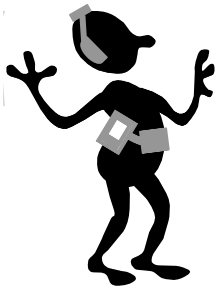

Figure 1: An illustration of our hardware set up: a wearable PC is attached to the user's waist, as is a PDA; a pair of headphones with a head tracker attached is on the user's head.

The wearable device itself (a Xybernaut MA V running Windows XP) sits on the user's belt. Additionally, as shown in Figure 1, the user has a PDA (in this case, a Compaq iPAQ) attached to the belt via a clip. The PDA is connected to the wearable via a cable or wireless connection. Using a finger on the screen of the iPAQ, users can make 2D gestures. A tracker could also be mounted on the PDA so that it too could be used for 3D gestures but that was outside the scope of this research. Although not within the concern of this investigation, the PDA could be removed from the belt and serve as the screen of the wearable should the need arise to present information visually rather than audibly.

3.1 Head Gestures

To enable users to select, control and configure mobile applications, there needs to be an interaction paradigm that supports (or is suited to) item choice from menus or lists. We therefore developed 3D audio radial pie menus as a vehicle to test the ability and suitability of 3D head gestures to meet this interaction need.

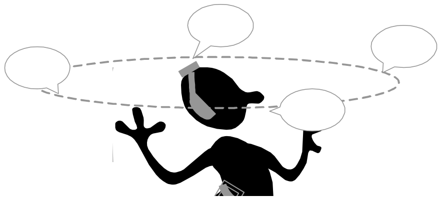

Figure 2: Multiple sound sources are presented in space around the listener.

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 user's head (see Figure 2) at the level of the ears (to achieve the best spatialisation for the largest group of listeners). 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). The following sections outline the nod recogniser and soundscape designs implemented to support the above.

3.1.1 Head Gesture Recognition

A simple 'nod recogniser' was built to allow us to recognise selections. Since the recogniser has to be sufficiently robust to accommodate and deal with head movements from the user walking, much 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 of more than 7^o, then this signifies the head is moving forward (avoiding small movements of the head which are not nods). For example, if the head started at 5^o (from vertical) and then moved to 15^o, then a nod has potentially started. Allowing for differences 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 by 7^o or more within 600ms a nod is registered; outside this time frame, 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 do not want to choose anything). The same method works for nods in all directions, but uses roll for left and right nods. This method is simple but fairly robust to the noise of most small, normal head movements, movements due to walking, and gross individual differences in nodding.

3.1.2 Soundscape Design

As an application for our 3D audio radial pie menus, we chose to present current affairs information options to users. Four menu items were presented - Weather, News, Sport, and Traffic – the scenario being that a user wearing the device might want information about one or more of these when out and about 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 theme tune of a UK news program;
- *Sport:* A clip taken from the theme tune of a UK sports program;
- *Traffic:* A mix of various busy street samples, including cars, trucks, engines, horns and skids.

Three soundscapes were designed. These looked at different placements of the sounds in the audio space and whether the space was ego- or exocentric (our 3D sounds are rendered by Microsoft's DirectX 8 API). The designs were:

1. *Egocentric:* Sounds are placed at the four cardinal points (every 90^o 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, in order rotating clockwise around the head. This is a simple design but does necessitate many backward nods that are hard on the neck muscles. It is also hard, with this method, to have more than 4 items in the soundscape as nodding accurately at 45^o in the rear hemisphere is difficult.

2. *Exocentric, constant:* This interface has the four sounds arranged in a line in front of the user's head. The user can select any one of the items by rotating his/her head slightly until directly facing the desired sound, and then nodding. All nods are therefore basically forward nods, which are much easier to perform, 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 of which is 40^o) and a 'thump' is played when the segment at each end is passed (to let the

user know that the last sound has been reached). All sounds are played constantly and simultaneously; the sound currently directly 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 are no longer in front, but can be reset to the front again by nodding backwards. This is a more complex design than (1) but requires 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) with the exception that the sounds are played one after the other in a fixed order from left to right, similar to (1). This means there are fewer sounds playing simultaneously so the soundscape is less crowded but item selection may be more time consuming since the user may have to wait for a sound to play to know where to nod.

3.2 Hand Gestures

Pirhonen *et al.* [23] investigated the use of metaphorical gestures to control an MP3 player. For example, a 'next track' gesture was a sweep of a finger across the iPAQ screen from left to right and a 'volume up' gesture was a sweep up the screen, from bottom to top. Their experimental results showed that these were an effective interaction paradigm and more usable than the standard, button-based, 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 a stage further to investigate the use of audio feedback during the *progress* of the gestures. Like Pirhonen *et al.*, it was not our intention to develop a hand-writing recognition system (as it is very hard to handwrite on the move together with the fact that our aim was to investigate *novel* interaction paradigms) and we also concentrated on metaphorical gestures that could be used for a range of generic operations on a wearable device.

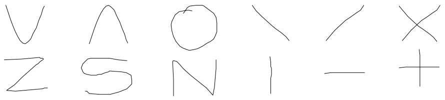

Figure 3: Gesture set used during investigation

For the purpose of our investigation, we focussed on a combination of 12 single- and multiple-stroke alphanumeric and geometric gestures (see Figure 3) encompassing those used by Pirhonen, that might potentially be used to control mobile applications.

3.2.1 Hand Gesture Recognition

We developed a gesture recogniser to allow a user to draw, simply using his/her finger, 2D gestures on the screen of a PDA (in our case, an iPAQ) without any need to look at the display of the PDA. The recogniser is generic in that it can be used to recognise any gesture that is predefined by an application developer 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 opted for a square layout as opposed to Friedlander's Bullseye concentric rings since it is a better fit with the shape of the iPAQ screen. Derived from a publicly available algorithm [26], 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.

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) The sounds used

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 complete 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 time-frame, 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, by double tapping the screen, the user can abort a gesture.

3.2.2 Audio Feedback Design

Audio feedback was designed to represent the 3 x 3 matrix. Unlike Friedlander *et al.'s* system wherein a single beep represented all menu items so navigation was based on counting, our sounds are designed to dynamically guide users correctly through gestures. Our sounds are 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. The notes C_x E_x G_x (where x corresponds to the octave for the selected row) would therefore be generated by a sweep left to right across a row. On the basis of the above basic 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 recognise them when heard. We developed two implementations of this basic design:

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 bounds of Bin 1, the C_6 will be played. This note will sound continuously until the user moves his/her finger into another bin (at which point the note being played will change to that corresponding to the new bin location) or until the user lifts his/her from the iPAQ screen.

2. *Complex Audio:* This implementation extends (1) 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 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 together with direction of movement, this information would allow users to pre-emptively avoid unintentionally slipping into incorrect bins for any given gesture, thus improving accuracy.

3.3 Experimental Design

An experiment was required to determine whether 3D audio menus combined with head-based gestures would be a usable method of selection in a wearable computer when the user is in motion, and to investigate which soundscape is most successful. Similarly, an experiment was required to investigate the extent to which 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.

Both experiments used a similar set up. Users had to walk 20m laps around obstacles set up in a room in the University of Glasgow – the aim being to test our interaction designs 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, an extensive range of measures was taken to assess the usability of the interaction designs tested. We measured time to complete tasks, error rates, and subjective workload (using the NASA TLX [16] scales). Workload is imperative in a mobile context: since users must monitor and navigate their physical environment, fewer attentional resources can or should be devoted to the computer. An interaction paradigm (and hence 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 since the inclusion of audio feedback in interface design is often considered annoying, due largely to the fact that it is oftentimes used inappropriately and in an *ad hoc* fashion.

To assess the impact of the physical device combined with the interaction techniques on the participants, we also recorded percentage preferred walking speed (PPWS) [22]: the more negative the effect of the device the further below their normal walking speed that users would walk. Pirhonen *et al.* [23] found this to be a sensitive measure of the usability of a gesture-driven 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 the wearable device.

The final measure taken was comfort. This was based on the Comfort Rating Scale (CRS) – a new scale developed by Knight *et al.* [17] – which assesses various aspects to do with the perceived comfort of a wearable device. For technology, and the associated interaction with and support offered by that technology, to be accepted and used the technology 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, anxiety, attachment, harm, perceived change, and movement. Knight *et al.* have used it to assess the comfort of two wearable devices that they are building within their research group. Using this will allow us to find out more about the actual acceptability or potential of our proposed interaction designs when used in motion with mobile technology.

3.3.1 Head Gestures – Experimental Design

A fully counterbalanced, within-groups design was used with each participant using the three different interface (soundscape) designs whilst walking. Preceding each condition, brief training was provided to the participants. Ten selections for each of the four menu items – that is, forty menu item selections in total – were required per condition. 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. Participants were not informed as to the correctness of their selections. Eighteen people participated: 13 males and 5 females, with ages ranging from 18 – 55. In addition to the measures described previously, we also collected information about the number of incorrect selections made and the distance walked.

Our primary hypothesis was that nodding would be an effective interaction technique when used on the move. Our secondary hypothesis was that soundscape design would have a significant effect on usability: Egocentric selection of items should be faster than Exocentric since with Egocentric presentation the user needs to nod at the chosen object whilst with Exocentric the user must first locate the sound, then nod.

3.3.2 Hand Gestures – Experimental Design

This experiment used the same basic setup as the head gesture experiment. This time, however, a Compaq iPAQ was used as the input device and participants drew gestures on the screen using 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 wearable using the Pebbles software from CMU (http://www2.cs.cmu.edu/~pebbles/overview/software.html). The sounds were not presented in 3D in this case. A fully counter-balanced, 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 two audio designs. Participants were allowed to familiarise themselves with the recogniser for use under each condition, but no formal 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 participants 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 had participated in the head gesture experiment. 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 initially (that is, until the users had gained familiarity with the system) 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.

4 Results & Discussion

This section outlines the results obtained from the two experiments comprising our investigation to date and discusses some of the implications therein.

4.1 Primary Findings

Consider first, the results of the head gesture experiment. A single factor ANOVA showed that total time taken was significantly affected by soundscape ($F_{2,51}=14.24$, $p<0.001$), as shown in Table 1.

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

Table 1: Mean time taken per condition when using audio pie menus with head-based gestures

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

There were no significant differences in the number of incorrect nods in each condition (approximately 80% accuracy rates were achieved across all conditions).

Consider now, the results of the hand gesture experiment. 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.

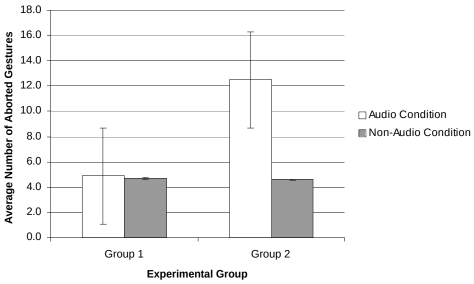

Figure 5: Mean number of aborted hand gestures

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 5 shows the average number of aborted gestures according to experimental group and condition.

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 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 will be required to confirm or counter this observation.

4.2 Workload

With respect to the head gesture experiment, there were no significant differences in overall workload across the experimental conditions. Only annoyance was significantly effected ($F_{2,51}=3.29$, $p<0.05$). Tukey HSD tests showed Exocentric periodic was significantly more annoying to participants than Egocentric ($p<0.05$) but no other differences were significant.

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

4.3 Comfort

The comfort ratings returned from both experiments were not significantly different. Like the NASA TLX, low ratings are desirable; of the six categories, the 'Attachment' of the wearable was shown to be the biggest obstacle to comfort. This category is concerned with the 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 localized manner. In the second experiment, participants also had an iPAQ attached to the belt, contributing extra weight. The pressure of the headphones against the participant's head further add to the feeling of attachment. It is interesting to note that, despite wearing the device (with added weight) for longer in the second experiment than in the first (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 the second experiment.

4.4 PPWS

For the head gesture experiment, an analysis of PPWS 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 other two Exocentric designs ($p<0.05$), but there were no significant differences between the latter two. The mean score in the Egocentric condition was 69.0% of PPWS, with 47.5% and 48.5% for Exocentric constant and periodic respectively.

PPWS varied considerably across the participants; some users found the wearable easy to use, whilst others slowed dramatically. One participant actually walked faster than normal when using the Egocentric design; two participants had problems and walked considerably slower than normal under all three conditions. Of the latter two participants, one found the distance needed to complete the experiment hard work and slowed down even after the initial assessment of PWS; the other stopped numerous times when selecting items, finding it hard to walk and nod simultaneously. We will investigate the issues these users exhibited in the next stage of our work to ensure that the head-gesture paradigm is usable by as many people as possible.

With respect to the hand gesture experiment, we had hypothesised that, as a result of increased levels of feedback, the audio designs would initially 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 PWS), a two factor ANOVA showed no significant affect of audio condition on PPWS.

It is interesting to note that walking speed was slower with head than hand gestures (which had no significant affect on walking speed). Perhaps this is unsurprising as nodding may make it harder for users to observe where they are going. Our more sophisticated head gesture recogniser (see Section 6) will allow us to recognise smaller head gestures more reliably which may reduce this problem and its effects on walking speeds.

5 Conclusions

Overall, the two experiments have demonstrated that novel interaction paradigms based on sound and gesture have the potential to address issues concerning the usability of, and standard of interaction with, eyes-free, mobile use of mobile or wearable devices.

Head gestures have been shown to be a promising interaction paradigm with the 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.

The technology required to support both these interaction designs was, when rated by our participants, considered comfortable and is therefore likely to be acceptable to real users. This is important since it is unlikely that an interaction paradigm will be accepted and used if the technology required to support the design is cumbersome and intrusive. That said, mobile technology is advancing so rapidly that a novel interaction paradigm that is prototypic and perhaps awkward at its inception is likely to be realistic and feasible not long afterwards. Hence, we should not, in our search for better interaction paradigms for use with mobile devices, be deterred unduly by current technology.

We have shown that non-visual interaction paradigms can be used effectively with 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. These are, however, only two examples of what is potentially possible in terms of alternative interaction for such devices. If we are to effectively embrace the *mobility* of mobile and wearable devices we need to acknowledge their limitations and the variability of conditions under which they are used and design new interaction paradigms that meet these very specific and challenging needs.

6 Further Work

As previously mentioned, the design of the Egocentric audio display encounters problems if more than four items are needed in a menu. A further experiment is needed to assess the maximum number of items a user *could* deal with in such a soundscape. It may be that four is the maximum given that the user has to handle the complexities of navigating round and listening to sounds from his/her environment in addition to interacting with the mobile device. During informal studies with seated participants, Savidis *et al.* [24] observed that users found it difficult to deal with 6 items placed around them. If it *is* possible for a user to deal with more than four items, then the Exocentric interface designs are likely to become more useful. It is also 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 nod in the correct direction.

The results suggest that, for faster performance, the audio cues (sounds) should be played simultaneously. This might not, however, be true when a larger number of items are included in the soundscape; further study is needed to investigate this issue.

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

The design of the menus could be extended to allow for hierarchical menu structures. If, as suggested previously, 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 sub-menu, and a backward nod could be used to return to the previous level. Given the lack of visual display, to ensure that users are aware of their position in such a structure, hierarchical earcons could be used to indicate position [4]. Care must be taken when designing such earcons so that they do not conflict with the sounds for the menu items themselves. A mix of auditory icons for menu items and earcons for navigation would help with separation.

Areas to investigate to try and lessen users’ awareness of the mobile technology and thereby render these novel interaction paradigms more transparent would include the style of the headphones used, the manner and location in which the device is physically attached to the body and the activity-specific requirements. One advantage these interaction designs have over visually-based interaction designs which require the use of head mounted displays is that many people currently wear headphones (for music players, cell phones or radios) making the technology required to support our interaction paradigms stand out less, lowering our CRS Anxiety scores. A further long-term study is needed to see if people would use these interaction paradigms in real situations. Even though the CRS ratings are good, nodding might very well be unacceptable in public unless we can make the nods required very small. This will be a focus for further investigation.

The results showed the potential for improved accuracy of 2D hand gestures when supported by dynamic audio feedback. Furthermore, the simpler the audio feedback design, the better able users appear to be able to interpret and respond to the dynamic feedback. Further investigation needs to be conducted into the potential for recognition and recall of the audio feedback; in particular, to enhance these elements of usability across the broadest range of users, investigation into the optimal earcon design needs to be completed.

On the basis of the results returned for the hand gesture recogniser, we are currently investigating similar audio enhanced support for the mobile use of unistroke alphabets – essentially a sophistication of the general notion of 2D gestures. In particular, taking as a basis the audio design for the gesture recogniser discussed here, we are investigating alternative audio designs to determine how best to support unstroke alphabet use when visual resource cannot be devoted to the use of the alphabet. Additionally, we are investigating how individual handwriting style (be it cursive, print, or mixed) impacts upon the use of unistroke systems with a view to personalization of such systems in terms of the manner in which audio feedback can be used to address inaccuracies inherited from natural writing style.

Acknowledgements

This work was funded in part by EPSRC grant GR/R98105 and ONCE, Spain. The authors would also like to thank Marek Bell and Malcolm Hall, research students in the Department of Computing Science at the University of Glasgow without whose dedicated efforts the research discussed would not have been possible.

About the Authors

Joanna Lumsden is a Research Officer with the National Research Council (NRC) of Canada where she works with the Atlantic IIT e-Business Human Web Group. Prior to joining the NRC, Joanna worked as a research assistant with Stephen Brewster in the Computing Science Department at the University of Glasgow, U.K. where she attained her Ph.D. in HCI.

Stephen Brewster is a professor in the Department of Computing Science at the University of Glasgow, U.K. He is head of the Multimodal Interaction Group (MIG) which collectively investigates the design and effective use of non-traditional interaction techniques (e.g. haptic, non-speech audio etc.). Stephen obtained his Ph.D. from the HCI Group at the University of York, U.K. in 1994.

References

[1] B. Arons, "A Review of the Cocktail Party Effect," *Journal of the American Voice I/O Society*, **12** (July), pp. 35 - 50, 1992.

[2] W. Barfield and T. Caudell, *Fundamentals of Wearable Computers and Augmented Reality*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2001.

[3] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. Cambridge, MA: Academic Press, 1994.

[4] S. A. Brewster, "Using Non-Speech Sound to Provide Navigation Cues," *ACM Transactions on Computer-Human Interaction*, **5** (3), pp. 224 - 259, 1998.

[5] S. A. Brewster, "Sound in the Interface to a Mobile Computer," presented at HCI International'99, Munich, Germany, pp. 43 - 47, 1999.

[6] S. A. Brewster, "Overcoming the Lack of Screen Space on Mobile Computers," *Per-*

sonal and Ubiquitous Computing, **6** (3), pp. 188 - 205, 2002.

[7] S. A. Brewster and P. G. Cryer, "Maximising Screen-Space on Mobile Computing Devices," presented at ACM CHI'99, Pittsburgh, PA, pp. 224 - 225, 1999.

[8] S. A. Brewster, G. Leplatre, and M. G. Crease, "Using Non-Speech Sounds in Mobile Computing Devices," presented at First Workshop on Human Computer Interaction with Mobile Devices, Glasgow, UK, pp. 26 - 29, 1998.

[9] S. A. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker, "Multimodal 'Eyes-Free' Interaction Techniques for Mobile Devices," presented at Human Factors in Computing Systems - CHI'2003, Ft Lauderdale, USA, pp. 473 - 480, 2003.

[10] S. A. Brewster and R. Murray, "Presenting Dynamic Information in Mobile Computers," *Personal Technologies*, **4** (4), pp. 209 - 212, 2000.

[11] M. Cohen and L. F. Ludwig, "Multidimensional Audio Window Management," *International Journal of Man-Machine Studies*, **34** (3), pp. 319 - 336, 1991.

[12] A. K. Dey and G. D. Abowd, "Toward a Better Understanding of Context and Context-Awareness," College of Computing, Georgia Institute of Technology, Atlanta, Georgia GVU Technical Report GIT-GVU-99-22, 1999.

[13] N. Fiedlander, K. Schlueter, and M. Mantei, "Bullseye! When Fitt's Law Doesn't Fit," presented at ACM CHI'98, Los Angeles, CA, pp. 257 - 264, 1998.

[14] E. Geelhoed, M. Falahee, and K. Latham, "Safety and Comfort of Eyeglass Displays," in *Handheld and Ubiquitous Computing*, P. Thomas and H. W. Gelersen, Eds. Berlin: Springer, 2000, pp. 236 - 247.

[15] B. Harrison, K. Fishkin, A. Gujar, C. Mochon, and R. Want, "Squeeze me, hold me, tilt me! An Exploration of Manipulative User Interfaces," presented at ACM CHI'98, Los Angeles, CA, pp. 257 - 296, 1998.

[16] S. G. Hart and C. Wickens, "Workload assessment and prediction," in *MANPRINT: an approach to systems integration*, H. R. Booher, Ed. New York: Van Nostrand Reinhold, 1990, pp. 257 - 296.

[17] J. F. Knight and C. Baber, "Physical Load and Wearable Computers," presented at 5th International Symposium on Wearable Computers, Atlanta, GA, 2000.

[18] G. Leplatre and S. A. Brewster, "Perspectives on the Design of Musical Auditory Interfaces," presented at Focus Conference on Anticipation, Cognition and Music, Liege, Belgium, 1998.

[19] R. Malkewitz, "Head Pointing and Speech Control as a Hands-Free Interface to Desktop Computing," presented at ACM ASSETS'98, Marina del Rey, CA, pp. 182 - 188, 1998.

[20] B. Meyers and A. Kern, "<Context-Aware> schema </Context-Aware>," presented at CHI'2000 Workshop on the What, Who, When, Where, Why and How of Context-Awareness, The Hague, The Netherlands, 2000.

[21] J. Pascoe, "The Stick-e Note Architecture: Extending the Interface Beyond the User," presented at International Conference on Intelligent User Interfaces, Orlando, FL, pp. 261 - 264, 1997.

[22] H. Petrie, S. Furner, and T. Strothotte, "Design Lifecycles and Wearable Computers for Users with Disabilities," presented at The First Workshop on Human-Computer Interaction with Mobile Devices, Glasgow, UK, 1998.

[23] P. Pirhonen, S. A. Brewster, and C. Holguin, "Gestural and Audio Metaphors as a Means of Control in Mobile Devices," presented at ACM-CHI'2002, Minneapolis, MN, pp. 291 - 298, 2002.

[24] A. Savidis, C. Stephanidis, A. Korte, K. Crispien, and C. Fellbaum, "A Generic Direct-Manipulation 3D-Auditory Environment for Hierarchical Navigation in Non-Visual Interaction," presented at ACM ASSETS'96, Vancouver, Canada, pp. 117 - 123, 1996.

[25] N. Sawhney and C. Schmandt, "Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments," *ACM Transactions on Computer-Human Interaction*, **7** (3), pp. 353 - 383, 2000.

[26] M. Willey, "LibStroke - A Stroke Translation Library," vol. 2002: ETLA Technical Services, 2002.