The GAIME project: Gestural and Auditory Interactions for Mobile Environments

Stephen Brewster, Roderick Murray-Smith, Andrew Crossan
Yolanda Vasquez-Alvarez and Julie Rico
Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
+44 (0)141 330 4966
[stephen, rod, ac, yolanda, julie]@dcs.gla.ac.uk / www.gaime-project.org

ABSTRACT
In this paper we discuss some of the key issues behind the need to move away from screens and keyboards on mobile devices. Our aim is to allow non-hand based gestures for input with 3D sound and tactile displays for output to create more effective mobile interactions that make mobile devices easier to use when on the move.

Categories and Subject Descriptors
H.5.2. [**User Interfaces**]: Haptic I/O, H.5.2 **[User Interfaces]**. Auditory (non-speech) feedback, H5.m. **[Information interfaces and presentation]** Miscellaneous.

General Terms
Human Factors.

Keywords
Hands-free, eyes-free, multimodal interaction, gesture input, 3D audio, tactile, pressure input.

1. INTRODUCTION
Most PDAs and smart phones have sophisticated graphical interfaces and commonly use small keyboards or styli for input. The range of applications and services for such devices is growing all the time. However, there are problems which make interaction difficult when a user is on the move. Much visual attention is needed to operate many of the applications, which may not be available in mobile contexts. Oulasvirta *et al.* [15] showed that attention can become very fragmented for users on the move as it must shift between navigating the environment and the device, making interaction hard. Our own research has shown that performance may drop by more than 20% when users are mobile [3].

Another important issue is that most devices require hands to operate many of the applications. These may not be available if the user is carrying bags, holding on to children or operating machinery, for example. A key research topic is therefore to reduce the reliance on graphical displays and hands by investigating other forms of input and output, for example gesture input from other locations on the body combined with three-dimensional sound for output.

Published by the British Computer Society

Little work has gone into making input and control 'hands free' for mobile users. Speech recognition is still problematic in such settings due to its high processing requirements and the dynamic audio environments in which devices are used. Much of the research on gesture input still uses hands for making the gestures. There is some work on head-based input, often for users with disabilities [11], but little of this has been used in mobile settings. Our own previous work has begun to examine head pointing and showed that it might be a useful way to point and select on the move [4].

Many other body locations could be useful for subtle and discreet input whilst mobile (e.g., users walking or sitting on a bumpy train). For example, wrist rotation has potential for controlling a radial menu as the wrist can be rotated to move a pointer across the menu. It is unobtrusive and could be tracked using the same sensor used for hand pointing gestures (in a watch for example). There has been no systematic study of the different input possibilities across the body.

Output is also a problem due to the load on visual attention when users are mobile. We and others have begun to look at the use of spatialised audio cues for output when mobile as an alternative or complement to graphics [13, 19] [8, 18]. Many of these use very simple 3D audio displays, but, with careful design, spatial audio could provide a much richer 'eyes-free' display space. Our AudioClouds project (www.audioclouds.org) built some foundations for 3D audio interactions, investigating basic pointing behaviour, target size and separation [4, 13].

Tactile feedback also has possibilities for freeing up the eyes when mobile. The whole of the body surface is available for information display. Sensory substitution has a long history in the area of accessibility [10, 20, 22] and has great promise for mobile interactions. The simple vibration motor currently in mobile phones is very popular, but is only used in a very simple way. Research is needed to really make use of tactile displays for rich mobile interaction.

This paper describes some of the research background in each of these key areas and interaction techniques that have been designed to use them with the aim of providing eyes and hands free interaction. This work is part of the UK EPSRC-funded GAIME project: Gestural and Auditory Interactions for Mobile Environments (www.gaime-project.org).

2. BACKGROUND AND TECHNIQUES
Many interactions in current mobile devices are based on those from desktops with buttons, scrollbars and dialogue boxes common. These were originally developed for high-resolution, large desktop displays operated by a mouse and are not always appropriate for mobile situations where the emphasis is on devices being small, portable and used when the user is in motion.

Our own work has shown that when using a PDA while walking or on bumpy public transport, selecting onscreen targets becomes significantly harder than when stationary [9]. The device moves as the user moves causing larger and more variable targeting errors and increasing the time to tap on a target [6]. This is particularly frustrating in situations requiring many clicks such as using an on-screen keyboard.

Gesture input has been successfully incorporated into both research and commercial devices [14, 16], with standard mobile phones now incorporating accelerometers that could be used for input. However, many previous projects have concentrated on gestures done with hands or fingers. For example, Strachan and Murray-Smith [8] describe the Bodyspace project where the user accesses applications by holding a phone and moving it to different areas of the body. The Nintendo Wii games console uses gesture-based interactions sensed through accelerometers in a controller held in the hand. In many mobile settings it may not be possible for users to operate devices that needs one or two hands; they may already be occupied by holding shopping bags or children.

Mobile devices often demand much visual attention from users. A user must look at the screen to type a message or target a button, but this is far from ideal if users wish to use the device on the move. They may be walking in a busy street or crossing a road which will require their visual attention to navigate safely. Oulasvirta *et al.* [15] examine how users divide their attention between looking at the screen and navigating a crowded area, showing that interactions are restricted to short bursts. This makes for slow and error-prone user interfaces. New ways of interacting are necessary so that users can make full use of the capabilities of their powerful mobile devices when on the move. We are investigating novel body-based gestures for input which do not require the user's hands. For output are studying the use of 3D sound and tactile feedback to create rich displays that reduce the need for visual attention.

Figure 1: The SHAKE sensor pack.

2.1 Sensing

An important issue for mobile interactions is the sensing technology used for the gestures. Accelerometers are one type of inertial sensor that allows a user to interact by gesturing, moving or tilting a device. **Figure 1** shows the SHAKE, a matchbox-sized sensor pack we developed as part of our AudioClouds research project to detect gestures. It contains accelerometers, magnetometer, gyroscopes and connects to a phone or other device via Bluetooth (www.dcs.gla.ac.uk/research/shake/). An advantage of using such sensors over most stylus-based gesture systems is that they can provide one-handed, screen free gesture control. Rekimoto [17] first suggested tilt as an input technique for mobile devices for interacting with menus and scrolling. An accelerometer-based screen orientation system has been integrated into Apple's iPhone and both Nokia and Samsung have phones on the market that include accelerometers. Murray-Smith examined tilt as a method of text entry on a mobile device [9]. He combined tilt input with a language model to allow the system to infer the current word being typed and adjust the dynamics of the system in order to make that word easier to enter. These examples are still based around visual displays with one hand used to operate the device. Our aim is to reduce the need for the use of the hands for some tasks, so that input can be made whilst mobile and carrying bags or holding on to children

2.2 'Hands free' gestures for input

There is currently little work examining the viability of body locations other than the hands for gesture input. 'Hands free' presently means using a headset to speak on the phone without holding it; the other interactions and applications a device can perform are inaccessible without the hands. Speech recognition has possibilities here, but is difficult to do on mobile devices due to processing requirements and dynamic audio environments. There are also situations where speech may be inappropriate (quiet environments, for example). Gestures are a good alternative and work well with speech.

Rekimoto describes GestureWrist in [16], which recognises user hand gestures through a device attached to the user's wrist without encumbering the user's hand with sensors. Previous work has examined pointing with different joints in the arm to control a cursor in a desktop situation. Zhai *et al.* [23] investigated the use of fingers, left/right motion of the wrist, elbow and shoulders in a Fitts' Law task for pointing in a graphics application. Balakrishnan *et al.* [1] showed that similar performance to can be obtained in a computer based pointing task with wrist and arm movements alone when compared to the same task with additional finger movement, for static locations. Similar Fitts' Law studies have examined head pointing for targeting. Our own work includes a study of mobile head pointing using one axis of rotation of the head to select menu items [4]. Users were able to select targets successfully when walking using head nods.

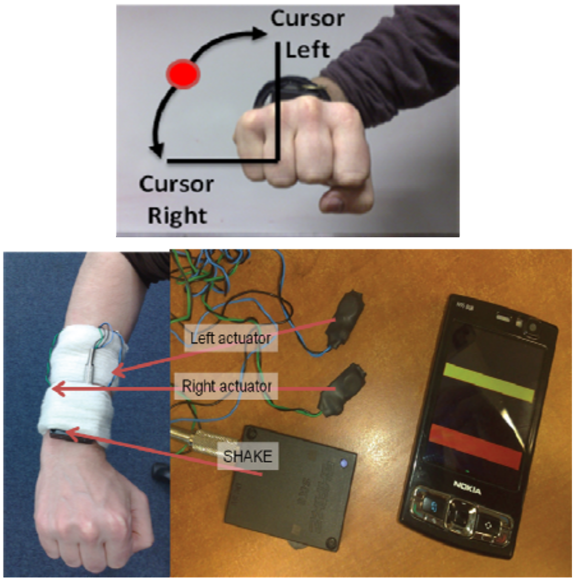

Figure 2: Wrist rotation for input. Top shows the range of rotation used, lower shows tactile actuators and phone display.

In more recent work we studied the use of wrist rotation as a form of input to drive a cursor across a screen or menu. This would be an effective interaction when the user was holding bag for example, as the wrist can still be rotated. Wrist rotation is also discreet: others will not notice a user doing it, avoiding issues of social acceptability. For this experiment a SHAKE was mounted on a strap on top of the wrist (like a watch). We used this in a Fitts' Law task to test pointing behavior with a graphical display on a phone and with tactile feedback on the wrist to indicate targets (see Figure 2). Results showed that it was an effective input technique with users able to achieve 90% accuracy when selecting targets of 9° [7].

One closely related area of work is in that of context sensitive systems. Previous work has examined inferring user activity (e.g. seated, walking, running or in a vehicle) or location through one or more sensors. Lukowicz *et al.* [12] used changes in muscle activity to infer the user's context. They were able to determine accurately whether a user was climbing or descending stairs, or walking normally. Our own work examined sensing users' gait through a mobile device instrumented with an accelerometer [6]. We were able to infer gait from their motion, allowing a detailed study of the interactions between walking and screen tapping behaviour. This could be adapted to look for small changes in gait that could be used as a form of input: users could slightly change the timing of a step to make input when hands were busy.

2.3 Spatial sound for 'eyes free' output

The display of information can be problematic on mobile devices as screens are small and it can be hard to look at them when on the move. Sound has the possibility to overcome some of these problems. In particular, 3D sound (usually based on Head-Related Transfer Functions [2] and delivered via headphones) can create a larger display area around the user and allow the use of spatial location for information display. This spatial element also gives a range of new possible uses for sound, rather than just being a notification system. 3D sound APIs are becoming more common on PDAs and mobile phones, due to the requirements of mobile gaming, and these give us the possibility for creating novel mobile interactions. Sound can reduce the burden on the visual display, allowing it to be used more effectively and also to allow users to look at the environment as they walk or move about.

We [13, 19] and others [8, 18] have begun to look at the use of 3D audio cues for output when mobile as an alternative or complement to graphics. Schmandt at MIT has done some significant work on 3D audio displays, in particular NomadicRadio [18] that used basic 3D sound for reminders and notifications in a mobile device. Our EPSRC-funded AudioClouds project laid the foundations for our own work by investigating low level issues with 3D audio displays for mobile interaction. We looked at issues such as size of audio targets, the effects of distracters and the user of simple gestures for the selection of audio targets [1,3]. We developed some simple interactions such as a 3D audio progress indicator where a sound moved around the user's head (starting and ending in front of the nose). The position around the head gave the amount of progress, the rate of movement the rate of progress. Results showed that this was an effective eyes-free way of delivering progress information [21].

One key issue to come out of our earlier work was the use of *Egocentric* and *Exocentric* audio displays. Egocentric displays are fixed to the user (so sounds stay fixed to the user as s/he turns). Exocentric displays are fixed to the world (so sounds stay fixed in place in the world when the user turns). Our results showed that egocentric displays were faster but more error prone, whilst exocentric displays were slower but more accurate [13]. We need to bring the speed of egocentric together with the accuracy of exocentric to get the best performance we can. As yet, both types have not been combined in a single interface. Good interaction techniques need to be developed for both of these types of sounds and for the interface between them, for example when an item from the local interface might be placed as a marker in the exocentric space.

2.4 Tactile feedback

Most current mobile phones include a vibrotactile actuator which is used as a vibration alert when a call or text message arrives. This is very popular with users but there is a lot more that the sense of touch can do for 'eyes free' interactions. The skin is the largest organ in the body and provides a large space for communicating information to the user. Tactile feedback is discreet as a message is delivered directly to the skin.

Tan has done significant work in this area, investigating different body locations and different types of hardware for delivering tactile feedback [20]. Our own work in the area of tactile displays has focussed on the design of Tactons, or tactile icons. These are structured forms of tactile feedback that can be used to deliver information to users [5]. We have used vibrotactile actuators such as that in Figure 3 to assess different body locations for feedback (including hand, forearm, waist, wrist and ankle) and to understand the types of vibrations users can perceive. We have also developed a range of interactions such as tactile progress bars and keyboards [9]. Further research is needed to fully understand the capabilities of the skin and how to make use of it for display and interaction.

Figure 3: The C2 Tactor from Engineering Acoustics (www.eaiinfo.com).

2.5 Social acceptability

One final important issue is social acceptability, both of the gestures and the attachment of sensors and actuators to different body locations. We need to ensure that people are happy to do the (hopefully small and discreet) gestures in different contexts and that they would be happy to wear the sensors. By evaluation in realistic contexts we can assess users' responses to these novel forms of body-based interaction. Our aim is to investigate both how the particular individual feels about doing a gesture and also about how others nearby perceive the individual doing it.This area has received little attention in the gesture input literature, where focus has been on the difficulties of recognition in complex environments. However, for gestures to be useful they have to be used and if they are too embarrassing to do then they will provide little benefit. In our research we aim to find a set of non-hand based gestures that are effective for interaction and socially acceptable.

3. CONCLUSIONS

The work we have reported here describes some of the work that we have been doing in the area of hands and eyes free interaction for mobile devices. Input has focused on using gestures made from body locations other than the hands as these are often busy in mobile settings. For output we are studying spatial audio and tactile based displays. Spatial audio allows the creation of a much richer display space than the single speaker or stereo headset currently used. Tactile feedback can be displayed across the body surface, which has a large area making it effective for information display. Key research questions remain in all of these areas, but particularly important issues are in the social acceptability of body-based gestures in different contexts of use.

4. ACKNOWLEDGMENTS

This work was funded by EPSRC project GAIME: Gestural and Audio Interactions for Mobile Environments (EP/F023405/1).

5. REFERENCES

[1] Balakrishnan, R. and MacKenzie, I.S., Performance Differences in the Fingers, Wrist, and Forearm in Computer Input Control. in in Proceedings of ACM CHI, (Atlanta Georgia, USA, 1997), ACM Press, 303-310.

[2] Blauert, J. Spatial Hearing. MIT Press, Cambridge, MA, USA, 1997.

[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., Lumsden, J., Bell, M., Hall, M. and Tasker, S., Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices. in Proceedings of ACM CHI 2003, ACM Press, Addison-Wesley, 463-480.

[5] Brown, L.M., Brewster, S.A. and Purchase, H.C., A First Investigation into the Effectiveness of Tactons. in Proceedings of Worldhaptics 2005, IEEE Press, 167-176.

[6] Crossan, A., Murray-Smith, R., Brewster, S.A. and Musizza, B. Instrumented Usability Analysis for Mobile Devices. in Lumsden, J. ed. Handbook of Mobile HCI, The Ideas Group Inc. , 2008.

[7] Crossan, A., Williamson, J., Brewster, S.A. and Murray-Smith, R., Wrist Rotation for Interaction in Mobile Contexts. in Proceedings of MobileHCI 2008, ACM Press, 435-438.

[8] Goose, S. and Moller, C., A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. In Proceedings of 7th ACM international conference on Multimedia, (1999), ACM Press, 363-371.

[9] Hoggan, E., Brewster, S.A. and Johnston, J., Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In Proceedings of ACM CHI2008, ACM Press Addison Wesley, 1573-1582.

[10] Kaczmarek, K., Webster, J., Bach-y-Rita, P. and Tompkins, W. Electrotacile and vibrotactile displays for sensory substitution systems. IEEE Transaction on Biomedical Engineering, 38 (1). 1-16.

[11] LoPresti, E., Brienza, D.M., Angelo, J., Gilbertson, L. and Sakai, J., Neck range of motion and use of computer head controls. In Proceedings of ACM ASSETS 2000, ACM Press, 121-128.

[12] Lukowicz, P., Hanser, F., Szubski, C. and Schobersberger, W., Detecting and Interpreting Muscle Activity with Wearable Force Sensors. In Proceedings of Pervasive, (2006), Springer-Verlag, 101-116.

[13] Marentakis, G.N. and Brewster, S.A., Effects of Feedback, Mobility and Index of Difficulty on Deictic Spatial Audio Target Acquisition in the Horizontal Plane. in Proceedings of ACM CHI 2006, ACM Press Addison-Wesley, 359-368.

[14] Oakley, I. and O'Modhrain, M., Tilt to scroll: evaluating a motion based vibrotactile mobile interface. in In Proceedings of World Haptics, (2005), IEEE, 40-49.

[15] Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J., Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In Proceedings of ACM CHI 2006, ACM Press Addison Wesley, 919-928.

[16] Rekimoto, J., Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In Proceedings of Fifth International Symposium on Wearable Computers (ISWC '01).

[17] Rekimoto, J., Tilting Operations for Small Screen Interfaces. In Proceedings of UIST'96, 167-168.

[18] 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.

[19] Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S. and O' Modhrain, S., GPSTunes - controlling navigation via audio feedback.In Proceedings of MobileHCI 2005, ACM, 275-278.

[20] Tan, H.Z. and Pentland, A. Chapter 18: Tactual displays for sensory substitution and wearable computers. in Barfield, W. and Caudell, T. eds. Fundamentals of wearable computers and augmented reality, Lawrence Erlbaum Associates, Mahwah, New Jersey, 2001, 579-598.

[21] Walker, V.A. and Brewster, S.A. Spatial audio in small screen device displays. Personal Technologies, 4 (2). 144-154.

[22] Wall, S.A. and Brewster, S.A. Sensory substitution using tactile pin arrays: Human factors, technology and applications. Signal Processing, 86. 3674-3695.

[23] Zhai, S., Milgram, P. and Buxton, W., The Influence of Muscle Groups on Performance of Multiple Degree-of-Freedom Input. In Proceedings of ACM CHI'96, ACM Press, 308-315.