Multimodal Mobile Interactions: Usability Studies in Real World Settings

Julie R. Williamson, Andrew Crossan, and Stephen Brewster
Glasgow Interactive Systems Group
University of Glasgow, Glasgow G12 8QQ
{julie, ac, stephen}@dcs.gla.ac.uk

ABSTRACT

This paper presents a study that explores the issues of mobile multimodal interactions while on the move in the real world. Because multimodal interfaces allow new kinds of eyes and hands free interactions, usability issues while moving through different public spaces becomes an important issue in user experience and acceptance of multimodal interaction. This study focuses on these issues by deploying an RSS reader that participants used during their daily commute every day for one week. The system allows participants on the move to access news feeds eyes free through headphones playing audio and speech and hands free through wearable sensors attached to the wrists. The results showed participants were able to interact with the system on the move and became more comfortable performing these interactions as the study progressed. However, participants were far more comfortable gesturing on the street than on public transport, which was reflected in the number of interactions and the perceived social acceptability of the gestures in different contexts.

Categories and Subject Descriptors

H5.2 **[User Interfaces]**: *Input devices and strategies*.

General Terms

Human Factors.

Keywords

Whole body interaction, mobile interaction, inertial sensing, user studies in the wild, wrist rotation.

1. INTRODUCTION

Multimodal mobile systems allow for improved interaction while on the go by allowing users to access functionality in new ways, giving users new methods of presenting themselves with technology, and bringing interaction out beyond the device to the whole body. However, usability in mobile settings suffers both practically and socially due to the fluid environments where these interfaces might be used. This paper presents a study that examines a multimodal mobile interface used while commuting where users can interact with and receive information eyes and hands free.

Previous work in this area has looked at many aspects of multimodal interaction without tackling the issues of usability and experience in the wild. There is a large body of work looking at how gestures might be recognized and how usability can be improved through the use of multimodal feedback. This also includes work that examines the feasibility of using different parts of the body for interaction and control. Other work has looked at the user experience of interacting with multimodal interfaces in public spaces, focusing on the social acceptability of adopting new interactions. There has been a large push to bring multimodal user studies out of the lab and into the wild, with advocates highlighting the importance of creating real world experiences.

In this paper, we present an in the wild study of a multimodal mobile system used during participants' everyday commutes. The Gesture RSS Reader is a multimodal system employing gesture controls with speech and non-speech audio feedback designed to be used while on the move. Seven participants used this system during their daily commutes over 5 days with a total of 21.9 participant hours included in our analysis. The results of this study describe the practical technical challenges of making the system robust when users are walking, how context of use affected frequency of interaction, and how user experience and social acceptability developed for system adoption in the wild.

2. RELATED WORK

The current generation of mobile phones offers interaction through traditional menus and buttons as well interaction using sensors such as touch sensitive screens, accelerometers, or GPS that allow the user to interact in a variety of new ways. Research in gesture-based interaction explores a wide variety of issues, from improving sensing technology by disambiguating user intention [27], incorporating multiple modalities to improve usability [8], to designing continuous feedback to enhance interaction [26]. However, we will focus here on previous work that looks at enabling eyes and hands free interactions that can be used while on the go in real world settings. This focuses on two styles of interacting; using the hands and using the body. We will then go on to discuss previous work in user experience, social acceptability, and evaluating multimodal interfaces in real world settings.

2.1 Gesturing Using the Hands

Interfaces that make use of the hands allow users to perform a wide variety of gestures, from simple actions like taps and shakes to complex actions such as drawing and writing. Interactions that involve gesturing with the fingers on a touch sensitive screen do not always require visual attention. One early example of this is described by Pirhonen *et al.* [16]. They demonstrate the success of a mobile music player controlled eyes free through touch alone. They use a touchscreen PDA as the player, and allow users to change track or control the volume by drawing gestures with a finger on the touchscreen using the physical form of the device to

ICMI'11, November 14–18, 2011, Alicante, Spain.

guide the gesture. While screen-based gestures can be designed for eyes free use they restrict users in that touch sensitive surfaces must be in contact with fingertips for interaction. Outside of the touchscreens, there is a growing body of work incorporating inertial sensing into hand-based gestures using accelerometers within the interaction device. Oakley and O'Modhrain [14] describe an eyes-free device tilt system that uses tactile feedback to allow a user to maintain position in the system. The goal of their work was to provide a system that allowed users to interact with menus non-visually. Linjama and Kaaresoja [11] describe tapping the device in different places to gesture, allowing low effort, discreet interactions ideal for discrete action events with tactile feedback to alert the user to the completion of an action. These systems might allow for eyes free interactions but still require users to touch [27], shake [26], or tilt their device [7] using their hands.

Recently, the idea of touchless gestures is becoming more prevalent. The Magiwrite system uses a hand-mounted magnet to interact with a magnetometer in a mobile device [9]. Here, the user can perform three dimensional gestures in the air around the magnetic sensor without directly touching the device. Gustafson et al. [6] describe a mobile camera-based hand tracking system to entirely remove the device from interaction, again with the goal of interaction without the need to remove the device from the pocket. These examples describe a selection of the work that examines how users might interact using hand-based gestures for eyes free interaction. However, they still require that the users' hands are free to hold the device or perform gestures around the device. This may not be possible if the user is carrying shopping bags, holding on to children or operating machinery as part of a job.

2.2 Body-Based Gesturing

The use of body-based gesturing, or those gestures that involve moving part of the body such as an arm or one's head, allows for hands and eyes free interactions in a variety of styles. Work in this area has focused on allowing for subtle interaction, keeping hands free, and determining the usability of different body-based gestures for menu selection. Costanza *et al.* [2] describe an interface designed for subtlety using electromyogram (EMG) based interactions, where the system can detect muscle flexing on a user's upper arm. Other work in body-based gesturing has focused on keeping hands free without compromising control. For example, Rekimoto [20] designed GestureWrist, a system that recognises user gestures without encumbering the user's hands. Also very relevant is Oakley & Park's motion-based marking menu system [15], which relies on wrist rotation (roll) to allow menu selection in a hands-free manner. Other work includes a study of mobile head pointing using ego-centric nod gestures to select items spatialised around the user with 3D audio [12]. These studies show that using different body locations for input is possible, but have not yet looked systematically at interaction across the body. One notable exception is Rahman et al. [17] who extensively investigated wrist tilt input techniques showing the potential of the technique using either flexion/extension or pronation/supination if appropriate care is taken with the design. They did not, however, consider the social issues of using body-based interaction while mobile in public spaces.

The work described in this paper builds from of a series of studies examining body-based gesturing for mobile interaction. Previous work has examined wrist rotation [4] as a position control mechanism for moving a cursor on a phone. Participants rotated their wrist within a 90° workspace to move a cursor the length of the screen and click a button when over a target to select. Results showed that an accelerometer could successfully be used to target in this manner in static conditions but was less successful when used while walking. A similar study demonstrated that head tilting sensed using a head mounted accelerometer was a viable technique for a mobile targeting task [5]. Results showed mobile performance could be improved with the use of velocity control and a deadzone. Foot tapping was also investigated for a menu navigation task [6] using discrete tap events to navigate a hierarchical menu. This work demonstrated an entirely hands and eyes free menu navigation system that achieved high rates of accuracy in a seated laboratory-based study. These previous studies provide important design recommendations for body-based gesturing, including the use of a dead zone, discrete selections, and providing a guide to the limits of control using these gestures while on the move. Here we build on these laboratory results by evaluating the techniques in 'the wild,' with an application designed from the ground up to be used hands and eyes free.

2.3 User Experience and Social Acceptability

As hand and body-based gestures become a part of mainstream mobile interaction, the experience of using these gestures and the social acceptability of performing gestures in public places becomes an important issue. Although gesture-based interface can be designed for discreet use, many gestures still have a high level of visibility. Previous work on the social acceptability of using gestures input has explored the issues of visibility and perceived meaning as important factors of acceptance. Rico and Brewster completed an on-the-street study where participants were asked to perform gestures in public settings and discuss their experiences [22]. Montero *et al.* ran a focus group study that looked at how the social acceptability of gestures is evaluated based on the visibility of the gesture with relation to the visibility of its effects [13]. These studies have focused on two important aspects of the user experience of performing gestures: the comfort and personal experience of the performer and the perceived opinions of spectators. Previous studies have also highlighted the importance of extended use and experience when making decisions about social acceptability [22]. This is an important aspect of acceptability because only during a longer deployment can users have the opportunity to explore the interaction technique in a variety of settings, experience the responses of divergent spectators, and appropriate the performance in their own way.

Much of the previous work in social acceptability has focused on user experience as a way of exploring how users make decisions about acceptability. Although the exact definition of user experience is still debateable, there are some aspects of user experience research that are widely agreed on. Firstly, that user experience is fluid and dynamic since the personal moods, social context, and past experiences are ever changing and clearly have and significant effect on the current experience [25]. Secondly, the user experience is inherently an individual and personal experience, albeit at times heavily influenced by others and social groups [10]. Because an individual user is the one creating the perceptions and responses that user experience is so interested in, this is where the experience exists. This study focuses on user experience by building on previous usability studies. Although user experience and usability are intertwined, this study looks first at qualitative experience and perceived success and accuracy, and secondly at quantitative usability metrics.

2.4 Mobile Evaluation in Real World Settings

Although the need for user studies while mobile or in the wild [23][22] has been indentified, there are still very few longitudinal deployments of multimodal mobile interfaces. However, there is significant work addressing how to deal with the issues of extended deployment and discussion of the benefits of longitudinal studies versus traditional lab based ones. When evaluating a mobile interface there are conflicting factors that we must consider. First of all, it is important to maintain control over the study to allow clear measurements to be taken to allow appropriate conclusions to be drawn. This is most easily done in the laboratory, where the different factors that influence the interaction can be measured and controlled. However, lab based studies do not give users a real world experience, which is key for understanding social acceptability. Sherwood *et al.* [24] address the issue of mobile usability advocating real world evaluation with instrumentation, monitoring and questionnaires for feedback. The advantage of real world evaluation is to improve the validity of the findings. Experimental participants will be facing the same challenges as the real target user group and will experience the same issues with the interactions. The concept of instrumented usability described by Crossan *et al.* [4] is a powerful tool to gather data for studies in the wild where constant observations or experimenter interventions are not practical. For our study, we instrument the participants with sensors that are later used to gain insight into the context of use of the mobile device when analyzing the interactions. This could be used for high level classification of location or current mode of transport, or low level classification of moment-to moment actions such as the current phase of gait [[1], [3]] for interactions while walking. The study we present in this paper combined instrumented usability techniques with qualitative interviews to identify context of use to better understand when and where and why users choose to interact in public spaces.

3. The Gesture RSS Reader Study

The study we present here used instrumented usability techniques to identify context of use to better understand when and where users choose to interact in public spaces. The study was designed to take place during our participants' daily commutes to provide a real world experience of interacting with a multimodal interface while moving through different public spaces. In designing the system and our user study, we were guided by three goals;

- Support hands and eyes free interaction while mobile;
- Create a usable gesture experience;
- Study usability and experience in the wild.

The Gesture RSS Reader application was designed to allow users to browse and listen to news feeds while on the move. We chose to create an RSS reader because reading news stories or social feeds is a common activity for commuters who walk, drive or take public transport to work. Additionally, this application requires continuous browsing and interaction unlike other applications areas that we explored, such as a music player, that only require initial or sparse interaction.

To support hands and eyes free interaction while mobile, this system utilized body-based gestures combined with audio feedback. The system used a hierarchical menu system where users could browse through news sources, such as the BBC, select a news category, such as Sport, and browse through a list of available news items to choose one to listen to. Each menu item was read out as speech and navigation was completed using wrist gestures. These gestures could be done even if the hands were holding bags or other items that might be carried when our participants we commuting to or from work.

For audio feedback, we used the Microsoft Speech API to read out menu items and news feed text through standard earphones. Each user could choose both the voice and the rate of playback. There was also non-speech audio played for confirmation of each 'Selection' or 'Back Up' action. The user was alerted to the state of the system through a low volume ambient background sound played when gesture interaction was active with silence when gesture interaction was not active. This was controlled by the user with a gesture to initiate or end interaction, known as a gating gesture.

Command	*Action*	*Hand and sensor*	*Gesture*
Next	Move to the next item	Dominant / Accelerometer	Rotate right wrist clockwise
Previous	Move to the previous item	Dominant / Accelerometer.	Rotate right wrist anti-clockwise
Select	Move down hierarchy or select node	Dominant / Accelerometer.	Shake right wrist
Back Up	Move up hierarchy	Non-dominant / Accelerometer	Shake left wrist
Gate	Turn on/off gesture recognition	Non -dominant / Magnetometer.	Rotate left wrist palm up

Table 1. The available commands in the system, their meaning within the program, how they are sensed, and the gesture required to perform them.

Browsing the hierarchical menu and selecting menu items could all be completed non-visually using wrist gestures. The system was activated using a gating gesture (turning the left wrist palm up), such that interactions would only occur when the user performed the gating gesture simultaneously to one of the action gestures. This simple gating motion meant that the system did not respond to the regular movements of walking or riding on public transport as inputs to the application. Table 1 shows the name of each interaction technique, the action associated, the sensing technique, and a description of the gesture.

In order to interact with the system, the user must first perform the gating gesture to begin. Then, a user may browse the hierarchical menu by rotating the dominant wrist clock-wise or anti-clockwise respectively to move down or up a menu. Menu selections can be made by performing a short rapid shake with the dominant wrist. Once a new item is selected, it is read as speech output. Figure 1 shows a user interacting with the system.

3.1 Creating a Usable Gesture Experience

To create a usable gesture experience, we had to consider both the user experience of performing gestures as well as the accuracy of our gesture recognition. We drew from previous work to choose which gestures to use, to design the actions the gestures would correspond to, and how the gestures would be recognized.

Because we designed our system to be used in public spaces, the user experience and acceptance of the system would depend heavily on selecting socially acceptable interactions. In previous work [22], both wrist rotation and foot tapping have been identified as socially acceptable body-based interactions. Because foot tapping would be impractical when walking, we chose to use wrist rotation and wrist shaking for hands-free interaction. Because this system would be used mainly while walking or in transit, previous work indicated this would be relatively acceptable, although different personalities can often produce outliers.

In selecting how gestures mapped to actions in the RSS Reader, previous work [6] suggested that discrete actions, rather than continuous ones, would be more usable in a mobile setting. Therefore, all menu browsing actions used discrete up/down commands where one action corresponded to moving a cursor one item up or down on the menu. Other gestures where mapped to actions such that the most frequent actions would be easiest to perform with the dominant hand and less frequent actions would be more difficult or be performed with the non-dominant hand. For example, "Back Up" was mapped to a non-dominant hand shake gesture since this was likely to be a less frequent interaction.

Figure 1. One participant using the system. A UMPC is in a backpack while two wrist bands contain sensors used for input. An additional sensor is worn on the back of the collar.

Previous work has also identified anxieties users have when interacting with a multimodal system in public [21], with false positive recognition being one of the top concerns. This is also a technical issue, known as the segmentation problem [24], where there is difficulty identifying the difference between everyday movements and actions directed towards the system. A common solution to this is incorporating an on/off switch that controls when interaction can occur. Although this is often done using a button or similar technique, we incorporated a gating gesture to keep our system hands-free. An ideal gating gesture would be one that is simple, quick and comfortable to perform and hold, unobtrusive but happens rarely in everyday life. It must be robust at filtering out general body movements, be discreet enough that a user will feel happy performing the gesture in public, and be physically comfortable enough to sustain during short periods of interaction. For our gating gesture, we exploit the sensor attached to the user's non-dominant wrist. No fine grain control is required as a gating gesture will simply return a binary on/off response. The 'gating on' command was to hold the wrist with the palm facing upwards. There are a number of discreet ways this gesture can be performed and it is not uncomfortable if held for short periods of time.

3.2 Equipment

The equipment used to run the Gesture RSS application is shown in Figure 2. The logging, processing of sensor data and gesture recognition was done on a Samsung Q1 Ultra, an ultra mobile tablet PC (UMPC). For interaction and context sensing, three JAKE sensor packs were used (code.google.com/p/jake-drivers). The JAKE, shown in Figure 2, is a very small, lightweight sensor pack (10x18x10mm, weight approx 7g) that connects through Bluetooth to the UMPC. It contains a three axis linear accelerometer and a three axis magnetometer that allow us to sense movement and heading information in three dimensions. The two interaction JAKEs were placed inside wrist bands, shown in Figure 3, and worn around each of the users wrists. The third JAKE, used for context sensing, was worn on the collar.

The UMPC is slightly larger than a phone and too large to fit in a pocket, so not ideal for a truly mobile interaction as it had to be carried in a rucksack. We chose it as it offered several advantages over a smartphone. Firstly, it runs a standard version of Windows XP. Critical to the success of this application was speech synthesis of the news stories as the user must be able to interpret the news feeds non-visually. The Microsoft Speech API is mature, works with many different voices and the UMPC allowed us access to this API. Secondly, the UMPC allows a reliable simultaneous connection for the three Bluetooth sensor packs at a sampling rate of 100HZ without processor overload or issues with dropped packets.

Figure 2. All the equipment used to interact with the Gesture RSS system. At the top, three JAKE sensor packs with two wristbands and one clip to attach the device to a user's wrists and back of the collar. The sensors connected to the UMPC (carried in a rucksack) via Bluetooth.

Figure 3. The JAKE sensor pack (left). A user shown wearing the JAKE (right) as used during the study.

3.3 Gesture Recognition

To recognize gestures in the Gesture RSS Reader, we use three basic inertial sensing techniques: orientation estimation using a magnetometer, orientation estimation using an accelerometer, and shake detection using an accelerometer.

For the gating gesture, our main goal was to make this system robust to the noise of general walking or transportation movements. This is not a trivial task since, unlike the discrete actions of back/next, the gating gesture must be maintained while walking and interacting. Here we use the magnetometer in a JAKE sensor pack (see Figure 3) to determine whether the device is upside down. The magnetometer signals are generally exocentric so less suited to mobile interaction where a change in the user's heading will affect the signal from the sensor. Here though, only the vertical component of the magnetic field is used. This is relatively robust to changes in heading and is not affected by lateral motion of the sensors. By using a simple thresholding algorithm we define a range of values that correspond to the sensor on the non-dominant wrist being upside down. Gestures with the dominant hand are only recognised when they are gated in this manner.

The accelerometer in the JAKE was used for orientation estimation with a low-pass filter allowing frequencies up to 2Hz used to remove noise from the signal. The tilt is then estimated using the sensed acceleration with respect to the fixed gravitational acceleration. Based on previous work [5], we designed our gesture recognition with a dead zone of control defined between -40 and +40 degrees where no interaction events occur. Once the orientation angle crosses the 40 degree threshold, a 'Next' menu item event occurs. When crossing the -40 degree threshold, a move to 'Previous' menu item event occurs.

For shake detection, we first approximate the rate of change of acceleration by taking the difference between successive sensor values. We then calculate the magnitude of this signal and high pass filter it using a Butterworth filter to remove the low frequencies in the signal that correspond mostly to non-shake movements. The remaining high frequencies correspond to fast movements, collisions or rapid changes in direction that would be characteristics of tapping or shaking. Shakes are then detected using a thresholding algorithm. Evaluation of this system led to a minimum length of shake set to 0.4 seconds before a second shake is detected. This was to avoid the potential problem of a single shake event being recognized twice. Because the "Back Up" interaction had to be completed while maintaining the gating gesture, the gesture recognizer included flexibility in recognizing a non-dominant hand shake in case performing this command interfered momentarily with gating. The system was designed to respond to gesture events that happened a fraction of a second after gating was turned off. This meant that if the shaking interfered with the gating, the back up gesture was still recognized.

3.4 Usability and Experience in the Wild

The goal for this study was to run the application in real world settings over an extended period of time. To give users a chance to interact with the system while on the move, we deployed the system during users' everyday commute over at least 5 days.

This style of study presents significant experimental challenges that need to be addressed to run the study successfully as well as collect appropriate data to analyze the results. Therefore, we gathered a variety of quantitative and qualitative data from usage logs, sensors, interviews, and questionnaires. In previous laboratory studies, researchers were able to use pointing accuracy and time to target metrics to give a measure of performance of each technique. Here, we have no set task and the users were free to customize their news feeds. Allowing users to browse and locate stories of interest is far more appropriate and engaging in the real world than specifying a regimented task to. We therefore decided not to have a set task for the participants to perform.

Participants were given instructions to use the system as much or as little as they wanted during their standard commute over the span of at least 5 days. No information was collected on any of the topics accessed by participants for privacy reasons. However, quantitative usage data were collected through usage logs and a collar mounted accelerometer that was used for context sensing. This method of determining context based on gait phase [3] has been a successful metric in previous work. The quantitative data was combined with quantitative interview and questionnaire data to make the context data more complete and provide insight into user experience.

As this task was likely to lead to a different experience from the users' normal commute, it was important that the equipment was used for an extended period of time. Previous studies have shown that two or more experiences [22] represent the point where users make decisions about social acceptability and user experience, so we extended this to five experiences to capture additional usage data and confirm these earlier results. An important concern for any system employing unusual techniques or technologies is how comfortable people feel using the system in public. This is difficult to measure in a purely quantitative way. For this reason, we use a combination of techniques to measure user acceptance. *Post hoc* interviews and questionnaires were presented to the users to gain insight into the social implications of using the system in public. As well as these measures, prior to the first session and after the third session and the final session, participants were asked to rate their experiences on a series of Likert scales and answer questions on perceived social acceptability.

3.5 Challenges

There were a number of challenges to be met for this study to succeed. First, there were four different pieces of technology (the UMPC and 3 sensor packs) that each needed to be charged at least every second day. The sensor packs needed to be placed in the correct wristbands and attached at the appropriate locations for the application to function correctly. Each sensor pack was labeled with a function and arrows demonstrating the correct orientation such that the participants could easily place the sensors themselves. Short daily meetings with the participants were used to ensure that the sensors remained charged and there were no issues with the technology. A decision was taken at the start of the study that the daily meetings with participants would also be used to resolve any interface issues they had. Much of the system was customizable, but features such as thresholds for rotation and shake recognition were initially standard for all users. These thresholds were adjusted if requested by the user. Although this obviously means that each participant had a slightly different experience, this also ensured that the user was not frustrated by interaction difficulties and allowed us to see the kinds and ranges of customizations needed to create a system usable by a wider audience in the real world.

4. RESULTS

A total of seven participants all took part in either five or six sessions (37 sessions in total), generating in total 11,544 interaction events for a total of 21.9 hours of commuting data. During this time 1536 stories were read and 6632 headlines were browsed. Figure 4 shows the interactions per minute during each session. Participants were recruited through email from the University mailing lists. Participants were limited to those whose normal daily commute was at least 30 minutes and who travelled either by walking or public transport. We discounted cyclists and drivers due to potential safety concerns. They were asked to use the interface for at least five sessions, although two participants completed a total of 6 sessions.

4.1 Where am I? Context and Interaction

When comparing different contexts of use, we group the data into walking or public transport. This was done through data gathered from the context sensing JAKE clipped to the users' collars combined with self-reported data from the participants collected during interviews. The data was categorised using a mixture of automated sensing techniques to identify features in the data and manual tagging completed by the experimenter using the self reported data for disambiguation of automatically derived contexts. There were 39 occasions where a user was walking, and 17 occasions where a user was on public transport. Walking was a more common context since only 3 of the 7 participants used public transport during their daily commute. The mean rates of interaction events were 10.0 interactions per minute (std. dev. = 6.8) while walking compared to 4.9 interactions per minute while on public transport (std. dev. = 2.7). A two tailed t-test was used to that interaction rates were significantly higher while walking (T_{54}=2.91, $P < 0.01$).

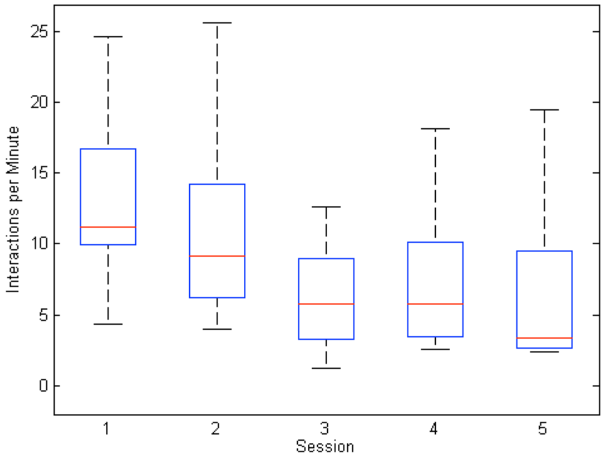

Figure 4. Mean number of overall interactions per minute over the five sessions.

The main reason for this significant difference between interactions on public transport and while walking is that spectators may be able to watch your actions while on public transport as opposed to only catching momentary glimpses when passing by on the street. This is discussed in more detail in the next section.

4.2 I'm Not Doing That: Social Acceptability

Even though the system included a very small gesture vocabulary, users experienced both positive and negative aspects of using the gestures on their daily commutes. The gestures were originally designed to be as discreet as possible to encourage acceptance [22] and it was therefore surprising when some of the gestures were deemed unacceptable in certain settings. However, participants developed different methods of performing gestures throughout the course of the study to improve their experience.

4.2.1 Strange Behaviour and Strangers

Although previous work in social acceptability has identified that strangers create situations with the lowest levels of acceptability, participants described two kinds of strangers with differing effects on social acceptability. Participants described their experiences of using this system on public transport as being significantly different from experiences walking on the street because the strangers around them could watch them for the duration of a bus or train ride, as opposed to momentarily passing them while walking. In both situations, participants became aware that observers might be noticing them but this was not an issue when the participant could simply continue walking and any unwanted attention quickly passed by. In comparison, users riding on a bus or train were aware that others might be watching them and the unwanted attention could extend for as long as the entire trip. One participant pointed out the people in places like underground or train stations are often very watchful and even suspicious of others. This participant felt extremely uncomfortable with this negative attention while on public transport, even doing something as simple as a wrist rotation. Another participant mentioned that the underground was the least comfortable setting because the gestures were particularly noticeable there due to constrained space.

The other interesting aspect was that participants were unable to predict their experiences in settings where spectators were sustained rather than transitory, where spectators might act in unexpected ways. The kind of interaction that one experiences with strangers and passersby seemed difficult for users to predict not only because those situations are often fluid and changeable but also because the appearance of the user and their attitudes towards the strangers' opinions varied widely. Several participants felt that spectators were becoming suspicious of their interactions of the system and were much more worried about attracting this kind of attention than others. On the other hand, one participant stated that "people look strange at you but you get used to and I didn't mind in the end". These differences in personality and appearance played a role in social acceptability, although more work would need to be done to look at this in greater depth.

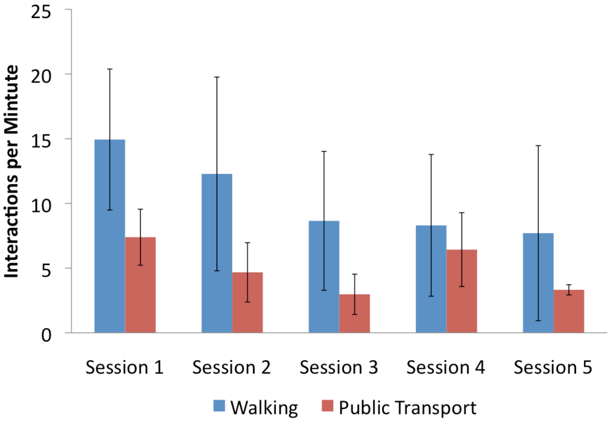

Figure 5. Interactions per minute grouped by context for all users across 5 sessions.

4.2.2 Appropriating Gestures: Usage Over Time

The gating gesture was in general well received by participants, and was the easiest gesture to appropriate into everyday movements by, for example, resting a hand at the appropriate rotation on a leg while seated in a bus, or holding a backpack strap with the non-dominant hand in a particular manner that looked natural but also turned the gating on. There were however some issues in specific situations that a laboratory-based study would not have detected. One user had issues on a rainy day when carrying an umbrella and carrying a bottle of water. In each instance rotation of the wrist has undesirable results that meant that for some forms of this technique cannot be considered truly hands free.

4.3 Mobility: Ease of Use While Walking

When moving between next and previous menu items, the data suggest a strong preference for moving in one direction. There were 7475 'Next' events compared with 1161 'Previous' events. This may be expected in a visual menu system where the user is conditioned to move down the menu due the visual layout of the system. As this system was used non-visually, there was no layout to guide the user. One possible reason for this is that the 'Next' event was more physically comfortable and natural to perform (users were required to rotate their right wrist clockwise). The right wrist has a larger range of rotation clockwise than anticlockwise, so "Next" gestures would be physically easier to perform than "Previous" gestures that utilized anti-clockwise rotation. In this system, the mapping of "Next" and "Previous" to clock-wise and anti-clockwise respectively meant that users strongly favoured "Next" events, which would typically be the most common action when browsing through a list of menu items. The fact that the system mapping matched existing practices is an important factor in creating an enjoyable gestural experience.

4.3.1 Thresholding Techniques and Mobility

Issues with gesture thresholds being set too high or low were a common complaint after the first session. During subsequent meetings with participants, it became clear that thresholds must be adjusted on an individual basis to provide comfortable and robust interactions. For example, no universally appropriate shaking threshold seemed to exist for all users, highlighting the importance of calibration and customization when using thresholds for gesture recognition.

Another unsolved issue involved the shake gestures being combined with rotation gestures. The menu browse and selection gestures were performed by rotating and shaking the right wrist respectively. In the initial training session, the users were instructed to shake with the palm facing down such that the sensor pack was on top of the wrist. When the user's wrist was rotated during a selection event, this could cause both a rotation and selection event to occur almost simultaneously. This was a common issue for participants particularly during the early sessions and resulted in the selection of an unexpected menu item. Clearly a danger with overloading multiple gestures on the same sensor is interference between the gestures. Due to the technology chosen for orientation estimation, it is impossible to separate out rotational and translational motions. By using additional sensors, it would be possible to improve recognition and solve this issue. For example, by combining the accelerometer signal with gyroscope and magnetometer data only shake actions that occur downwards, rather than to the side, could be recognized as a "shake" and rotations could still be successfully recognized using angle estimation.

5. DISCUSSION

The goal of this study was to complete a longitudinal deployment of a multimodal system in the wild that was usable both technically and socially while moving through a variety of public places. We deployed this system during users' daily commutes, with usage logs covering 21.9 hours of usage while walking and using public transport. During the course of the study, frequent meetings with participants meant we were able to identify and correct issues with usability and technical problems with the system as they arose. Most commonly, adjustments to gesture recognition thresholds had to be changed on an individual basis to create an algorithm accurate enough for each user. Users with an energetic style of walking reported more false positives and required higher thresholds to feel satisfied with the accuracy of the gesture recognizer. On the other hand, some users found it difficult to maintain the gating position or reliably complete a gesture and required thresholds to be lowered. These differences were partially due to physical attributes such as walking style or flexibility, but also heavily dependent on the user's willingness to perform larger or more performative gestures in public. Users who were less comfortable using the system in public would also have trouble successfully performing gestures if their movements did not produce enough acceleration to be recognized. Here, there is a clear tradeoff between successful recognition and an enjoyable user experience. However, designing interactions such that the technology supports both extravagant and subtle behaviour presents and interesting solution to this issue.

An important aspect of this study was that participants had the opportunity to perform gestures and interact with the system in a variety of real world settings. Previous work in the area of social acceptability and user experience has primarily focused on imagined or semi-controlled situations [13], [21], [22], whereas this study provides data about social acceptability and user experience of multimodal deployments in the wild. During interviews, participants described two kinds of strangers, sustained and transitory, creating two very different experiences. Montero *et al.* [13] describe how the visibility of actions and the resulting effects changes social acceptability based on the work of Reeves *et al.* [18]. Although this work identifies perceptions decreased social acceptability when actions are visible but the results are not, this is only an issue when you have sustained spectators. The discomfort caused in performing otherwise unexplainable actions in front of transitory strangers did not affect user acceptance and enjoyment of gesture interaction. This can be seen in the significant difference in the number of interactions per minute while walking as compared to while on public transport. While walking, users performed significantly more interactions than while on public transport. This is especially surprising when you consider that users typically have more idle time and fewer distractions while on public transport than while navigating a safe route while walking. During interviews, participants explained this by describing how uncomfortable it was to interact with the system knowing others might be watching you and could continue to watch you for the duration of the ride. This highlights the need for systems to provide ways of demonstrating interaction and expressing intent to spectators, an issue that has been identified in previous work and remains unsolved.

5.1 Design Recommendations

Based on these results, we present design recommendations with respect to the technical changes in recognition that came out of this study and the qualitative feedback from the user experience aspects of this study.

Recognition Flexibility – Users in this study often changed their performance of these gestures in order to increase their comfort and flexibility. One way of increasing the flexibility of gesture performance is to correct the relative angle of inertial sensing using a gyroscope. This allows users to perform the same actions in different ways. For example, wrist rotations can be performed with the wrist held parallel or perpendicular to the ground if the accelerometer values are normalized using the gyroscope.

Designing for Strangers – Users often described how sustained spectatorship was significantly less comfortable than transitory spectatorship. Users did not consider this before the study, showing that they found it difficult to imagine how different spectators might affect their experience. This shows the benefits of usability studies in the wild, and demonstrates an important aspect of context that must be considered for systems that will be deployed in public places where sustained or transitory spectatorship is likely.

Evaluation in the Wild – There are many examples here of results that would be difficult to discover in a lab setting. Despite the difficulties of evaluating in the wild, this work argues that these evaluations are important for systems that are likely to be used in a social context. To run such a study however, researchers should consider being flexible in their approach such that small issues resulting from unexpected situations can be resolved as the study progresses. For example, this study required that recognition thresholds be adjusted throughout the study for individual participants.

6. CONCLUSIONS

The results of this study demonstrate the successful use of body-based gestures for eyes and hands free interaction on the move. Much of the difficulty in deploying such a system is in creating gesture recognition that is robust enough for use while walking where sensor noise is high while also creating a socially acceptable and enjoyable experience. If gesture recognition requires extravagant movements, this will limit the acceptability of those interactions in daily life. This is due as much to context of use and personal comfort as to practical attributes such as flexibility and style of walking. However, these issues can be considered and designed for when selecting gestures and appropriately mapping these to actions in the interface.

Previous work has identified that the presence of strangers influences acceptability, but this work identifies the difference in comfort when interacting in front of sustained versus transitory strangers. In situations where strangers have the opportunity to watch an interaction occurring over a sustained period of time, such a during a train ride, participants were much less comfortable interacting through gesture. This is clearly shown in the significantly lower level of interactions per minute while on public transport versus while walking.

This study also puts to the test some recognition techniques in the wild that have been previously evaluated in lab settings. The use of a dead zone of control, discrete interactions, and sensing that combines accelerometers and magnetometers was successful while walking but required individual changes and adjustments for each participant. These changes improved usability and experience, and this kind of flexibility would be necessary for future deployments of real world systems.

7. ACKNOWLEDGEMENTS

This work was supported by a National Science Foundation Graduate Research Fellowship and the EPSRC funded GAIME Project (EP/F023405). Equipment was provided by Nokia.

8. REFERENCES

[1] Byrne, R., Eslambolchilar, P., Crossan, A., Health Monitoring using Gait Phase Effects, in PETRA'10: PErvasive Technologies Related to Assistive Environment, June 2010.

[2] Costanza, E., Inverso, S.A., and Allen, R., Toward subtle intimate interfaces for mobile devices using an EMG controller, in *Proc. of the ACM CHI*. 2005. 481 - 489.

[3] Crossan, A. Murray-Smith, R. Brewster, S. Kelly, J. Musizza, B. Gait Phase Effects in Mobile Interaction, *In Proc CHI 2005.*

[4] Crossan, A., Williamson, J., Brewster, S., and Murray-Smith, R., Wrist Rotation for Interaction in Mobile Contexts, in the *Proc. of Mobile HCI*. 2008, ACM Press.

[5] Crossan, A., McGill, M., Brewster, S., and Murray-Smith, R. Head Tilting for Interaction in Mobile Contexts. in *Proc. of MobileHCI*, 2009, ACM Press, (Bonn, Germany).

[6] Crossan, A., Brewster, S., and Ng, A. Foot Tapping for Mobile Interaction. in *Proc. of BCS HCI*, 2010, (Dundee, UK).

[7] Harrison, B., Fishkin, K., Gujar, A., Mochon, C., and Want, R. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proc. of* CHI '98, ACM, New York, NY, USA, 17-24.

[8] Kajastila, R. A. and Lokki, T. 2009. A gesture-based and eyes-free control method for mobile devices. In *Extended Abstracts CHI '09*. ACM, New York, NY, 3559-3564.

[9] Ketabdar, H., Roshandel, M., and Yüksel, K. A. 2010. MagiWrite: towards touchless digit entry using 3D space around mobile devices. *In Proc. of MobileHCI'*10.

[10] Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., and Kort, J. Understanding, scoping and defining user experience: a survey approach. In *Proc. of CHI 2009.* ACM, New York, NY, USA, 719-728.

[11] Linjama, J. and Kaaresoja, T. Novel, minimalist haptic gesture interaction for mobile devices. in *Proc. of NordiCHI*. 2004. Tampere, Finland: ACM Press, 457-458.

[12] 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 Proc. of ACM CHI* '06. ACM Press Addison-Wesley, 359-368.

[13] Montero, C. S., Alexander, J., Subramanion, S., Marshall, M. Would you do that? – Understanding Social Acceptance of Gestural Interfaces. *in Proc. of ACM MobileHCI*. 2010 (Lisbon, PR) ACM Press.

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

[15] Oakley, I. and Park, J. A motion-based marking menu system. in *Extended Abstracts of ACM CHI*. 2007. San Jose, CA: ACM.

[16] Pirhonen, A., Brewster, S.A., and Holguin, C. Gestural and Audio Metaphors as a Means of Control for Mobile Devices. In *Proc. of ACM CHI*. 2002. Minneapolis: ACM Press Addison-Wesley, 291-198.

[17] Rahman, M., Gustafson, S., Irani, P. and Subramanian, S. Tilt Techniques: Investigating the Dexterity of Wrist-based Input. In *Proc. of ACM CHI*, 2009, 1943-1952.

[18] Reeves, S., Benford, S., O'Malley, C., and Fraser, M. Designing the spectator experience. *In Proc. of CHI*, ACM Press (2005), 741-750.

[19] Rekimoto, J. Tilting Operations for Small Screen Interfaces. In *Proc. of UIST*. 1996, 167-168.

[20] Rekimoto, J. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Proc. of ISWC '01*. 2001.

[21] Rico, J., and Brewster, S.A. Gesture and Voice Prototyping for Early Evaluations of Social Acceptability in Multimodal Interfaces. In *Proc. of ICMI 2010.* ACM Press.

[22] Rico, J. and Brewster, S.A. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In Proc. of A*CM CHI*, 2010, ACM Press Addison-Wesley, pp 887-896.

[23] Sherwood, S., Reeves, S., Maitland, J., Morrison, A. and Chalmers, M. Adapting Evaluation to Study Behaviour in Context. *International Journal of Mobile Human Computer Interaction. 1(2).* 2009. 37-55.

[24] Strachan, S., Murray-Smith, R. and O'Modhrain, S. *BodySpace: inferring body pose for natural control of a music player.* In: CHI '07 Extended Abstracts. ACM.

[25] Vermeeren, A., Law, E., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. 2010. User experience evaluation methods: current state and development needs. In *Proc. of NordiCHI '10*. ACM, New York, NY, USA.

[26] Williamson, J., Murray-Smith, R., and Hughes, S. Shoogle: Multimodal Excitatory Interaction on Mobile Devices. in *Proc. of ACM CHI*. 2007. San Jose.

[27] Zhai, S. and Kristensson, P. Shorthand writing on stylus keyboard. in *Proc. of CHI* '03. Florida, USA: ACM, 97 - 104.