

# Chapter 14

## Gesture-Based Interfaces: Practical Applications of Gestures in Real World Mobile Settings

Julie Rico, Andrew Crossan, and Stephen Brewster

**Abstract** In the past, the design of gesture-based interfaces has focused on issues of gesture recognition with consideration of social or practical factors that affect the ability of users to perform gestures on the go largely missing. This work describes two important aspects of gestures design for mobile gesture and body-based interaction. First, this paper discusses the social acceptability of using gesture-based interfaces in the variety of locations where mobile interfaces are used. This includes a discussion of a variety of methods that can be used to evaluate social acceptability early on in the development process. Second, this paper discusses the practical implications of creating gesture recognition using accelerometer based sensing given the challenges of gesturing in mobile situations. This includes a discussion of body-based interactions and the scenarios where these might be used successfully.

### Introduction

Although gesture-based interfaces have become a popular topic for research since the success of the “Put-That-There” system in 1980 [1], gestures have not seen successful widespread use. Even though many smart phones now have the capabilities, through accelerometer based sensing, to recognize gestures, users seem unwilling to accept gesture-based interactions outside of the gaming or novelty applications. This paper discusses some reasons for this, including the social acceptability of gesturing in public and the practical implications of using gestures as part of a mobile interface.

---

J. Rico (✉), A. Crossan, and S. Brewster  
Glasgow Interactive Systems Group, School of Computing Science,  
University of Glasgow, G12 8QQ Glasgow, UK  
e-mail: julie@dcs.gla.ac.uk; ac@dcs.gla.ac.uk; stephen@dcs.gla.ac.uk

## Understanding Gestures

The use of gestures in interfaces has ranged widely from conversational interfaces with speech and gestures used together to interfaces using arbitrary gestures languages. However, when thinking about using gestures as part of an interface, it is important to consider what is being considered as a gesture, and what aspects of the working definition are important to the interface being designed. Although there is no clear or widely accepted definition of ‘gesture,’ the scope of gestures chosen for an interface has clear implications to the kinds of interactions that interface will facilitate. The following definitions of gesture seek to explore these issues.

Kendon gives a general definition of gestures as voluntary and expressive movements of the body [10]. Kendon’s definition of gesture includes what would commonly be thought of as “conversational” gestures, but does not provide guidance in determining which gestures belong to this set. Specifically, this definition fails to address how context, perception, meaning, and relationship to speech affect how an action is identified as a gesture. Kendon does discuss the role that perception plays into determining whether a given action is a gesture [11] even though this isn’t included in his definition. Kendon states that individuals are able to understand gestures and identify expressive intent simply by watching how the action is performed. This results in the circular definition that a gesture is simply an action that is considered a gesture by others. This kind of definition is even less helpful than others in determining what is a gesture. In order to narrow down the definition of gesture, Kendon’s work only considers those gestures that are used along with speech and are perceived by observers as part of the meaning of the speech. Accidental gestures and fidgeting are not included in his analysis. This difference between the general definition of gestures and the applied definition gesture can be seen in many of the following examples.

Cassell defines gestures as hand movements occur during speech [2] although she criticizes this limited definition. Cassell states that many systems that use gesture recognition focus on “gesture languages” rather than gestures that naturally occur with speech. However, she goes on to state that the communicative elements of gestures that occur with speech are important elements needed to create natural user interfaces. Väänänen and Böhm define gestures as “body movements which are used to convey some kind of information from one person to another” [17]. This definition is then further refined, including that information conveyed by gestures is easily understood by observers yet vague and implicit in nature. For example, an individual might use a gesture alongside the speech “I only want a little bit a milk in my coffee.” The gesture will communicate that a small amount of milk is desired, but does not say how much is considered “a little bit” by the speaker. For the purpose of their system, Väänänen and Böhm highlight that gestures used in human computer interfaces must have defined meanings, which is in direct opposition with gestures as they are used in daily life. In order to accommodate the technical restrictions of their gesture-based interface, a limited definition of gestures is used.

For the final system, a set of postures is implemented in place of more fluid or natural gestures. These hand positions included finger pointing, an open palm, and fist. While they are easy to perform movements, they are arbitrary movements and fall short of the original definition that revolved around information exchange.

A major issue with gesture definition is the difference between “conversational gestures” that occur naturally with speech and “gesture languages” that are commonly used in gesture-based systems [19]. Gesture systems such as the general use hand pose recognition system [20] describe a gesture language that can be used to define gestures for an interface. These gestures, however, would not fall into the everyday meaning of the term “gesture” but rather a “gesture system.” For example, Clara Rockmore’s “aerial positions” for playing the Theremin might be considered gestures using one of the above definitions, but this is more like a gesture system in that it is a set of hand positions used in a formally structured manner. These gesture systems might better describe how gestures are used in computing science, but this definition might take away the natural interaction that gestures originally sought to provide [18]. Some researchers argue that gestures cannot be examined without the linguistic context where they occur [12], but multimodal systems have both incorporated speech [1] and used gestures on their own [9]. Since speech plays a significant role in individuals’ understanding of gesture meanings [6], using gestures without speech could in some cases remove any natural or cognitive advantages that gestures might have over other forms of communication. Because of this, system designers must carefully evaluate their applied definition of gesture to ensure they are not stripping away the benefits of using gesture.

## **Gestures and Social Acceptability**

Previous work in multimodal interaction has mainly revolved around issues in recognition and detection, as well as advancements in sensing technologies. However, in order to design multimodal interfaces to provide a satisfying and enjoyable experience, interface designers must also consider the social acceptability of using these interfaces in public spaces. This is not simply an issue of “acceptable” or “unacceptable,” but a dynamic decision process that occurs in different social contexts at different stages of experience.

Our previous work in understanding social acceptability has utilized a variety of methods, including surveys [15], on-the-street user studies [15], experience prototyping and focus groups [16]. The on-the-street study required participants to perform a set of gestures in both a public outdoor setting and a private indoor setting over three repeated trials. The gestures were selected from a previous survey study to include both highly acceptable and unacceptable gestures. Multiple trials were completed to test the survey results over time, and observe if and how participants’ opinions of the gestures changed. The results of this study showed there were significant changes in opinion over time, and also gave insights into the reasons

why individuals liked or disliked gestures. The next study involved a survey that incorporated both gesture and voice commands. For the purpose of this survey, a categorization of gestures and voice commands was developed. The survey asked respondents to imagine different situations where they might use a gesture and voice command, giving clear scenarios including the location and audience. The results of this survey showed how gesture and voice ranked, with gestures being more accepted overall than voice commands. This survey also showed the affect that different audiences in the same locations have on social acceptability. This survey was followed by a focus group study, which expanded the gesture and voice lists and used a variety of low-cost prototypes developed for the groups. Participants were also grouped into separate age categories, with groups including participants aged 18–29 and groups including participants age 70–95. The results of this study demonstrated how participants imagined these interfaces working in the real world, and what anxieties they had about using them. This also showed an important difference in the way that older adults approached the gesture and voice commands as compared to the younger adults, resulting in different preferences and different concerns. In each of these studies, the importance of early evaluation and methods for evaluating large sets of gestures at low cost has been emphasized. The importance of gesture selection and evaluation before development can lead to more usable and enjoyable interfaces overall.

Although gestures can provide a rich interaction experience, using gestures in the context of an interface poses challenges. Researchers argue that gestures cannot be examined outside of the context of speech, yet they are used this way in many multimodal interfaces. The gestures commonly used in multimodal interfaces are often sets of arbitrary hand positions designed for the system, rather than something resembling our everyday conversational gestures. While this issue may occur due to technological constraints, it is also due to an incomplete understanding of how users understand and adopt gestures as part of an interface.

### ***Social Acceptability and Gesture-Based Interfaces***

Because gesture-based interfaces require users to adopt new and possibly strange behaviors in public spaces, the design of gestures used in these interfaces must take into account the possible meanings that gestures might have when used in this context and the social acceptability of performing that gesture in social settings. This means that designers must understand a variety of factors that influence the way that gestures will be used, including issues of performance, the influence of spectators, and the ways in which technology influences gesturing. These factors each contribute to an overall concept of social acceptability, where the acceptability of a certain action is being constantly evaluated and reevaluated by the performer given continuous feedback from spectators. Thus, social acceptability is not simply a matter of acceptable or unacceptable, but an ongoing decision process.

## Performance

Goffman describes every action that takes place in a public setting as a performance [1], and as mobile phones become increasingly integrated into our personal appearance, mobile phone usage becomes a performance. The variety of places where mobile interfaces are used means that performances are constantly changing and being reevaluated. With respect to gesture-based interfaces, the performative aspects of these interactions are accentuated given the often highly visible nature of these interactions. The required performance of a given gestures varies both on the gesture itself and the performer perceptions of that gesture. Evaluations of performance and perceptions of gestures can be done early in the development process through the use of surveys and video prototypes. Our previous work using surveys to evaluate social acceptability based on visual aspects of the gestures and the places they might be used proved a valuable evaluation tool [15].

## User Experience and Spectators

Although the exact scope and definition of user experience is still debated, it is clear that an understanding of an individual's thoughts, feelings and reactions to an interface are important factors that designers must consider [13]. With respect to gesture-based interaction, an understanding of the user experience of these interactions is especially important because these interactions often require users to try new and possibly unfamiliar actions. The experience of using an interface develops and changes over time as the user is continually exposed to the interaction and experiences it in different settings with different people. User experience, however, is essentially an individual experience [13]. Although other people and spectators heavily influence the social context where an interaction takes place, the decision to interact and the experience of doing so is a personal and individual experience.

Because mobile phones are commonly used in public settings, the presence of spectators and the performative aspects of multimodal interactions play an important role in user acceptance. Following from Goffman's assertion that all actions done in a public setting are performances [8], the performance of an interaction with a mobile device can range from unconscious, automatic actions to explicit and deliberate performance on a stage. The presence of spectators and their affect on the performers has a major influence on the type of interaction the performer will experience [14]. Because of this, performer and spectator roles should play an important part in the design of multimodal mobile interfaces and the evaluation of social acceptability. Early evaluations involving user experience can be completed using low-cost prototypes in focus group settings [16]. The prototypes used in this study allowed users to experience different gestures and voice commands without the need for sophisticated systems. The simple prototypes also allowed for a large number of interaction techniques to be tested at once. An on-the-street study also allows for early evaluations of social acceptability that take into account the affects which spectators have an evaluations of social acceptability [15].

## **Technology and Social Acceptability**

When individuals imagine the experience of performing gestures in public places, the role that technology plays in facilitating or obstructing an interaction plays an important role in how acceptable a certain interaction will be. In particular, failure scenarios, even when simply imagined rather than experienced, are enough to make an action unacceptable if the user perceives that these failures are likely to occur or likely to cause embarrassment. For example, our previous work shows that users were concerned that an interface would be unable to successfully recognize inputs, results in repeated and increasingly erratic motions to successfully complete a gesture [16]. While the error could happen to any gesture, errors executing a foot tap versus a shrug would lead to very different behaviors. Users were also concerned about the possibility of false positive recognitions by a system, and those gestures that users felt they were more likely to perform ‘by accident’ were less acceptable [16]. These issues can be discussed even when sophisticated sensing and technology is not part of a user evaluation. In our previous focus group study, participants brought up these issues and anxieties when presented with a variety of low-tech prototypes that portray the devices and sensors that would be part of an interface [16].

## **Social Acceptability**

Individuals make decisions about the social acceptability of their actions by gathering information about their current surroundings and using their existing knowledge [8]. Appropriate actions are then carried out and feedback is gathered through the reactions of observers. The process of experiencing an interface and determining the social acceptability of performing is an ongoing process that changes over time. The social acceptability of technology usage is not a simple matter of embarrassment or politeness, but a continuous evaluation that is influenced by a variety of factors. The factors that influence these decisions include the performance of the action, the experience the user hopes or expects to take away, and the perceived reactions of spectators. These can be evaluated using methods that can be completed early on in the development process and encourage evaluation of more interaction techniques than traditional user studies.

## **Body-Based Gesturing**

While social acceptability is an important aspect of designing gesture-based interfaces, we must also be aware of the practical and technical implications of using the gestures in the variety of places where mobile interfaces are used. Can we design gesture interfaces that are usable while the user is on the move; on a bus or train or while walking? The gestures should be robust to the noisy input channels that a mobile setting invariably creates, and allow a user to use the interface in a low effort,

comfortable and safe manner. When designing interactions for use in a mobile context, we cannot assume that the interactions are the user's primary task. For a user crossing a road or walking down a busy street, the primary goal is to navigate the environment safely, avoiding cars and other pedestrians. Users need to use their eyes to identify obstacles and safety hazards in the environment and avoid them. To a lesser extent users may also use their auditory sense, although the widespread use of music players while on the move suggests that the auditory sense is secondary to vision when navigating. To design an interface that can be used while on the move, there may be benefit in shifting the feedback away from the mobile device's screen and on to the other senses such as hearing and touch. From the point of view of the user's input into the system, again we must take into account mobile contexts. When away from the desktop and involved in other tasks, users may be encumbered, carrying bags or children or holding onto handles to stabilize themselves while standing on public transport. Interfaces designed to be used while encumbered in this manner could lead to very different ways of interacting with a device.

In almost all interfaces, including current gesture-based interfaces, interactions happen through the device. This could be through finger gestures on a touchscreen, pressing physical buttons on the device, or even by moving or orientating the device in a specific way. By shifting away from device-based interactions towards input techniques where the phone remains in the pocket, it is possible to envisage hands free interaction techniques. By employing a whole body interaction mechanism, we can start to take advantage of other areas of the body to provide fast, low effort interaction for mobile situations. For example, users could interact through wrist rotation, or nodding, pointing or shaking their head. For scenarios where the user may be seated or standing, for example on public transport for example, we might also consider foot tapping. These input channels will be lower bandwidth than the more traditional hand and finger-based interactions, but provide a mechanism to allow common actions to be performed with minimal effort and without the need to remove a phone from a pocket or bag. Here we describe three studies examining body-based interaction as an input technique for mobile interaction. In these studies we examine wrist rotation, head pointing and foot tapping as potential whole body interaction techniques for use while on the move. To allow interaction without removing the phone from a pocket, we use wearable sensor packs for input. The SHAKE is a small lightweight Bluetooth sensor pack that can be attached to different parts of a user to sense movement. It contains a three axis linear accelerometer, gyroscope and magnetometer that can be used for inertial sensing. In general, we can consider the inertial input from the user to be discrete action events or continuous steering control. For discrete action events, a gesture classifier continually monitors the stream of data from one or more sensors and attempts to identify preset patterns, such as movements corresponding to a tap, shake or preset trajectory. Once a pattern is detected and recognised, the appropriate control event is generated in the system. For a continuous control interaction, the user controls actions that happen fluidly over time, for example steering a cursor through a menu. A common continuous control input technique, which has been used for this work,

uses orientation estimation of the sensor pack. The orientation of the sensor in one or more axes is used to control the position of a cursor within a workspace.

### ***Testing Body-Based Discrete Action Event Control***

We tested foot tapping as a mobile interaction technique. Foot tapping is a common action that a user may perform while listening to music for example so does not require unusual movements. We tested the technique using a menu navigation task over a two level hierarchical menu where the root nodes represent common tasks performed on a mobile device. As well as a visual representation, the menu item name was read out as the user moves over an item using high tempo speech, known as spearcons [18], to present the audio. The menus were cyclical such that the currently selected item loops at the bottom and the top of the menus. The final item for each of the sub-menus was 'back', which returns the user to the top of the root menu. During the experiment, participants were asked to complete a set of menu selection tasks, with each task being prompted using a spearcon through the headphones to select a specific menu item. When the spearcon for the menu item was played, the user would then navigate to and select the menu item requested. To navigate through the menus, users tapped their right foot to move down through the menu and tapped their left foot to make a selection. Taps were detected using a high pass filtered accelerometer signal from a sensor pack attached to each foot. For comparison, both a visual control condition and a pocket condition were also completed. In the visual condition the participants were seated and held the phone in their dominant hand. To navigate they used the up/down keys on the phone keyboard, while a selection was made using the central select key. For the in pocket condition, the participants all wore the same jacket with an inside breast pocket. Before each selection, the phone was placed in the pocket. When prompted, the participants removed the phone from the pocket, and navigated to the appropriate menu item as with the visual condition. Twelve users navigated through the menus using all three conditions in a counterbalanced order.

Detailed statistical analysis of the results can be found in [5]. When examining the overall results, it is clear that as a general mobile input technique, foot tapping by itself cannot be considered superior to the traditional mechanisms. Both the visual condition and the in pocket condition demonstrated higher accuracy and faster selection times than the foot tapping condition. This is hardly surprising given the high number of taps or button presses to reach some of the menu items. Additionally, having access to the visual display of the menu allows the user to scan all the menu items quickly and move to the correct location without having the necessity of listening to each item each in turn. If we examine the trends in the data however, it is possible to see where foot tap could provide some benefit. Linear regression of the data suggests that for less than five taps, foot tap can be faster than removing the device from the pocket. In the case of common or simple menu selections, the benefit of leaving the device in the pocket would add to the benefit



of quicker selection times. There are also social scenarios where users might not be willing to remove the device from their pockets. For example, in a crowded area where space is limited or theft is common users may feel more comfortable controlling their devices using foot tapping while keeping the device securing stowed.

### ***Testing Body-Based Continuous Control***

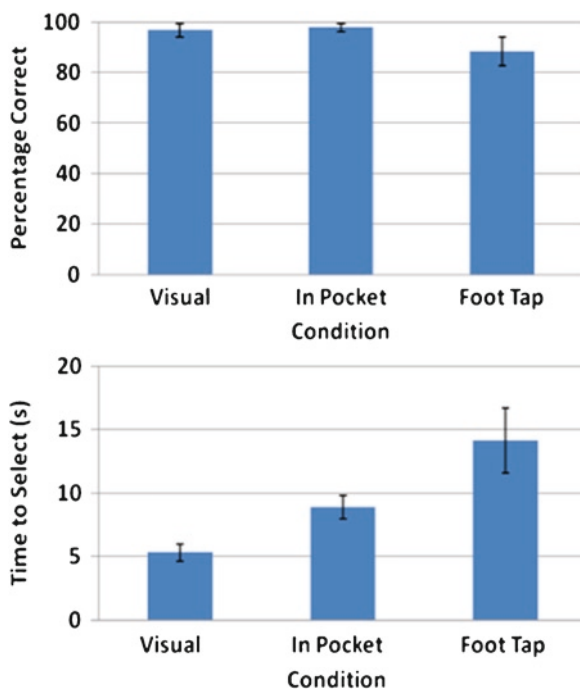
In order to apply whole body interaction for continuous control, we investigated two areas of the body to provide non-hand-based control in mobile settings; wrist and head input. To evaluate the feasibility of the techniques we use used a Fitts' Law paradigm [7], a common method of characterising performance in a one dimensional targeting task. In this type of evaluation, participants repeatedly move between two targets of varying widths and separations with movement time and accuracy used as metrics of performance. As these interfaces are designed to be used while mobile, we tested targeting performance both while static and while walking a figure of eight route.

#### **Wrist Rotation**

Wrist rotation was sensed using the accelerometer of a SHAKE sensor pack attached to the user's wrist. The forearm was held approximately parallel to ground and rotated in a 90° workspace. For a right handed user, palm facing down corresponded to the left of the workspace, and palm facing left corresponded to the right of the workspace. Participants viewed their interactions on phone screen held in one hand. A visual display was used so that results would not be affected by a potentially poor choice of non-visual interface design. Participants selected targets by pressing a button on the phone. A button press was used so that results would not be affected by a potentially difficult or inaccurate gesture-based selection mechanism. This allowed us to investigate if wrist rotation was effective for input without other factors influencing the results. Twenty-four users performed the study both a static and a walking condition. Detailed statistical analysis can be found in [3].

Results showed that while walking, users were significantly slower to target and significantly less accurate. Figure 14.1. Percentage of correct selections (top) and mean time to select in each condition (bottom). Error bars show one standard deviation.

Figure 14.2 illustrates the effect that changing target width and separation had on participants mean movement time and accuracy while static and mobile. In the static condition, participants achieve a high level of accuracy (~90%) for the targets of 9° of rotation wide and larger. This suggests that participants could successfully target using wrist rotation. The walking condition was however was significantly harder for all participants who were both slower and far less accurate than in the static condition. These results for the walking conditions applied to all target width

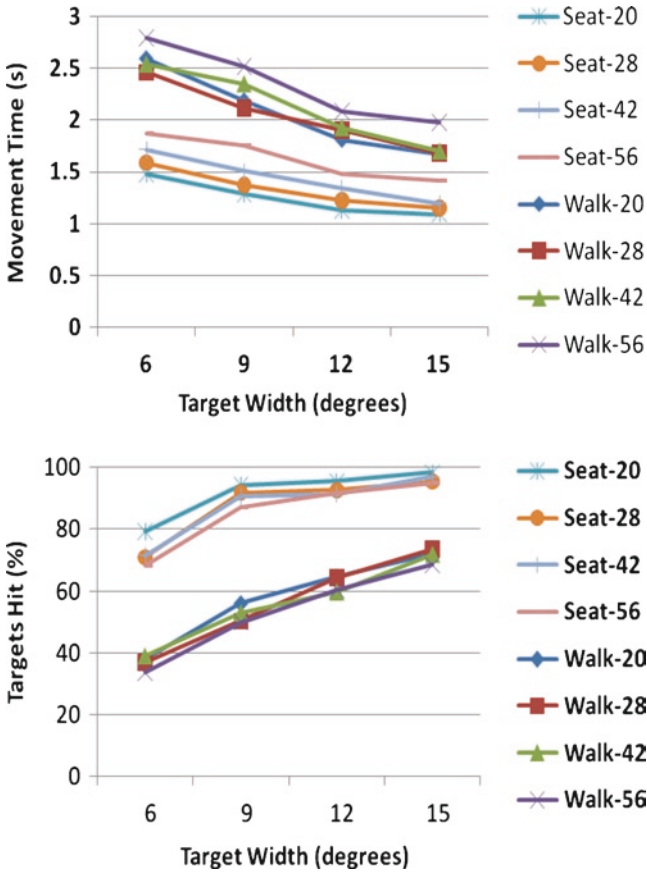


**Fig. 14.1** Percentage of correct selections (*top*) and mean time to select in each condition (*bottom*). Error bars show one standard deviation

and separation combinations. It is also interesting to note that the accuracy values for this condition are less than 80% in all cases. All participants commented on the difficulty of the task and expressed low confidence in their performance. The accelerometer signal contains both the tilt from the user's targeting, and noise generated by the walking behaviour, which caused the cursor to oscillate in time with the user's walking speed making it difficult to target.

## Head Tilt

In order to examine the possibility of continuous control using the body, we also completed a study using head tilt for interaction. In this study, users controlled the cursor by tilting their heads left or right within a range of  $\pm 40^\circ$ . This was estimated through a sensor pack attached to a cap worn by the user. Two different mechanisms were used to control the cursor. In the first, a position control mechanism was used where the position of the cursor changed linearly with respect to the angle of the participant's head tilt. The cursor positions correspond to head positions, where holding your head at a set angle will move the cursor to the corresponding place. The second, a velocity control mechanism was used where the velocity of the cursor



**Fig. 14.2** Movement time (*left*) and percentage of targets hit (*right*) for  $C_{SE}$  the seated and  $C_{WA}$  walking conditions for all width-separation combinations

movement was determined by the angle of head tilt. The greater the tilt, the faster the cursor moves.

For the position control mechanism, the central head position corresponded to the cursor being in the centre of the phone screen. By tilting their heads left, participants could move the cursor left, with the leftmost screen position corresponding to a tilt angle of  $40^\circ$  (and *vice versa* for tilting right). For the velocity control conditions, the velocity of the cursor changed linearly with respect to the head tilt angle with the central head position corresponding to a stationary cursor. Unlike the position control condition, using velocity control allows robustness to be built into the control by using a dead zone as there is no one-to-one mapping between head tilt and cursor position. A dead-zone of  $\pm 5^\circ$  was included to provide some robustness to noise from normal small head movements. Within the dead zone the cursor did not move. Similarly to the wrist rotation study, users interacted both while static and mobile. Detailed statistical results can be found here [4].

There were significant differences found in the accuracy data between the static and mobile conditions using both position and velocity control. Position control did consistently exhibit lower movement time than its velocity counterpart, and was significantly more accurate under the standing condition. However, position control was significantly worse than velocity control when mobile, with performance deteriorating significantly under the walking condition. There is the trade-off with cursor gain in the velocity controlled condition with a low cursor gain giving high accuracy and a longer target time, and high cursor gain giving a potentially faster time to target but lower accuracy. It will be possible for interface designers to shift performance between the two extremes by tuning this gain parameter.

In the static conditions, participants achieved a high level of accuracy, with approximately 80% for the targets with 7° of rotation wide and over 90% accuracy for larger targets. This suggests that participants could successfully target using head tilting. The walking conditions were however reported to be far more difficult for all participants who were both slower and far less accurate than in all other conditions for all target width-separation combinations, and particularly so for the position control condition which was significantly poorer than the others. The tendency for the cursor to oscillate with the walking was again noted. For the velocity control condition, the effect of noise seems to be reduced. This could be due firstly to the inclusion of a dead-zone around the central position making it easier for the participant to stop the cursor, and secondly since it is the velocity that is affected by the noise and not the actual cursor position.

## Conclusions

There are many issues that have arisen out of this work in gesture-based interfaces that may also have an effect on practical usage but that have not yet been thoroughly investigated. With respect to social acceptability, factors such as appearance of the performer and cultural context may play a large role in social acceptability. Goffman presents the idea that appearance plays a large role in the types of actions that both performers and observers find acceptable [8]. With respect to gesture-based interfaces, the same may hold true and manifest itself through surprising and unexpected actions and responses from users. Another important issue revolves around the adaptation and appropriation of gesture-based interfaces over time. This includes not only the changes in performance from a sensing and detection point of view but also the motivations and methods that bring about that change. The types of changes in performance that users would create in order to make an interface work practically and socially in different contexts of use is relatively unknown, which makes it difficult for designers and implementers to add the right kind of flexibility and customization to an interface. Because it is important for users to feel comfortable and in control while using an interface, flexibility in personal performance while maintaining accuracy could greatly improve the usability of gesture-based interfaces overall.

Overall, this work focuses on practical usage of gestures as part of a mobile interface. We argue that successful design of gesture-based interfaces requires not only a consideration of the social acceptability of gesturing in public, but also a practical consideration of the usability and robustness of using gestures on the move. These issues can and should be addressed at different stages of development, from early conceptualization and prototyping to development, usability testing, and deployment. With a combination of prototyping with divergent interaction techniques, early evaluations of social acceptability, consideration of constraints on vision, hearing, and movement while on the go, the widespread adoption of mobile gesture-based interfaces may soon become a reality.

## References

1. Bolt, R.A.: “Put-that-there”: voice and gesture at the graphics interface. In: Proceedings of SIGGRAPH 1980, pp. 262–270. ACM Press, New York (1980)
2. Cassell, J.: A framework for gesture generation and interpretation. In: Cipolla, R., Pentland, A. (eds.) *Computer Vision in Human–Machine Interaction*, pp. 191–215. Cambridge University Press, Cambridge (1998)
3. Crossan, A., Williamson, J., Brewster, S., Murray-Smith, R.: Wrist rotation for interaction in mobile contexts. In: *The Proceedings of Mobile HCI 2008*. ACM Press, New York
4. Crossan, A., McGill, M., Brewster, S.A., Murray-Smith R.: Head tilting for interaction in mobile contexts. In: *Proceedings of Mobile HCI 2009*. ACM Press, New York
5. Crossan, A., Ng, A., Brewster, S.: Foot tapping for mobile interaction. In: *The Proceedings of BCS HCI, Dundee* (2010)
6. Eisenstein, J., Randall, D.: Visual and linguistic information in gesture classification. In: *SIGGRAPH: ACM Special Interest Group on Computer Graphics and Interactive Techniques*, San Diego. ACM, New York (2007)
7. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* **47**(6), 381–391 (June 1954)
8. Goffman, E.: *The Presentation of Self in Everyday Life*. Penguin, London (1990)
9. Jin, Y., Choi, S., et al.: GIA: design of a gesture-based interaction photo album. *Pers. Ubiquit. Comput.* **8**(3), 227–233 (2004)
10. Kendon, A.: Gesture. *Annu. Rev. Anthropol.* **26**, 109–128 (1997)
11. Kendon, A.: Current issues in the study of gesture. In: Nespoulous, J.L., Perron, P., Lecour, A.R. (eds.) *The Biological Foundations of Gestures*. Lawrence Erlbaum Associates, Hillsdale (1986)
12. Kendon, A.: Language and gesture: unity or duality? In: McNeill, D. (ed.) *Language and Gesture*. Cambridge University Press, Cambridge (2000)
13. Law, E.L., Roto, V., Hassenzahl, M., Vermeeren, A.P., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: *Proceedings of CHI 2009*, pp. 719–728. ACM Press, New York (2009)
14. Reeves, S., Benford, S., O’Malley, C., Fraser, M.: Designing the spectator experience. In: *Proceedings of CHI 2005*, pp. 741–750. ACM Press, New York (2005)
15. Rico, J., Brewster, S.A.: Usable gestures for mobile interfaces: evaluating social acceptability. In: *Proceedings of CHI 2010*, pp. 887–896. ACM Press, New York (2010)
16. Rico, J., Brewster, S.A.: Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *Proceedings of ICMI 2010, Beijing*. ACM Press, New York
17. Väänänen, K., Böhm, K.: Gesture driven interaction as a human factor in virtual environments – an approach to neural networks. In: Earnshaw, R.A., Gigante, M.A., Jones, H. (eds.) *Virtual Reality Systems*. Academic, London (1993)

18. Walker, B. N., Nance, A., Lindsay, J.: Spearcons: speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the 12th International Conference on Auditory Display (ICAD2006). Department of Computer Science, Queen Mary, University of London, London, pp. 63–68 (2006)
19. Wexelblat, A.: Research challenges in gesture: open issues and unsolved problems. In: Wachsmuth, I., Fröhlich, M. (eds.) Proceedings of the International Gesture Workshop, 17–19 September 1997. Lecture Notes in Computer Science, vol. 1371, pp. 1–11. Springer, London (1997)
20. Zhenyao, M., Ulrich, N.: Lexical gesture interface. In: Fourth IEEE International Conference on Computer Vision Systems. IEEE (2006)