

Multimodal Interactions for Expressive Interfaces

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
[stephen, rod, ac, yolanda, julie]@dcs.gla.ac.uk / www.gaime-project.org

This paper presents an overview of a range of gestural, audio and tactile interaction techniques that can be used to create expressive interfaces that allow rich communication while reducing the demands on the visual display. We also discuss issues of social acceptability and how these new kinds of interactions will be accepted by different types of users in different types of settings.

Keywords: Multimodal interaction, gestures, haptics, tactile feedback, non-speech audio, social acceptability.

1. INTRODUCTION

Most mobile devices 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.* [13] 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]. The device moves as the user moves causing larger and more variable targeting errors and increasing the time to tap on a target. This is particularly frustrating in situations requiring many clicks such as using an on-screen keyboard to enter commands. Alternative techniques are needed for controlling mobile devices that allow a richer, more expressive interaction for future mobile interfaces.

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 new forms of expressive, multimodal input and output, for example gesture input from other locations on the body combined with three-dimensional sound and tactile displays for output.

Little work has gone into making input and control 'hands free' for mobile users. New forms of rich, expressive input are needed if the keyboard is not available. 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 and new methods of display are needed when eyes are busy.. 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 [16] [15]. Many of these use very simple 3D audio displays, but, with careful design, spatial audio could provide a much richer 'eyes-free' display space.

Tactile feedback also has possibilities for freeing up the eyes when mobile and giving another rich form of output to work with audio feedback. The whole of the body surface is available for information display. Sensory substitution has a long history in the area of accessibility [10] 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, but tactile feedback can be much more expressive.

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. INTERACTION TECHNIQUES

2.1 'Hands free' gestures for input

Gesture input has been successfully incorporated into both research and commercial devices, with standard mobile phones now incorporating accelerometers and magnetometers that could be used for mobile input. However, many previous projects have concentrated on gestures done with hands or fingers. The body itself is very expressive so our work is investigating how we can use different body locations for input.

Rekimoto describes GestureWrist in [14], 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.* [19] 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 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.

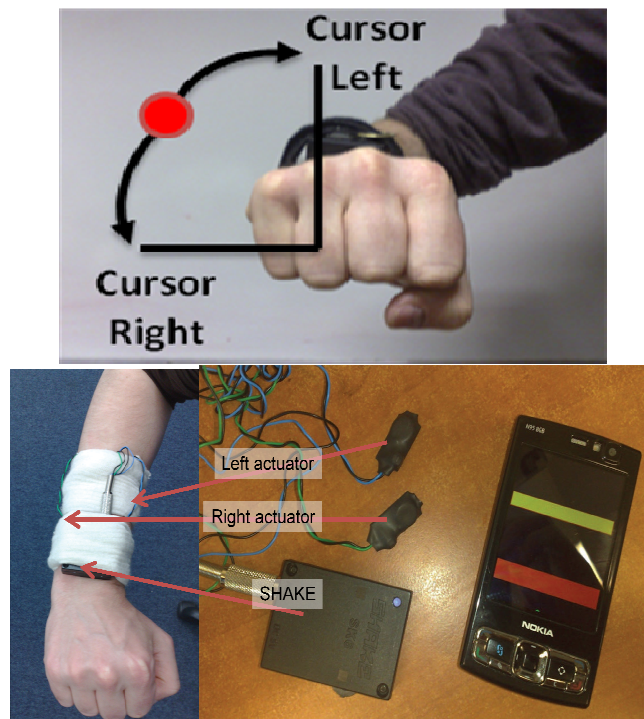


Figure 8: Wrist rotation for input. Left shows the range of rotation used, right shows tactile actuators and phone display.

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 an accelerometer 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 8**). Results showed that it was an effective input technique with users able to achieve 90% accuracy when selecting targets of 9° [8].

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 [7]. 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.

2.2 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. Expressive 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 richer 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 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. The technology to render spatial sound is now becoming common in mobile devices, with the Nokia N95, for example, containing a 3D sound rendering engine (see Figure 2).

Sounds can be presented as coming from around the user in a plane around the head (Figure 3), with different applications having different parts of the audio space, creating 'Audio Widows' [6]. Schmandt at MIT has done some significant work on 3D audio displays, in particular NomadicRadio [15] that used 3D sound for reminders and notifications in a mobile device.

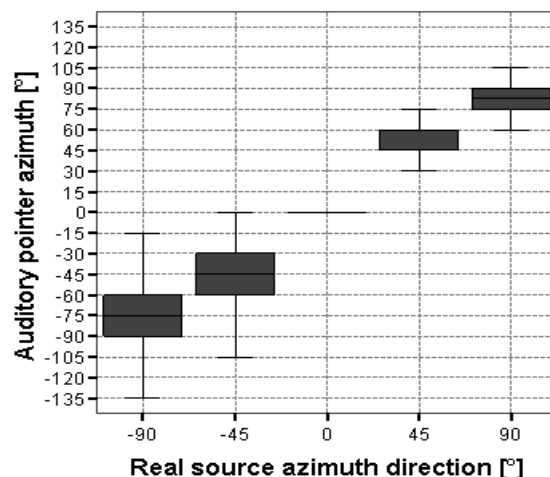


Figure 2: Accuracy of target location identification of 3D audio targets presented via the Nokia N95 3D sound engine (baseline performance with no head-tracking).

We have looked at issues such as size of audio targets, the effects of distracters and the user of pointing 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. We also looked at using the 3D audio to present a clock face so that a daily diary could be presented in a 'RADAR-sweep' around the clock face. An experiment comparing this to a standard linear menu showed that it performed better [18].

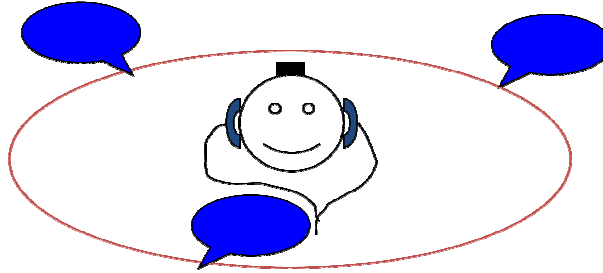


Figure 3: Sounds are presented in a plane around the user's head through headphones. Distance can be manipulated to create a 2.5D display.

Our present work in this area focuses on how to manage the audio display. Users need to be able to minimise sounds and audio sources when they are not needed. In visual displays applications can be minimised to a bar at the bottom of the screen. Our approach for an audio display is to move a sound source to the side and further away. In that way users can still hear it if they want to access it again, but it does not get in the way of their current interaction or activity.

2.3 Tactile feedback for 'eyes free' output

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 expressive '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 [17]. 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 4 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 4: The C2 Tactor from Engineering Acoustics (www.eaiinfo.com).

2.4 Social acceptability

One final important issue is social acceptability, both of the expressive 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. 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 for expressive interaction.

As a first step towards understanding the social acceptability of gestures, a Web survey was conducted to examine how location and audience affect user willingness to perform gestures. Both of these aspects of social situations affect how we behave and make decisions about appropriate actions and therefore provide a useful base from which to begin exploring social acceptability.

This survey examined a set of eighteen gestures which included both device-based gestures and body-based gestures. A device-based gesture refers to any gesture which directly involves touching or moving typical mobile device, in this case a mobile phone. A body-based gesture refers to any gesture that directly involves movements of the body without the use of a typical mobile device. The names and descriptions of these gestures are given in Figure 5. Fifty-one people completed the survey, ranging in age from 22 to 55, 43% of participants were living in the United Kingdom, 45% were living in the United States, and 12% declined to state or were the only participant in a given country. 29% of the participants were female and 71% were male.

Device Based Gestures	
Whip	Whip-like movement of the device in the hand.
Device Shake	Shaking the device in the hand, side to side.
Pocket Tap	Tapping the device with the hand in a pocket.
Device Squeeze	Squeezing the device with the hand.
Table Tap	Tapping the device with the hand setting on a table.
Body Tap	Tapping the shoulder with the device.
Rhythm	Shaking the device in the hand up and down on an even rhythm.

Body Based Gestures	
Shoulder	Rotation of the shoulder, forwards and backwards.
Hand Proximity	Measuring the space between the hands.
Nose Tap	Single tap on the nose with the hand.
Head	Nodding of the head, side to side.
Wrist	Rotation of the wrist, side to side.
Balance	Shifting the balance while sitting
Clap	Clapping of the hands.
Foot Tap	Tapping the foot, sitting or standing.
Belt Tap	Tapping of the belt with the hand.
Finger Proximity	Measuring the space between two fingers on one hand.
Arm Squeeze	Squeezing of the forearm with the opposite hand.

Figure 5: Names and descriptions of all the gestures used in the survey.

For each gesture, survey participants were asked to watch a video of the gesture being performed and answer multiple-choice questions. Because survey participants were asked to imagine the locations and audiences where they might perform these gestures, the videos were designed to focus solely on the gesture itself. The videos used in this survey intentionally portrayed a plain scene without a defined context so that the setting would not distract viewers from evaluating the gesture. The questions in this survey examined the social context of these gestures through the locations where gestures might be used and the kinds of people they might be used in front of. For

each gesture, users were asked to select from a list all of the locations where they would be willing to perform the given gesture as part of a mobile interface. Users were then asked to select from a list all of the types of audiences they would be willing to perform the gesture in front of (see Figure 6).

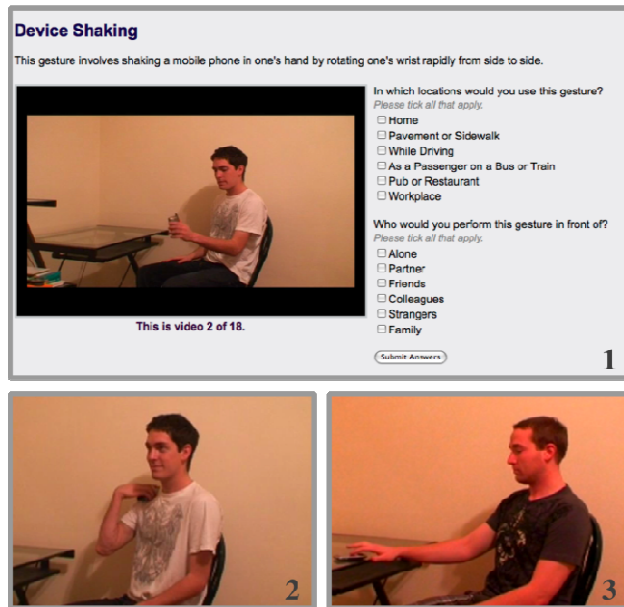


Figure 6: The questionnaire and two screen shots of different gestures (body tapping and table tapping).

By comparing average acceptability rates between locations, we can determine which settings were the best facilitators of gesture usage and which settings were the most controversial. Figure 7 shows the average acceptance rate and standard deviation for each location. The differences show that location plays an important role in how users determine acceptability of a given gesture. For example, the difference between pavement and workplace shows that these places are socially different. As compared to the pavement setting, gestures were 12% more likely to be used in the workplace than on the pavement and 34% more likely to be used at home. The pavement, as a public location, provides few opportunities for privacy and exhibits a restricted set of social norms due to compromises of using a shared space. Home, on the other hand, offers many opportunities for privacy and a clearly defined social norms as a familiar place, with the workplace being somewhere in between the two.

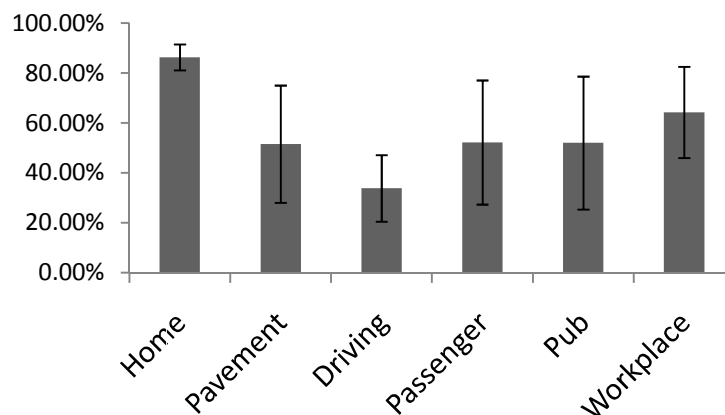


Figure 7: Average percentage of gesture acceptability by location. Error bars show one standard deviation of the acceptability for all gestures at each location.

Familiarity with the audience played a significant role in whether or not a given gesture would be acceptable to use (see Figure 8). Strangers, the least familiar audience with an average acceptability of 51%, were significantly lower than partners, with an average acceptability of 77%. This shows that the more familiar audiences gave performers more confidence in using gestures than less familiar audiences.

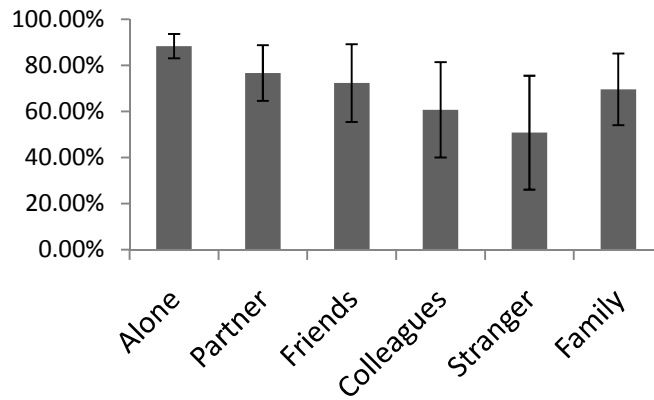


Figure 8: Average percentage of gesture acceptability by audience. Error bars show one standard deviation.

These results show that social acceptability is an important issue and must be considered when designing gesture-based expressive interfaces. If care is not taken then developers could spend a lot of time creating gestures and gesture recognition algorithms for gestures that users will not use as they are too embarrassing to do in public. We are also investigating what other factors may also influence acceptability. Clearly cultural issues will be important as some gestures are rude in one culture but acceptable in another.

3. CONCLUSIONS

This paper has described a range of interaction techniques to facilitate new forms of expressive interactions that do not rely on the hands or eyes. In real world situations hands may be busy holding shopping bags or children, but users still need to operate their mobile devices. In a similar way, users may not be able to look at the screen (or only be able to glance at it) as they need to keep their eyes on where they are going. The techniques we have presented using gestures, audio and tactile feedback have been shown to improve interaction in mobile environments. These are quite different ways of using mobile devices so care must be taken that people will accept them and will not be too embarrassed to do the gestures we create, for example. To avoid this problem we are study the social acceptability of our new interactions to ensure they are both useful and will be used.

Acknowledgements

This work was funded by the EPSRC as part of the GAIME project: Gestural and Audio Interactions for Mobile Environments (EP/F023405/01). Parts of this work were also funded by the Nokia University donations programme.

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*, (Fort Lauderdale, FL, USA, 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*, (Pisa, Italy, 2005), IEEE Press, 167-176.
6. Cohen, M. and Ludwig, L.F. Multidimensional audio window management. *International Journal of Man-Machine Studies*, 34. 319-336.
7. 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.

8. Crossan, A., Williamson, J., Brewster, S.A. and Murray-Smith, R., Wrist Rotation for Interaction in Mobile Contexts. in *Proceedings of MobileHCI 2008*, (Amsterdam, Netherlands, 2008), ACM Press, 435-438.
9. Hoggan, E., Brewster, S.A. and Johnston, J., Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. in *Proceedings of ACM CHI2008*, (Florence, Italy, 2008), 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 *in Proceedings of the fourth international ACM conference on Assistive technologies ASSETS*, (Arlington, Virginia, USA, 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 *In Proceedings of Pervasive*, (Dublin, Ireland, 2006), Springer-Verlag, 101-116.
13. 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*, (Portland, OR, USA, 2006), ACM Press Addison Wesley, 919-928.
14. Rekimoto, J., Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. in *In Proceedings of Fifth International Symposium on Wearable Computers (ISWC '01)*, (2001).
15. Sawhney, N. and Schmandt, C. Nomadic Radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Human-Computer Interaction*, 7 (3). 353-383.
16. Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S. and O' Modhrain, S., GPSTunes - controlling navigation via audio feedback. in *Proceedings of MobileHCI 2005*, (Salzburg, Austria, 2005), ACM, 275-278.
17. 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.
18. Walker, A., Brewster, S.A., McGookin, D. and Ng, A., Diary in the sky: A spatial audio display for a mobile calendar. in *Proceedings of IHM-HCI 2001*, (Lille, France, 2001), Springer, 531-539.
19. Zhai, S., Milgram, P. and Buxton, W., The Influence of Muscle Groups on Performance of Multiple Degree-of-Freedom Input. in *in Proceedings of ACM CHI*, (Vancouver, BC, Canada, 1996), ACM Press, 308-315.