

# A Study on Gestural Interaction with a 3D Audio Display

Georgios Marentakis and Stephen A. Brewster

Glasgow Interactive Systems Group  
Department of Computing Science  
University of Glasgow  
Glasgow, G12 8QQ, UK  
{georgios, stephen}@dcs.gla.ac.uk  
www.audioclouds.org

**Abstract.** The study reported here investigates the design and evaluation of a gesture-controlled, spatially-arranged auditory user interface for a mobile computer. Such an interface may provide a solution to the problem of limited screen space in handheld devices and lead to an effective interface for mobile/eyes-free computing. To better understand how we might design such an interface, our study compared three potential interaction techniques: head nodding, pointing with a finger and pointing on a touch tablet to select an item in exocentric 3D audio space. The effects of sound direction and interaction technique on the browsing and selection process were analyzed. An estimate of the size of the minimum selection area that would allow efficient 3D sound selection is provided for each interaction technique. Browsing using the touch screen was found to be more accurate than the other two techniques, but participants found it significantly harder to use.

## 1 Introduction

Designing a user interface for a handheld device to be used on the move is a challenging task. The lack of screen space for information display in combination with the disturbances incurred by walking makes most of the techniques that are used in desktop user interface design problematic. Anyone who has tried to read a piece of text on a handheld computer while sitting in a taxi or to target a menu item while walking can verify that this task is a difficult one.

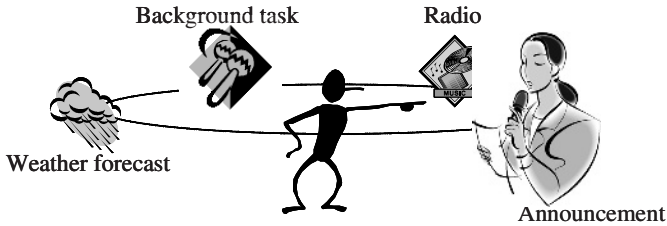
We are taking an alternative approach to interface design for mobile devices by creating multimodal interfaces based on sound and gestures. Multimodal interfaces allow the user to use multiple senses to interact with a mobile computer. It is an objective of our work to use the human senses so that they act in a complementary way to each other. No sense can replace all of the others and each can outperform the rest for certain tasks. For example, listening to text is much more efficient than reading it when walking but on the other hand performing corrections and editing the result can be more efficiently done using the visual sense.

The study reported here examines the potential of designing an interface based on the auditory sense for information display and the use of gestures for control. Moreover, three-dimensional (3D) sound is used as it enables better separation between multiple sound sources and increases the information content of an audio display. It

also allows the spatial nature of the audio space to be used, which we hope will be as beneficial as the spatial display of information in a Graphical User Interface (GUI).

The spatial aspects of our auditory sense have been little explored in human-computer interaction. The ability of the auditory system to separate and apply focus to a sound source in the presence of others (commonly known as the ‘Cocktail Party’ effect [1]) is very helpful for interface design. It implies that simultaneous streams of information can be presented, with users choosing to focus on what is most important (just as occurs visually in GUIs). This phenomenon is greatly enhanced if the sources are spatially separated and thus suggesting the use of three-dimensional sound in auditory user interface design. Other interesting audition properties include omnidirectionality and persistence.

Gestures have the potential to be effective input techniques when used on the move because they do not require visual attention (as do most current mobile input techniques such as pens or soft keyboards). Our kinaesthetic system allows us to know the position of our limbs and body even though we cannot see them. This means that for a mobile application the user would not need to look at his/her hands to provide input, visual attention could remain elsewhere, for example on navigating the environment.



**Fig. 1.** Example of a gesture controlled 3D auditory user interface. A range of different audio sources are presented around a listener and they can be selected using a gesture.

As can be seen in the Figure 1, we are planning to build a 3D audio system where the user will be able to monitor a number of tasks simultaneously, discriminating between foreground and background ones and interacting with them using gestures. The user will hear a range of different sounds but will be able to tune in to the one that is most important, selecting items and interacting with them using gestures. The sound locations in this study are not truly three-dimensional. We place sounds on a plane around the user’s head at the height of the ears to avoid problems related to elevation perception. This results in a 2.5D planar soundscape.

## 2 Previous Work on Auditory and Gestural Interfaces for Mobile Devices

Applications of audio in user interface design have been examined by many researchers. Gaver [8] introduced the notion of Auditory Icons in user interface design. Auditory Icons are based on the notion of everyday listening and they have been used in systems such as the SonicFinder and the ARKola system [15]. Blattner *et al.* [2] have proposed designing audio displays that are based on structured musical listening, re-

sulting in the notion of Earcons that have been examined and proved to be usable by Brewster [5].

The notion of an audio window system was introduced by Cohen and Ludwig [7]. Cohen also introduced the concept of using 3D sound to increase the auditory display space and proposed simple gestural interaction with 3D sound for input [6]. According to Cohen, sounds are positioned in the space around the user and a mapping between the sounds and the elements of the interface is performed. Users can subsequently interact with the sounds by pointing, pitching, catching and throwing them. By using these interaction techniques users can organize the system so that it suits their needs. Another idea developed by Cohen has been the an audio pointer as an aid in the cluttered audio space to assist localization and help the user disambiguate current position in relation to the position of the sounds in the display. The concept of 'filtars' has been also introduced by Cohen. According to this idea sounds slightly change as a result of filtering when being in different states such as selected, caught etc. This cue has been designed to assist the user in understanding the state of the display elements as he/she is interacting with them.

Another attempt to construct a system based around spatialised audio was Nomadic Radio by Sawhney and Schmandt [14]. It is targeted primarily at messaging. It is enabled with speech recognition and synthesis to allow the user to communicate and receive feedback from the system. It is also enabled with 3D audio to enhance simultaneous listening and conferencing. Another interesting issue about this application is the fact that it works based on loudspeakers mounted on the shoulders of the user and a directional microphone on the chest of the user; the user is able to listen to his/her real audio environment at the same time as when interacting with the system. The system also uses a space to time metaphor to position different messages around the user depending on the time of arrival. It works using a limited set of commands that can be recognized through the speech recognizer.

Brewster *et al.* [4] tested a three dimensional gesture controlled audio display on the move. They used an auditory pie menu centred on the head of the user and compared fixed to the world versus fixed to user sound presentation. They found that fixed to user sound presentation performs better in terms of time required to perform tasks as well as in terms of the walking speed the users could maintain. In another study by Pirhonen *et al.* [12] gestural control of a MP3 audio player was found to be faster and less demanding than the usual stylus based interaction when on the move.

Goose *et al.* [9] presented a system using 3D audio and earcons and text to speech for browsing the WWW. Finally, Savidis *et al.* [13] used a non-visual 3D audio environment to allow blind users to interact with standard GUIs. Different menu items were mapped to different locations around the user's head.

The ideas in the literature shape a framework for working with sound in a gesture controlled 3D audio display. Speech control has been used to control a 3D audio display, however it is known to require a silent environment to operate, users to be able to remember the command repertoire and can be indiscrete. Gesture control seems like a more feasible solution for systems to be used on the move and in a social context. Cohen as well as other researchers, have proposed designs for 3D audio interface development. However, with the exemption of [4], no formal evaluation of these ideas was done. We believe that given the ambiguity that can occur in such interfaces, further empirical research is necessary to allow us to design 3D audio interfaces in a formal way.

### 3 Three-Dimensional Audio Issues and Definitions

Designing a user interface based on 3D sound and controlling it by gestures poses a number of questions that must be answered before successful interfaces can be created. It is the case that when asking people to locate a sound, there is a certain extent of ambiguity in their answers. This ambiguity, called Localization Blur [3], has been measured for users listening to sounds from different locations in space and has been shown to be bounded (for a full review see [3]). As found in Blauert [3] localization blur can range from  $\pm 3.6^\circ$  in the frontal direction,  $\pm 10^\circ$  on the left/right directions and  $\pm 5.5^\circ$  to the back of a listener under well controlled conditions. Localization blur also depends on the position of the sound source and the spectral content of the source. Virtual sound positioning using headphones is realized using HRTF filtering [3]. Head Related Transfer Functions (HRTF) are functions that capture the frequency response of the path between a sound source and the listener's tympanic membrane. These functions are estimated experimentally usually using a dummy head and torso. By filtering a sound signal using these functions it is possible to apply to it directional characteristics. However, problems related to non-individualized HRTF's (using a set of filters not created from your own ears) and HRTF interpolation and reproduction reliability affects the quality of the result so that performance is commonly poorer than for real-world listening.

In the light of these facts, it is interesting to try to define what we mean by asking a person to interact with a spatially positioned sound source, utilizing cues such as the source's direction. It is necessary to associate a certain area of the display to each of its elements. This mapping is not obvious as it is in graphical displays, since a person cannot judge exactly where the sound source is located or what its dimensions are. For example, consider the setup where non-overlapping sounds are presented around the user in the horizontal plane. In this case, we could map an angle interval to each display element. Any type of interaction that occurs in this area could be mapped to a specific display element positioned in the centre of this angle interval. By estimating this interval a design principle is obtained that can be used to partition the audio space. The estimation of such quantities can be problematic though, due to the unfamiliarity of many users with the sound localization task as well as with virtual 3D sound environments. Both when using real sound sources and when using virtual ones untrained subjects respond with great variation to questions related to the direction of a sound source.

Localization accuracy can also be improved by using feedback. Feedback could help in assisting the whole localization procedure by guiding the user towards the source and by reassuring the user that he/she is on the target area, thus making the selection process more effective. It could help overcome the poorer localization that occurs with virtual 3D sound to allow it to be used effectively in a user interface.

Two design techniques are positioning the sound sources egocentric versus exocentric. Egocentric or fixed to the listener sources, can be localized faster but less accurately, due to the absence of active listening. By active listening we refer to the process of disambiguating sound direction by small head movements. Active listening enhances localization accuracy but results in computationally intensive updating of the sound source positions (which may be a problem in a lower-powered mobile device) as well as in increasing the time required for a person to localize a sound stimulus. This is because the process of active listening involves moving and converging

towards the target sound using the information provided by the updated sound scene. There is a trade-off in localization accuracy and time required to make a selection when deciding between fixed to the listener versus fixed to the world sound sources. We chose the better accuracy of exocentric or fixed to the world over egocentric to overcome the limitations of the non-individualized HRTF's we used.

A key issue in 3D audio design is the number of sources that can be presented simultaneously. It has been shown that human performance degrades as the number of audio display elements increases [11] when sounds stem from the same point in the display. Spatial separation, however, forms a basic dimension in auditory stream segregation and thus can possibly increase the number of sources users can deal with. The study we present here uses just one sound source as we wanted to gain an idea of selection angles in the simplest case, before we move on to more sophisticated sound designs later in our research.

To handle the ambiguity in the aforementioned tasks, we decided to use adaptive psychophysical methods. Adaptive methods are characterized by the fact that a stimulus is adjusted depending on the course of an experiment. They result in measures of performance on psychophysical tasks as a function of stimulus strength or other characteristics. The result constitutes what is called a psychometric function [10]. The psychometric function provides fundamental data for psychophysics, with abscissa being the stimulus magnitude and the ordinate measuring the subjective response. One commonly used psychophysical method is the Up-Down method. Up - Down procedures work by setting the stimulus to a certain level at the beginning of an experiment and then decreasing or increasing the stimulus based on the observation of a specific pattern in the subject's response. The phenomenon that occurs when the direction of stimulus change is reversed is called a reversal. Up-Down methods that decrease the stimulus after a valid answer and increase stimulus after an invalid answer converge to the 50% point of the associated psychometric function. A point of this function that corresponds to 50% would imply that at this stimulus level, 50% of the answers would be expected to be 'valid'. By altering the rule of stimulus change, different points of the psychometric function can be estimated. However, full sampling of the function is often impossible due to the large number of experimental trials required.

## 4 Experiment

An experiment was designed to answer some fundamental questions about the design of audio and gestural interfaces, in particular: what is the minimum display area needed for the effective selection of a sound source, and what selection technique is the most accurate. We estimated the angle interval that would result in 67% of a user's selections being on target. To do this we used an adaptive psychophysical method, more specifically a two-down one-up method (for a review of adaptive psychophysical methods see [10]). We investigated three different browsing and selection gestures that could be used by users to find items in a soundscape and select them. We used head/hand tracking to update the soundscape in real time to improve localization accuracy.

The three browsing gestures were: browsing with the head, browsing with the hand or browsing using a touch tablet. These gestures differ with respect to how common

they are in everyday life. The first is the normal way humans perform active listening, with the position of the sound being updated as the user's head moves, so should be very easy to perform. The second is more like holding a microphone and moving it around a space to listen for sounds. The location of the sounds in the display is updated based on the direction of the right index finger. Direction is inferred by a 2D vector defined by the position of the head and the position of the index finger of the user. The third gesture can be thought as an extreme, in the sense that it cannot be mapped to a real world case. The user moves a stylus around the circumference of a circle on a tablet (the centre of the tablet marks the centre of the audio space) and the position of the sound source is determined by the stylus direction with respect to the centre of the tablet. In early pilot testing this type of sound positioning proved to be confusing if a user was to start a selection from the lower hemisphere. This was due to the fact that sounds moved as if the participant was looking backwards, although the participant was actually looking forwards. For this reason, we decided to reverse left and right in case the user began browsing in the lower hemisphere. By doing this, the optimal path to the next sound could be found by always moving on the circle towards the direction in which the sound cue was perceived to be stronger.

The selection gestures were: nodding with the head, moving the index finger as if clicking a non-existent mouse button, and clicking a button available on the side of the stylus to indicate selection. In this experiment, three combinations of the above were examined: browsing with the head and selecting by nodding, browsing with the hand and selecting by gesturing with the index finger, and browsing with the pen on the tablet and selecting by clicking.

#### 4.1 Sound Design and Apparatus

The aim of the experiment was to look at how the minimum angle interval that allows efficient selection of an audio source varies with respect to direction of sound event and interaction technique used. We used a single target sound placed in one of eight locations around the users head (every  $45^\circ$  starting from  $0^\circ$  in front of the user's nose) at a distance of two meters. This stimulus was a 0.9 second broadband electronic synthesizer sound, repeated every 1.2 seconds.

We used very simple audio feedback to indicate that the user was within the target region and could select the sound source. This was a short percussive sound that was played repeatedly while the user was 'on target' (i.e. within the current selection region) to assist each user in localizing the sound. This was played from the direction of the target sound. Sounds were played via headphones and spatially positioned in real time using the HRTF filtering implementation from Microsoft's DirectX 9 API. Sound positions were updated every 50msec.

To perform gesture recognition and finger tracking we used a Polhemus Fastrack to get position and orientation data, and two sensors (see Figure 2). One sensor was mounted on top of the headphones to determine head orientation and allow us to recognize the nod gestures. A second sensor was mounted on top of the index finger to determine the orientation of the hand relative to the head and to recognize the clicking gesture in the hand condition. A Wacom tablet was used for the tablet condition. We determined nodding and clicking by calculating velocity from the position data.



**Fig. 2.** A participant making a selection in the hand pointing condition.

## 4.2 Experimental Design and Procedure

The experiment used a two-factor within-subjects design with each participant using each of the three interaction techniques in a counterbalanced order. There were two independent variables: sound location (eight different levels) and interaction technique (three levels). The dependent variables were deviation angle from target and effective selection angle. Participants were also asked to rate the three interaction used for browsing and selecting on a scale from one to ten with respect to how comfortable and how easy to use they found them. Our hypotheses were that the effective selection angle would be affected by interaction technique, with no effect of location because participants always faced the targets when selecting them.

Twelve participants took part: five females and seven males with ages ranging from 19 to 30.

The participant's task was to browse the soundscape until the sound was in front and then select the target sound using the interaction techniques described. The target sound repeated until the participant performed a selection. Upon selection, the stimulus was presented in a different location randomly out of the set of available positions. The whole process was repeated until all up-down methods for each position converged. According to the up-down rule the effective selection angle was varied between trials; it was reduced after two on-target selections and increased after one off target selection. The step was initially  $2^\circ$  but was halved to  $1^\circ$  after the third reversal occurred. It should be noted that participants were unaware of this process; they were instructed to perform selections based only on audio feedback and localization cues.

The experiment lasted approximately one hour. Participants stood wearing the headphones and tracker. They could turn around and move/point as they wished and were given a rest after each condition. The experiment could not be conducted in a fully mobile way with users walking (as in previous studies such as [4]) due to the tracking technology needed for gesture recognition – participants had to stay within range of the Polhemus receiver. The results may therefore be different if the techniques were used in a fully mobile setting, but they will indicate if any of them are usable and should be taken further. Participants were trained for a short period before being tested in each condition to ensure they were familiar with the interaction techniques. They performed eight selections before embarking on the experiment. Prior to

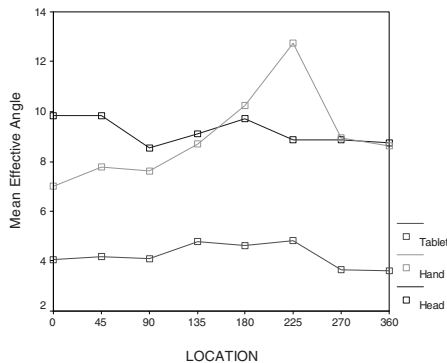


testing, participants' localisation skills were checked to rule out hearing problems and to familiarise them with the sound signal they would hear. During this 3D sound training, participants were asked to indicate verbally the direction they had perceived the sound source was coming from. The experimenter subsequently corrected them in case they were wrong and tried to direct their attention to the relevant cues.

## 5 Results

The up-down method was expected to converge on the point of the associated psychometric function where 67% of the selections would be on target. To estimate this point we averaged the angle intervals as these were updated by the up-down rule. Averaging included only the angle intervals that occurred after the second reversal.

A 3x8 two factor ANOVA was performed to examine whether sound location and interaction technique affected the effective selection angle. Sound location was not found to have a significant main effect ( $F(2.314, 77) = 2.241, p = 0.121$ ). However, there was a significant main effect for interaction technique ( $F(2, 22) = 10.777, p = 0.001$ ). There was no interaction between location and technique. Pair-wise comparisons using Bonferroni confidence interval adjustments showed that the tablet condition was significantly more accurate than the other two techniques, but no significant differences were found between the hand and head. Figure 3 shows the mean effective angle intervals for the three interaction techniques with respect to direction of the sound. These results define the one side interval around a source. To give an example of how these data could be applied, if an exocentric 3D audio user interface (enabled with active listening) using audio feedback and controlled by a stylus on a touch tablet, was developed, the designer should allow at least  $4^\circ$  on *each side* of a sound positioned at  $90^\circ$  relative to the front of the user so that a user would be able to select the sound effectively.



**Fig. 3.** Effective selection angle for each sound direction.

The deviations of the users' selections from target were also analyzed. Ninety measurements for all different directions were analyzed. A 3x8 two factor ANOVA showed a significant main effect for interaction technique ( $F(2, 192) = 7.463, p = 0.001$ ). Direction also had a significant main effect ( $F(7, 672) = 7.987, p = 0.001$ ).



There was a significant interaction between technique and direction ( $F(14,1344) = 7.996$ ,  $p = 0.001$ ). Pair-wise comparisons using Bonferroni confidence interval adjustments showed that the tablet condition was significantly better than the others, but there was no significant difference between head and hand. With respect to the direction of the sound event, direction  $225^\circ$  was significantly different from directions  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $270^\circ$ ,  $315^\circ$  and direction  $180^\circ$  was different from  $45^\circ$ ,  $270^\circ$ ,  $315^\circ$ . Figure 4 illustrates mean deviation from target and its standard deviation.

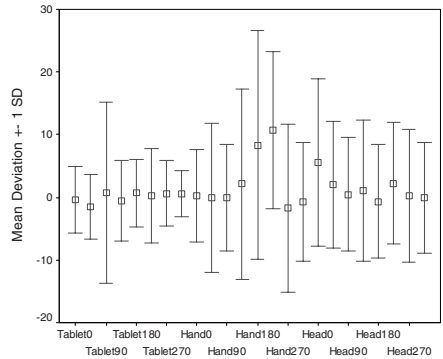


Fig. 4. Mean deviation from target versus sound direction.

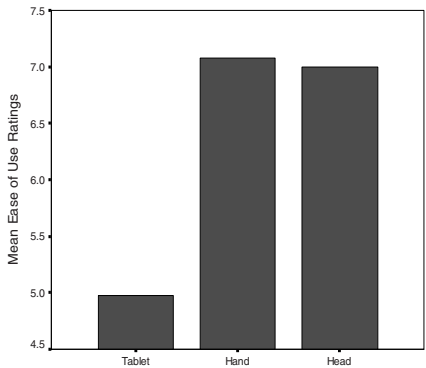


Fig. 5. Mean ease of use ratings for each interaction technique.

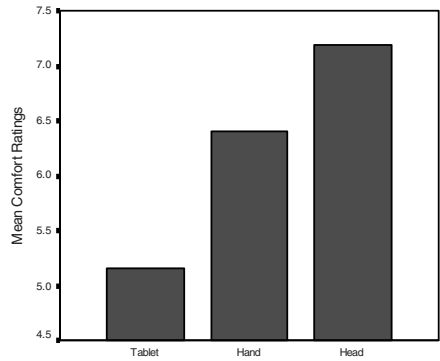


Fig. 6. Mean Comfort ratings for each interaction technique.

As mentioned, each participant was asked to rate each of the interaction methods in terms of how easy and how comfortable he/she found them to be, on a scale from 1 to 10. Figure 5 shows the means of the results for ease of use. A statistical analysis of variance showed interaction method to be a significant factor ( $F(36) = 7.386$ ,  $p = 0.002$ ). Bonferroni t-tests verified mouse to be significantly harder to use, but showed no statistical difference between hand and head.

A similar analysis on how comfortable the use of the three devices was showed no significant difference between devices. Figure 6 shows comfort means for the three interaction methods. It should be noted that participants have performed a large num-

ber of selection using the three interaction methods to allow the up-down methods to converge in the three different conditions and the eight different sound positions. In that sense, when observing the graphs, absolute values should be taken into account carefully. However, the ratings of the three devices relative to each other can be used to infer how they are ordered relative to each other with respect to ease of use and comfort.

## 6 Discussion

The results of the study showed that interaction with a 3D audio source can be done effectively in the presence of localization feedback. They also showed that novel methods of browsing can be as effective (and even more effective) in locating sounds than 'natural' ones in the presence of feedback. Users were able to perform active listening using the tablet and the hand without any particular difficulty. It was also surprising that they could do the active listening operation more accurately when using the tablet than when using their heads. This can be explained in terms of the resolution that the three different mechanisms provide. For example, a stylus controlled touch tablet provides a much better minimum possible displacement compared to the head or the hand of a person. By constructing histograms of deviation data we verified that the results of the up-down procedure would indeed allow 67% percent of the selections to be on target. It should be mentioned, however, that more reversals would result in having more accurate results. This was not possible to do since we tried to maintain a within-subjects design and keep the experiment duration in the order of one hour to avoid effects caused by fatigue. Effective selection angles are likely to reduce with practice and improved feedback design.

When considering the three interaction methods, one would not expect the direction of sound to be a significant factor in the results of this study. This is due to the active listening operation; that is, users selected a sound when it was in front of them. This was verified in the effective angle case where no location was found to be a significant factor. However, in the deviation analysis, certain angles were significantly different from others. This was mostly in the direction of 225° degrees. The reason for this difference can be described by the mechanics of the browsing and selection modalities. A closer look at the graphs reveals the technique that caused this difference was browsing by hand. As was observed during testing, some right-handed participants found it difficult to point to that location, if they had not turned their bodies first (they had to reach around their body causing them to stretch, reducing the accuracy of their selections). A significant number of participants indeed tried to point without turning their bodies, a result that influenced the accuracy of the browsing and selection processes.

By analyzing how the ease of use ratings are ordered, we see that users find browsing the sound space to be equally easy either using the head or using the hand. The touch tablet however, although more accurate, was not rated highly. This can be associated with the unnaturalness of the browsing process. In the other two cases, participants used a natural process for browsing the space, such as moving their heads or simulated one by moving their hand in a synchronous way with their head.

When considering the effective angles, we can observe that if accuracy was the only factor to be taken into account, an audio user interface could be constructed hav-

ing all eight sounds locations, and possibly more. Our next study will investigate the presentation of multiple sounds and the design of a more sophisticated soundscape such as would be needed for a real application of a wearable device based around 3D sound and gestures. If studies show that listeners cannot use sounds from eight locations then we can increase the selection angles for our sound sources which will further increase selection accuracy.

## 7 Conclusions

In this paper a study on gestural interaction with a sound source in the presence of feedback was presented. Three different gestures for browsing and selecting in a 3D soundscape were examined and their effectiveness in terms of accuracy was assessed. Browsing and selecting using a touch tablet proved to be more accurate than using a hand or a head gesture. However, browsing and selecting using the hand or the head were found to be easier and more comfortable by the users. Effective selection angles that would allow efficient selection were estimated for each interaction technique and on 8 sound locations around the user using an adaptive psychophysical method. The results show that these different interaction techniques were effective and could be used in a future mobile device to provide a flexible, eyes free way to interact with a system.

**Acknowledgements.** This study was supported by the EPSRC funded Audioclouds project grant number GR/R98105.

## References

1. Arons, B., *A Review of the Cocktail Party Effect*. Journal of the American Voice I/O Society, 1992. **12**: p. 35-50.
2. Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M., *Earcons and Icons: Their Structure and Common Design Principles*. Human-Computer Interaction, 1989. **4**(1): p. 11-44.
3. Blauert, J., *Spatial Hearing: The psychophysics of human sound localization*. 1999: The MIT Press.
4. Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. *Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices*. In *ACM CHI*, 2003. Fort Lauderdale, FL: ACM Press, Addison-Wesley. p. 463-480
5. Brewster, S. A., *The design of sonically-enhanced widgets*. Interacting with Computers, 1998. **11**(2): p. 211-235.
6. Cohen, M., *Throwing, pitching and catching sound: audio windowing models and modes*. International Journal of Man - Machine Studies (1993), 1993. **39**: p. 269 - 304.
7. Cohen, M. and Ludwig, L., *Multidimensional Audio Window Management*. International Journal of Man - Machine Studies, 1991. **34**: p. 319-336.
8. Gaver, W. W., *The SonicFinder: An Interface that uses Auditory Icons*. Human-Computer Interaction, 1989. **4**: p. 67-94.
9. Goose, S. and Moller, C. *A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure*. In *7th ACM international conference on Multimedia*, 1999. Orlando, Florida, United States: ACM Press. p. 363 - 371

10. Leek, M. R., *Adaptive procedures in psychophysical research*. Journal of Perception & Psychophysics, 2001. **63**(8): p. 1279-1292.
11. McGookin, D. K. and Brewster, S. A. *An Investigation into the Identification of Concurrently Presented Earcons*. in *ICAD 2003*. (Boston, MA). p. 42-46
12. Pirhonen, A., Brewster, S., and Holguin, C. *Gestural and audio metaphors as a means of control for mobile devices*. In *ACM CHI*, 2002. Minneapolis, Minnesota, USA: ACM Press New York, NY, USA. p. 291-298
13. Savidis, A., Stephanidis, C., Korte, A., Rispin, K., and Fellbaum, C. *A generic direct-manipulation 3D auditory environment for hierarchical navigation in non-visual interaction*. In *ACM ASSETS '96*, 1996. Vancouver, Canada, 1996: ACM Press. p. 117-123
14. Sawhney, N. and Schmandt, C., *Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments*. ACM Transactions on Computer-Human Interaction, 2000. **7**(3): p. 353-383.
15. William W. Gaver, Randall B. Smith, and O'Shea, T. *Effective Sounds in Complex Systems: The ARKOLA Simulation*. in *Conference on Human Factors in Computing Systems archive*. In *ACMCHI*. 1991. New Orleans, Louisiana, United States: ACM Press New York, NY, USA. p. 85 - 90