Effects of Reproduction Equipment on Interaction with a Spatial Audio Interface

Georgios N. Marentakis and Stephen A. Brewster
Glasgow Interactive Systems Group
Department of Computing Science
University of Glasgow, G12 8QQ, UK
e-mail: {georgios, stephen}@dcs.gla.ac.uk Web: www.audioclouds.org

ABSTRACT
Spatial audio displays have been criticized because the use of headphones may isolate users from their real world audio environment. In this paper we study the effects of three types of audio reproduction equipment (standard headphones, bone-conductance headphones and monaural presentation using a single earphone) on time and accuracy during interaction with a deictic spatial audio display. Participants selected a target sound emitting from one of four different locations in the presence of distracters whilst wearing the different types of headphones. Target locations were marked with audio feedback. No significant differences were found for time and accuracy ratings between bone conductance and standard headphones. Monaural reproduction significantly slowed interaction. The results show that alternative reproduction equipment can be used to overcome user isolation from the natural audio environment.

Author Keywords
Auditory I/O and Sound in the UI, Interaction Design.

ACM Classification Keywords
H.5.2 [User Interfaces]: Auditory (non-speech) feedback.

INTRODUCTION
Spatial auditory interfaces have much potential for facilitating interaction where standard displays are difficult to use. Human hearing [7] has useful properties such as the ability to localize a sound source, omni-directionality, omni-presence and the ability to process multiple streams of information simultaneously (known as the *Cocktail Party* effect [1]). In addition, in application areas where vision should be directed to a primary task, eyes free interaction can be more effective [3].

Most spatial auditory interfaces build on the audio window concept [4]. Such systems use a mapping between spatially positioned sounds and display elements to define interaction. Each display element appears in a certain position in space, enabling space-based interaction techniques like pointing. Spatial positioning of display content also helps in differentiating sounds and contributes to comprehension of concurrent audio streams [1, 2, 7]. In the literature there are examples of hierarchical content organization such as in Brewster *et al.* [3], where users nod to select spatially positioned sounds in an auditory pie menu placed around their heads. Other applications include textual content presentation by means of synthesized speech [6, 9]. Such applications commonly use spatial audio to sonify speaker position and structural information within a document.

Head Related Transfer Functions (HRTFs) are often used to position sounds in space. HRTFs encode the properties of the path from a sound source position into each of the listener's eardrums. When applied to a sound signal they produce the impression of space. In most of the studies in the literature sounds are presented to the user using headphones because loudspeaker presentation would confine the user to a fixed position and thus hinder mobility.

Headphone presentation can, however, be a disadvantage in some application areas for spatial audio displays. Our auditory sense is valuable when mobile both for communicating and as an alerting mechanism. Blocking it can be irritating and possibly dangerous, depending on the interaction context. For example, being able to hear cars when crossing the road is important to avoid accidents. One way to overcome this problem is by using alternative reproduction devices. Nomadic Radio [8], a spatial audio interface targeted primarily at messaging, was designed to work on shoulder mounted speakers to overcome this problem. This approach, however, can be indiscrete due to other people overhearing sounds emitting from the loudspeakers. In noisy environments intelligibility is also likely to be reduced by the interference of other sounds. Goose and Safia [6] used speakers for spatial audio presentation since their proposed system was designed for inside a car. This option is a context specific solution and cannot be applied generally in more mobile situations.

Although reproduction device is an important factor from a user satisfaction, usability and safety point of view, so far no study has reported on the effects of different types of reproduction equipment for interaction with spatial audio displays.

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.
ACM 1-59593-002-7/05/0004.

In this paper we present a study evaluating two alternative reproduction devices and compare them with standard headphone presentation. In particular, we evaluate monaural presentation and bone conductance headphones versus presentation using normal headphones. Both of these options are interesting since they provide the possibility of unblocking the audio channel at the same time as the user participates in a digital audio experience.

Monaural presentation is achieved by playing the sound using only a single earphone (such as most mobile phone speakers or hands-free kits). This technique has the advantage that it allows for one of the two ears to monitor the real audio environment. However, sound localization is based on binaural cues, i.e. differences between the signals arriving at both ears, so the spatial impression is degraded in monaural situations. The impression of space in monaural presentation does occur (mainly due to the effect of the pinnae) but localization judgments are far from accurate [2]. Therefore, we need to investigate if the localization cues are strong enough to make a successful spatial audio interface.

Reproduction using bone conductance headphones is accomplished by transmitting vibrations through the skull of the user. Such headphones feature a vibrating surface that is mounted on the side of the head in front of each ear. The mounting mechanism is very similar to standard headphones, with the difference that the outer ear is completely open. Vibrations propagate through the skull to stimulate the ear and thus become audible. The perceived sound signal will, however, be distorted by the transmition path, increasing the signal to noise ratio. Reproduction fidelity is thus lower than normal headphones. Nevertheless, this reproduction technique can lead to intelligible impressions of speech or music. This may not be the case for spatial audio, due to the fact that the subtle pinnae effects applied through HRTF filtering will be distorted. Some cues will remain, in particular inter-aural intensity and time differences. These cues can produce a spatial impression similar to stereo reproduction which may be enough to form an overview of the spatial structure of a simple audio display, for example one that is based on the horizontal axis in front of the user.

Given the disadvantages of these reproduction devices over standard headphones, an experimental evaluation is necessary to see how interaction will be affected. Studying these alternative reproduction devices is useful since it can provide insight on how to combine the real with a digital audio environment, as well as how feedback can be used to facilitate interaction in the presence of weak localization cues.

EXPERIMENT

The aim of the experiment was to assess the effects of the three different reproduction devices on target selection performance in a spatial audio interface. Figure 1 shows the different devices used. These were: standard Sennheiser HD250 closed-back headphones, Vonia EZ-3200P bone conductance headphones and a Panasonic RP-HS50 earpiece (reproduction using standard headphones will also be referred to as binaural listening).

Figure 1. The different headphone types used in the experiment. Bottom left is the single earpiece, top left are the bone conductance headphones and to the right are the Sennheiser headphones.

Experimental Task and Design

The experimental task was designed to represent a common scenario in interaction with a deictic spatial audio display where users must select a sound emitting from somewhere the space in front of them using a tracker held in their hand [4]. Participants initially had to listen for a target sound, played in isolation from a certain position in space. When finished, the target sound played continuously together with three distracter sounds. An angle span was associated with each sound. To select a sound participants had to point at its location and make a downwards wrist gesture to indicate selection. Participants received audio feedback (the sound of people cheering) when they were within the target sound's area. An XSENS MT-9B orientation tracker (www.xsens.com) was used to track user orientation and the selection gesture.

To avoid any effects related to timbre, the same sound was used for both distracters and the target sound. This was a short (0.5 sec.) segment of white noise. During each trial this sound played from four different locations in the display, one of which was the target position for the trial. Sound positions were at 45 degrees intervals, starting from -67° to +67° in front of the listener (with 0° in front of the user's nose), in the horizontal plane. The target sound was located in the leftmost position in every second task to equalize the distance pairs. Target sound location for every other task was selected randomly from the three remaining ones. This target selection procedure resulted in three distance pairs of 45°, 90° and 135° arc length respectively.

To improve intelligibility, we introduced a 300ms onset difference between neighboring sounds. Counting from left to right, this resulted in the second sound starting 300ms later than the first sound, the third 600 ms later and the

	Equipment	Distance to Target
Group A	HD	45°,90°,135°
	MA	45°,90°,135°
Group B	HD	45°,90°,135°
	BC	45°,90°,135°

Table 1. Experimental factors and their levels (HD = Headphones, BC = Bone Conductance, MA = Monaural).

fourth 900 ms later than the first sound. Sounds repeated after a 500 ms period of silence.

Participants performed the task according to the design presented in Table 1. There was a short training session prior to testing, during which their performance was monitored to make sure that they understood the task. After participants successfully completed four consecutive trials during the training session, the testing started. Participants were tested in the two experimental conditions associated with the groups shown in Table 1, one followed by the other in a counterbalanced order.

Sixteen participants were tested (17 to 27 years of age, 2 females and 14 males). Participants were paid £5 for their participation. Time to complete trials, angular deviation from target and movement pattern for each trial were recorded during the experiment. After testing in each experimental condition participants completed a NASA TLX subjective workload assessment. The experiment lasted half an hour. In total 128 measurements were available per level combination for the monaural and bone conductance cases and 256 for the standard headphone case.

Results

The analysis involved two within-subjects comparisons, one for each participant group and one between-subjects comparison to compare bone conductance and monaural presentation.

Time Analysis

The taken time to complete trials was analyzed using a repeated measures ANOVA. Monaural presentation proved to slow interaction significantly both when compared with binaural ($F(1,127) = 120.498$, $p < 0.001$) and with bone-conductance presentation ($F(1,254) = 118.941$, $p<0.001$). Time was not significantly different between the binaural and bone-conductance cases. Figure 2 shows the timing results.

As would be expected, time to complete tasks was significantly affected by distance to target in all cases ($F(2,254) = 58.594$, $p<0.001$ between binaural and bone-conductance cases, $F(2,254) = 10.856$, $p<0.001$ between binaural and monaural cases and $F(2,508) = 16.699$, $p<0.001$ between bone-conductance and monaural cases).

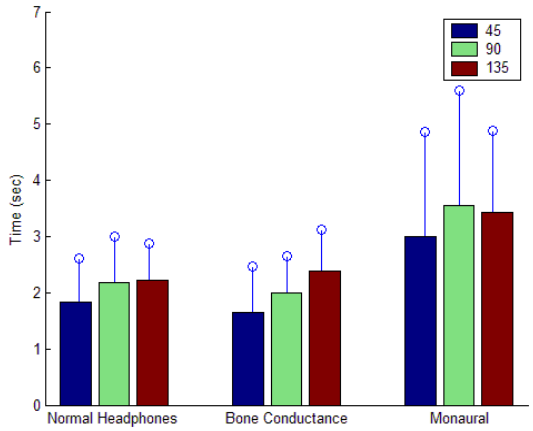

Figure 2. Mean time and standard deviation for the three presentation methods and the three distance paths.

Accuracy & Workload Analysis

Based on the orientation measurements for each selection, percentage correct ratings for each condition were calculated and can be found in Table 2. As can be observed the wide angle span of each target resulted in high success rates for all reproduction types. Reproduction device was not found to affect accuracy.

Data from NASA TLX forms measuring subjective workload were also analyzed. No significant difference in overall workload was found in any of the comparisons of reproduction equipment. Therefore, subjective workload was not affected by reproduction type.

Discussion

The results of this study show that alternative reproduction techniques can replace standard headphones in spatial audio systems. In the context of our experimental task, bone conductance presentation was found to be as fast and in the same range of accuracy as binaural presentation. Although we cannot argue that bone conductance headphones can produce a comparable spatial impression to standard headphones, it is the case that the more 'stereo like' cue was sufficient to guide the users to the target sound. It should also be stressed that the easily perceptible audio feedback cue contributed significantly to the success of users both in the bone-conductance and the monaural cases. Rapidly presented and perceived feedback is very important with low-fidelity spatial audio reproduction techniques, because it can compensate for the weaker localization cues. Consequently, a 'stereo' like directional cue combined with good feedback can successfully guide users to a spatially positioned target sound.

Equipment/Distance	45°	90°	135°
Headphones	99.6%	94.14%	93.75%
Bone Conductance	100%	93.7%	94.53%
Monaural	97.65%	98.43%	93.75%

Table 2. Percentage of on-target trials across the different conditions of the experiment.

As expected, the results showed that monaural presentation slowed interaction. This is due to the fact that binaural differences proven to be important in making judgments of sound direction [2, 7] are not available under monaural conditions. This results in extended search times for the target display elements, a fact that slows interaction. In fact, as can be observed from the results, monaural presentation slowed interaction by one and half to two times in the context of our task. However, the time to complete tasks was not completely unrealistic from a usability point of view. This suggests that in a display where the user is familiar with the positions of the elements, interaction speed under monaural reproduction is likely to be within an acceptable range. This reproduction technique could also be used for presenting speech, playing music and other content presentation tasks.

In addition, our auditory system is still sensitive to loudness, pitch and other sound attribute differences under monaural conditions. These could also be used to guide the user to a target sound, compensating for the lost directional cues. However, extra care must be taken if presenting simultaneous audio streams under monaural conditions because the phenomenon of masking is much stronger in monaural cases than in binaural ones [1, 2, 7]. Masking is defined as the process by which the threshold of audibility for one sound is raised by the presence of another. Binaural presentation benefits from lower masking thresholds and has been proven to contribute to the Cocktail Party effect [1, 2, 7]. This leads to the conclusion that under monaural conditions the amount of content that can be rendered simultaneously in the display will definitely be lower than in the binaural case.

As expected, we found that trial completion times depended on distance to target. This can be explained in terms of Fitt's law, which relates distance to target to selection time. From our results it seems that users used two different strategies for selection. In the standard headphone case users tended to perform in a manner that varied logarithmically with distance (standard Fitt's law). This implies that users were moving fast and often overshot the target prior to homing to it. In the bone conductance case a linear dependency to distance was found, implying a more cautious selection strategy. These two modes of interaction are similar to those suggested by Friedlander *et al.* [5] for interaction in a similar display. However, a more thorough analysis is beyond the scope of this paper.

CONCLUSIONS

This study has shown that interaction with spatial audio displays is feasible using bone-conductance and monaural presentation in the presence of good feedback. The proposed solutions are promising for overcoming the problem of user isolation from the real audio world. The results showed that, with appropriate design, interaction with a spatial audio display using bone conductance headphones can be as fast and accurate as interaction using standard headphones. Although monaural presentation was found to slow interaction, the selection times in our study were within an acceptable range. Based on the results of this paper further research can be stimulated, aiming at designing spatial audio displays that benefit from user exposure to our natural audio environment.

ACKNOWLEDGEMENTS

The authors would like to thank Lorna Brown and Iain Darroch for reviewing the paper and providing helpful comments. This study was supported by the EPSRC-funded Audioclouds project (www.audioclouds.org), 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. Blauert, J., *Spatial Hearing: The psychophysics of human sound localization*. 1999: The MIT Press.
3. Brewster , S., Lumsden, J., Bell, M., Hall, M. and Tasker, S. *Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices*, in *ACM CHI*, 2003, p. 473-480.
4. Cohen, M. and Ludwig , L., *Multidimensional Audio Window Management.* International Journal of Man - Machine Studies, 1991. **34**: p. 319-336.
5. Friedlander, N., Schlueter, K., and Mantei, M. *Bullseye! When Fitt's Law doesn't fit*, in *ACM CHI*, 1998, p. 257-264.
6. Goose, S. and Safia, D., *WIRE: Driving Around the Information Super-Highway.* Personal and Ubiquitous Computing, 2002. **6**(3): p. 164-175.
7. Moore, B. C. J., *An introduction to the Psychology of Hearing*, 3rd edition 2001: Academic Press Limited, San Diego, CA, USA.
8. 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.
9. Sifelman, L., Arons, B., and Schmandt, C. *The Audio Notebook. Paper and Pen Interaction with Structured Speech*. in *SIGCHI 01*, 2001. Seattle, WA, USA: ACM. p. 182-189