Evaluation of Hear-Through Sound Localization

Georgios Marentakis
marentakis@iem.at

Rudolfs Liepins
rudolf.liepins@student.kug.ac.at

Institute for Electronic Music and Acoustics
University of Music and Performing Arts Graz
Inffeldgasse 10/3, A-8010, Österreich

ABSTRACT

Listening and interacting with audio relies commonly on using earphones which limit the ability of users to perceive their auditory environment. Earphone sets that integrate miniature microphones on their exterior can, however, be used to hear-through the auditory environment. We present an evaluation study in which sound localization when wearing such a hear-through system is compared to normal earphones, open headphones and unblocked ears. Although localization performance is improved compared to open headphones, we find that it is compromised in comparison to listening without earphones because confusions of sound direction increase and localization judgment distributions are more dispersed and show a weaker correlation to the test directions. The implications of the results to human computer interaction and possible improvements to hear-through system design are discussed.

Author Keywords

auditory augmented reality;hear-through systems

ACM Classification Keywords

H.5.2 User Interfaces: Auditory (non-speech) feedback; H.5.1 Multimedia Information Systems: Artificial, Augmented, and Virtual Realities

INTRODUCTION

Earphones are used quite commonly both for listening to music and interactive audio display with applications, in navigation [15, 17], interaction design for the visually impaired, and eyes-free interaction, in which binaural technology is used to enable gesture interaction with spatialized sound [2]. Performance in eyes-free interaction using audio can be comparable and in some applications even better compared to interaction with a visual display [19]. Earphones and headphones however, attenuate the sound from the environment. This is sometimes desirable, but can be both inconvenient (e.g. when trying to communicate with a person) and dangerous (such as when crossing roads). Injuries caused by wearing headphones have almost tripled since 2004 [10]. Importantly, although interaction with real world objects using vision has received attention within HCI research [14], interaction with our auditory environment has not received enough of researchers' attention. This is not surprising as most auditory display designs rely on using earphones and headphones which occlude sound from the environment.

This shortcoming could be overcome by using microphones integrated onto the earphone capsule, similar to what happens in noise cancelation headphones. These could pick up sound from the environment which rather than used for noise cancellation, could be simply played back by the earphones (or mixed with music or an auditory display) thereby improving its audibility. Müller and Karau [12] coined the term *Transparent Hearing* to describe the situation of hearing *through* our headphone sets. Such earphones received attention within auditory augmented reality [8]. It was pointed out that sound quality in such systems can be improved by compensating for resonances due to the blocked ear-canal using an inverse filter. As ear-canal transfer functions (ECTF) depend on earphone placement and the individual's anatomy, often a compromise has to be achieved by using generalized ear-canal functions, either from an individual or a dummy head [8, 16].

Despite the amount of sound quality evaluation of earphone based hear-through systems, sound localization has not been formally evaluated. Listeners wearing a hear-through earphone reported an accurate spatial impression when the experimenter spoke or made finger snaps at different locations in a room [16]. Similarly, Albrecht et al. [1] mentions that localization works well with such systems. Härmä et al. [8] did not perform a formal localization study, rather focused on whether virtual spatialized sounds rendered through the headset could be differentiated from real ones played through loudspeakers. Recent measurements [4] question these observations by showing that the location of the microphone outside the ear canal distorts the frequency response of the pinnae in an individualized way above 4kHz. Given the importance of sound localization in the aforementioned everyday and interactive settings, we performed and present a localization study that assesses the extent to which sound localization is affected by the hear-through earphone system.

EXPERIMENT

In a user study, we obtained localization judgments for sounds originating on, above, and below the horizontal plane in four different conditions: A. Unblocked, i.e. without earphones, B. Blocked, i.e. with earphones, but hear-through disabled, C. Hear-through, where microphone signal was played back by the earphone (Roland CS-10EM) and D. Open Headphones

CHI'14, April 26–May 1, 2014, Toronto, Canada.

http://dx.doi.org/10.1145/2556288.2557168

CR(%)	UBL	BLK	HTR	OHD
Front-Back	14.5(1.1)	30.7(3.1)	29.8(3.4)	51.1(3.6)
Back-Front	10.4(1.3)	22.6(2.8)	24.1(2.4)	27.6(2.9)
Total	12.5(1.5)	26.7(2.5)	26.9(2.9)	39.4(2.9)
Up-Down	0(-)	1.0(0.3)	2.0(0.3)	8.6(1.1)
Down-Up	4.0(0.4)	36.1(2.2)	61.6(2.3)	45.6(1.9)
Total	1.4(0.4)	14.1(2.1)	24.3(2.2)	22.5(2.0)

Table 1. Mean confusion rate and standard error in the four listening conditions averaged over the test directions in the experiment. UBL = Unblocked, BLK = Blocked, HTR = Hear-through and OHD = Open Headphones.

(Senheiser HD600), which was included in the experiment as a baseline headphone condition.

Apparatus & Materials

Participants sat on a height adjustable chair at the centre of a 2m-radius sphere, their nose aligned with the horizontal plane. As directional hearing is approximately symmetrical around the median plane, sixteen small loudspeakers were placed on their left side and were rendered invisible using acoustically transparent screens (Figure 1). Loudspeakers used the Peerless PLS-P830983 driver and were driven by a Class-D amplifier (TI-TPA3122 chipset) connected to a RME Fireface 800 audio interface and a notebook computer which controlled the experiment. Frequency response of the system remained within ±3dB from about 200 Hz to 12kHz. The Roland CS-10EM headset (Blocked and Hear-through) and the Senheiser HD600 headphone (Open Headphone) were used. The experiment was performed in a normal room (∼6x4m^2) without acoustic treatment. RT_{60} was 1.03, 0.70, 0.80, 0.63 sec at 50, 63, 80, 100 Hz respectively and less than 0.6 sec thereafter. Using a graphical user interface running on the laptop, participants placed two indicators corresponding to the perceived azimuth and elevation of the sound source on a graphical model of a head projected on the azimuth and the elevation plane respectively.

Stimulus was 500ms of white noise. Impulse responses for the through, blocked, and unblocked conditions were measured at (0°,0°) using the sine-sweep method on a Brüel&Kjær dummy head. Their RMS Level was adjusted, relative to the unblocked condition, at -5 dB in the through and -10 dB in the blocked condition. Unblocked SPL at the listening position was 60 dB for all tested directions. In the through condition, there was a latency of ∼10ms between the time the signal was picked up by the microphone and playback started.

N	Az	El
1	0	45
2	-40	45
3	-90	45
4	-140	45
5	180	0
6	0	0
7	-40	0
8	-60	0
9	-90	0
10	-120	0
11	-140	0
12	180	0
13	140	0
14	0	-45
15	-40	-45
16	-120	-45

Figure 1. Left: Experiment setup and Right: the sound directions used in the experiment.

Participants

Twelve participants participated, 19-45 years of age, five female and seven male; all reported having normal hearing.

Procedure

Participants could listen to stimulus twice in each trial to facilitate the separate azimuth and elevation judgments. Presentation of the four conditions was counterbalanced and stimulus direction was randomized in each trial. Participants responded using the GUI and clicked to continue. First, 16 training trials were given, thereafter 5 repetitions for each sound direction and test condition, a total of 90 trials per condition and 270 trials for the experiment which took about half an hour.

Results

Confusions: Up-Down and Down-Up confusions (excluding directions with 0° azimuth), as well as Front-Back and Back-Front confusions (excluding directions with 90° azimuth) are presented in Table 1.

There were more front-back confusions for all systems and test directions in comparison to the Unblocked condition. A two-way (System x Direction) repeated-measures ANOVA yielded significant main effects of System, F(3,33) = 24.368, $p < 0.001$, Direction, F(13,143) = 8.618, $p < 0.001$ and a significant interaction between System and Direction F(39,429) = 2.03, $p < 0.001$. In t-tests, front-back confusion rate was higher in all conditions compared to the Unblocked Condition ($p < 0.001$), and lower in the Hear-through and Blocked conditions compared to the Open Headphone Condition ($p < 0.01$). There was no difference between Blocked and Hear-through conditions. Direction 15 yielded the highest confusion rate ($p < 0.01$), followed by and 4, 5 and 14, with no significant difference for the rest of the stimulus directions. The interaction between System and Direction is because confusion rate was similar for all directions for Open Headphone condition, while it varied consistently with direction in the other conditions.

Interestingly, there were significantly more down-up than up-down confusions on average, t(11)=4.5066, $p < 0.001$, implying that sounds below the horizontal plane were significantly more likely to be confused as coming from above than the opposite. A two-way (Condition x Direction) repeated-measures ANOVA, performed for up-down and down-up confusions separately, yielded only a main effect of Condition, F(3,33) = 5.659, $p = 0.003$ and F(3,33)=13.385, $p < 0.001$ respectively. In pairwise comparisons (t-tests), there were significantly more up-down confusions for Open Headphones in comparison to all other systems ($p < 0.01$), but no other difference. There were significantly more down-up confusions in Hear-through and Open Headphone condition ($p < 0.05$) compared to the other two, and significantly less down-up

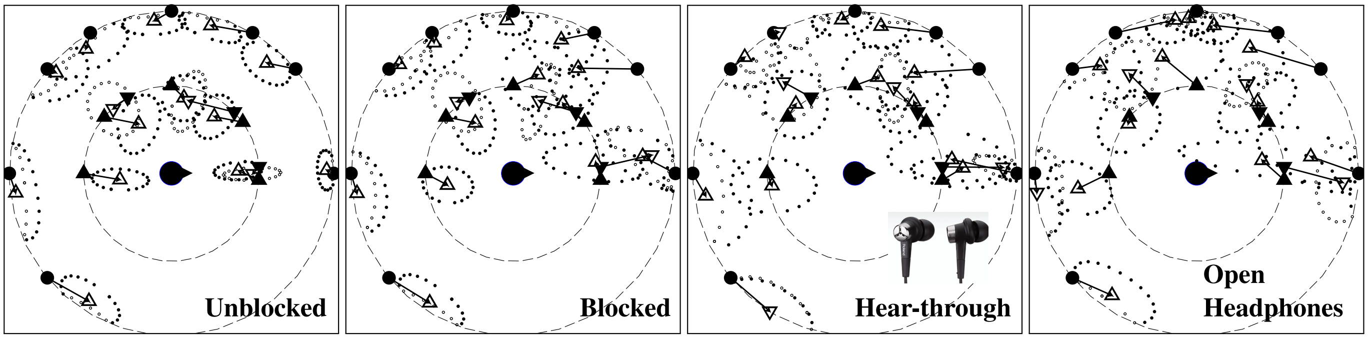

Figure 2. Distributions of localization judgments in the four conditions in the experiment. Open Symbols denote original listening locations and closed the centroid of the localization judgments. Dotted ellipsoids contain one standard deviation of measurements, along the two major axes of measurements dispersion. Triangles facing upwards/downwards indicate locations/centroids above/below the horizontal plane and circles at the horizontal plane. A photo of the Roland CS-10EM is included in the hear-through panel.

confusions in the Unblocked condition compared to all other conditions ($p < 0.01$).

Localization Error: Pooled estimates of the localization judgment distributions (Figure 2) were calculated after excluding trials in which confusions occurred, to avoid distorting the distributions by correcting confused judgements [3]. Spherical correlation coefficients between the judgment centroids in the four conditions in the experiment and the original test directions were 0.96, 0.97, 0.91, 0.85 for the Unblocked, Blocked, Hear-through and Open Headphone conditions respectively. According to the test proposed in [11], the correlation coefficient for the unblocked condition was found to be significantly higher than the rest at the 0.05 level; there was no significant difference between the other three.

Localization judgment dispersion around their centre of gravity was estimated using the concentration parameter K [6]. Dispersion increased in all conditions compared to the Unblocked one. K values were approximately normally distributed across locations. A one-way ANOVA with Condition as independent variable and K in each of the test directions as repeated measures variable yielded a main effect of Condition F(3,45)=7.40, $p < 0.01$. In pairwise comparisons (using t-tests across locations), it was found that the control condition resulted in significantly smaller dispersion than the rest, (t(15) = 2.39, $p = 0.03$ vs. Blocked, t(15) = 2.63, $p < 0.01$ vs. Hear-Through, and t(15) = 2.63, $p = 0.02$ vs. Headphones), but no other differences proved significant.

No effects were observed when comparing azimuth error in any of the conditions in the experiment, despite a small tendency for test directions in the horizontal plane to be pulled towards $90°$. While elevation judgments in the Unblocked condition were slightly overestimated for sources below or above the horizontal plane, there was a tendency for perceived directions to be gradually squeezed in between the horizontal plane and the $\pm 45°$ elevation in the other conditions. Elevation error decreased from $11.92°$ in the Unblocked, to $4.0°$ in the Blocked, to $-3.9°$ in the Hear-through, and to $-5.4°$ in the Open Headphones Condition for sources above and increased from $-3.6°$, to $8.7°$, $18.2°$ and $20.1°$ for sources below the horizontal plane. A one-way ANOVA with System as independent variable and elevation error for each direction as repeated-measures variable performed separately for test directions above and below the horizontal plane, yielded a significant effect of System; F(3,12) = 5.25, p = 0.02 for sources above and F(3,6) = 5.51, $p = 0.04$ for sources below the horizontal plane.

DISCUSSION

The presence of the earphone/headphone systems that were examined results in degradation in sound localization performance as evidenced by the weaker spherical correlation coefficients, the higher number of front/back and up/down confusions, and the increased dispersion of the localization judgements observed in the conditions in which an earphone/headphone was used. There was no significant influence on perception of azimuth, arguably due to the preservation of Interaural Time Differences which dominate azimuth perception in the horizontal plane thereby overruling distorted level and spectral cues [18]. Source elevation was underestimated and in comparison to up-down, significantly more down-up confusions occurred for all systems. On average, elevation error was higher below the horizontal plane. Consequently, the induced distortion is direction-dependent and affects strongly locations below the horizontal plane. Acoustical measurements would be necessary to fully disambiguate this aspect, which is challenging as the frequency range within which elevation cues are contained depends on the individual and is therefore difficult to establish.

The spectral cues that support elevation perception and front/back discrimination were poorly preserved by the systems. Concerning the hear-through system, we attribute the spectral distortion primarily to the microphone placement away from the ear-canal entrance. Microphone positioning affects the accuracy of the HRTF sampling [4], therefore microphones should have been placed at, or few millimeters away from the ear canal entrance [7]. Equalizing the blocked ear-canal frequency response could potentially yield further improvements [13]. Sensors on the ear canal side of the earphone could enable the dynamic estimation of the the ECTF; such equalization however, did not reduce confusion rates significantly in HRTF reproduction [9]V. The extent to which the leakage, that inadvertently occurs with earphones, influenced localization judgments is not clear. We asked participants to fit the earphones as well as they could and included a delay of 10ms (256 samples@44100Hz) typical when routing through a computing device or applying audio pre-processing. An investigation of the impact of latency in hear-through systems was outside the scope of this study.

Our findings suggest that auditory mixed reality applications that use hear-through systems should not be used in safety critical contexts (e.g. biking or driving) in which effective and efficient auditory localization of environmental sound is important. In familiar environments, in the presence of visual cues, or when sounds are long enough to allow for head-movements, the localization problems noted here will manifest themselves less; localization with hear-through earphones was sufficient in the presence of visual cues [16]. Such systems could therefore potentially extend the scope of interactive auditory displays [17, 2, 15] by enabling increased awareness and interaction with environmental sound. The visually impaired would also benefit by such systems in that they could both interact with computers and maintain contact with their auditory environment while avoiding the privacy and noise issues that occur when using loudspeakers in an office for example. Importantly, the application of well-developed frameworks concerning interaction with virtual spatialized audio [5] on interaction with environmental sound would enable designers to explore exciting possibilities for interaction design in a structured way.

CONCLUSION

We presented a study that evaluated sound localization using a hear-through earphone system. We found that although an advantage appears in comparison to listening with open headphones, sound localization as well as front-back and up-down discrimination are compromised due to poor reproduction of spectral cues. Although such systems already make a difference in environmental sound awareness, it appears that better earphone designs are necessary to maintain good sound localization and support interaction in the absence of visual cues.

ACKNOWLEDGMENTS

This work was supported by project Klangräume: Situated Usability Evaluation of Interactive 3D Audio Environments (PN:6039) awarded from Zukunftsfonds Steiermark, Land Steiermark, Österreich.

REFERENCES

1. Albrecht, R., Lokki, T., and Saviola, L. A mobile augmented reality audio system with binaural microphones. In *Proceedings of IWS* (Stockholm, Sweden, 2011), 7–11.
2. Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal 'eyes-free' interaction techniques for wearable devices. In *CHI '03*, ACM (2003), 473–480.
3. Carlile, S., Leong, P., and Hyams, S. The nature and distribution of errors in sound localization by human listeners. *Hearing Research 114*, 1-2 (1997), 179–96.
4. Christensen, F., Hoffmann, P. F., and Hammershøi, D. Measuring directional characteristics of in-ear recording devices. In *Audio Engineering Society Convention 134* (May 2013).
5. Cohen, M., and Ludwig, L. F. Computer-supported cooperative work and groupware. Academic Press Ltd., London, UK, UK, 1991, ch. Multidimensional Audio Window Management, 193–210.
6. Fisher, N. I., Embleton, B. J. J., and Lewis, T. *Statistical analysis of spherical data / N.I. Fisher, T. Lewis, B.J.J. Embleton*. Cambridge University Press, Cambridge [England] ; Melbourne :, 1987.
7. Hammershøi, D., and Møller, H. Sound transmission to and within the human ear canal. *The Journal of the Acoustical Society of America 100*, 1 (1996), 408–427.
8. Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., and Lorho, G. Augmented reality audio for mobile and wearable appliances. *J. Audio Eng. Soc 52*, 6 (2004), 618–639.
9. Hiipakka, M. *Estimating pressure at the eardrum for binaural reproduction*. PhD thesis, Department of Signal Processing and Acoustics, Aalto University, Finland, 2012.
10. Lichenstein, R., Smith, D. C., Ambrose, J. L., and Moody, L. A. Headphone use and pedestrian injury and death in the united states: 2004 - 2011. *Injury Prevention* (2012).
11. Meng, L.-X., Rosental, R., and Rubin, D. Comparing correlated correlation coefficients. *Psychological Bulletin 111*, 1 (1992), 172–175.
12. Mueller, F., and Karau, M. Transparent hearing. In *CHI 2002* (2002), 730–731.
13. Pralong, D., and Carlile, S. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *The Journal of the Acoustical Society of America 100*, 6 (1996), 3785–3793.
14. Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., and Chin, J. An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In *UbiComp'06*, Springer-Verlag (Berlin, Heidelberg, 2006), 87–104.
15. Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., and O'Modhrain, S. Gpstunes: Controlling navigation via audio feedback. In *MobileHCI '05*, ACM (2005), 275–278.
16. Tikander, M. *Development and Evaluation of Augmented Reality Audio Systems*. PhD thesis, Helsinki University of Technology, 2009.
17. Vazquez-Alvarez, Y., Oakley, I., and Brewster, S. A. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing 16*, 8 (2012), 987–999.
18. Wightman, F. L., and Kistler, D. J. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America 91*, 3 (1992), 1648–1661.
19. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P. Earpod: Eyes-free menu selection using touch input and reactive audio feedback. In *CHI '07*, ACM (New York, NY, USA, 2007), 1395–1404.