ICSV14
Cairns • Australia
9-12 July, 2007

14th International Congress on Sound & Vibration

PSYCHOACOUSTIC EVALUATION OF A NEW EARPIECE FOR AUGMENTED-REALITY AUDIO

Craig T. Jin[1], Aengus Martin[1] and André van Schaik[1]

[1]School of Electrical and Information Engineering, University of Sydney
Sydney, NSW 2006, Australia
craig@ee.usyd.edu.au

Abstract

An augmented-reality audio system ideally provides a means to present virtual auditory space stimuli to a listener, without distorting the listener's normal spatial hearing. In this work, we conduct a psychoacoustic experiment to evaluate both the virtual and real sound localisation fidelity of an augmented-reality audio system based on acoustically-transparent ear shells. In addition to being acoustically-transparent, the ear shells are comfortable to wear and relatively discrete and thus provide an attractive platform for presenting augmented-reality spatial audio. The results of the sound localisation tests indicate that high-fidelity spatial audio, i.e., spherical correlation coefficient generally greater than 0.87, but with individual variations, can be obtained.

1. INTRODUCTION

Mobile augmented-reality audio presents a relatively new growth area within personal audio systems. The last decade has certainly witnessed extensive growth in personal audio systems in the form of mobile telephony and personal music devices and now a leading researcher at the Nokia Research Center predicts that the next big paradigm shift in mobile telephony is the convergence of mobility and the internet [1]. Within this framework, it is also claimed that mobile virtual and augmented-reality audio will play an increasing role in the future diversification of audio user interfaces. An all-inclusive review of current mobile and wearable augmented-audio applications and methods can be found in [2, 3].

Current methods for augmented-reality audio rely on a binaural earphone-microphone system that consists of an earpiece worn in each ear and in which the earpiece consists of the combination of an earphone and microphone. Using current methods, the real acoustic environment is recorded by the microphone and played directly back via the earphone. The great difficulty with this situation is that there is invariably latency associated with the simulated real acoustic environment and also leakage of the real acoustic environment through the earpiece. The combination of leakage and latency can lead to comb filtering in the perceived audio. The fact that this simulation of the real acoustic environment is not quite veridical is indicated by the fact that it is frequently referred to as the pseudo-acoustic environment [2]. The issue of dealing with the problem of leakage and latency is extremely difficult. Sealing the ear reduces the leakage, but introduces the occlusion effect, which accounts for the hollow or booming sound of the listener's own voice and the unpleasant loudness when chewing food that occurs because low frequency vibrations cannot escape the ear canal. This effect generally becomes intolerable with prolonged use of the earphones.

In this paper, we present a novel technique for augmented-reality audio in which we use an "acoustically-transparent" ear shell coupled with an earphone. This device can pass virtual audio and is relatively open or transparent to the real acoustic environment. Thus, there is no need for a microphone to directly pass the real acoustic environment to the earphone. In this way, we eliminate all of the leakage and latency issues that are associated with rendering the real acoustic environment as described above. The focus of this paper is an evaluation of the proposed method for augmented-reality audio in terms of the fidelity of the spatial hearing that can be rendered in both the real, free-field acoustic environment and in virtual auditory space (VAS).

2. A NEW EARPIECE FOR AUGMENTED AUDIO

The new augmented-audio headset that we consider consists of a combination of an acoustically-transparent ear shell manufactured by Surefire, LLC, either the CommEar™ Comfort EP1 (EP1) and the CommEar™ Boost EP2 (EP2), and also the Etymōtic Research ER4P MicroPro earphones (ER4P) (see Figure 1). The acoustically-transparent ear shells are designed to fit snugly within the conchal cavity of the ear and they are made with a resilient polymer. The left and right ear shells are mirror symmetric and come in three sizes for the EP1s and two sizes for the EP2s. The ER4P earphones have a manufacturer-cited frequency response of 20 Hz -16 kHz ± 4 dB using the supplied ear tips, and a 1 kHz sensitivity of 108 dB SPL for a 0.2 V input and a nominal impedance of 27 ohms.

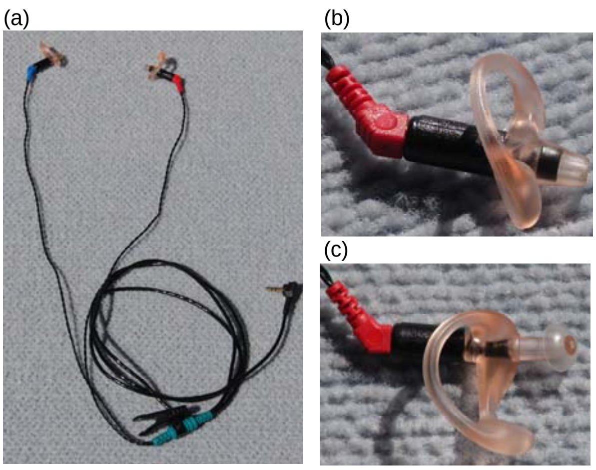

Figure 1. (a) The augmented-audio headset consists of the Etymotic Research ER4P device connected with the acoustically-transparent Surefire EP1 ear shell. The EP1 and EP2 Surefire acoustically-transparent ear shells are shown in (b) and (c), respectively. The EP1 ear shell is open while the EP2 ear shell has a flange on the end.

The transfer function of the augmented-audio headset comprising the ER4P drivers and the EP1/EP2 earpieces was measured using a log-sine sweep signal (10s, 50 Hz – 22 kHz) and a Brüel and Kjær Head and Torso Simulator (HATS 4128C) mannequin. It was found that the transfer function of the ER4P-EP2 headset varied substantially more than that for the ER4P-EP1 headset. The variation of the ER4P-EP2 headset for repeat measurements is shown in Figure 2a. For the synthesis of the VAS stimuli, we compensate for the transfer function of the augmented-audio headset. More specifically, a minimum-phase compensation filter is applied such that the system comprising the augmented-audio headset plus compensation filter has a transfer function equivalent to that of the Etymōtic Research ER1 earphones. The ER1 earphones are designed to have a flat response except for a simulated ear-canal resonance and are ideal for the synthesis of VAS. The compensation filter was created using transfer function measurements that were made with the ER1 earphones and HATS. Repeat measurements of the transfer function of the ER4P-EP1 headset with compensation filter applied is shown in Figure 2b. Because the EP1 ear shells are open, there is a significant roll-off of 15 dB/octave or so in the transfer function below 800 Hz.

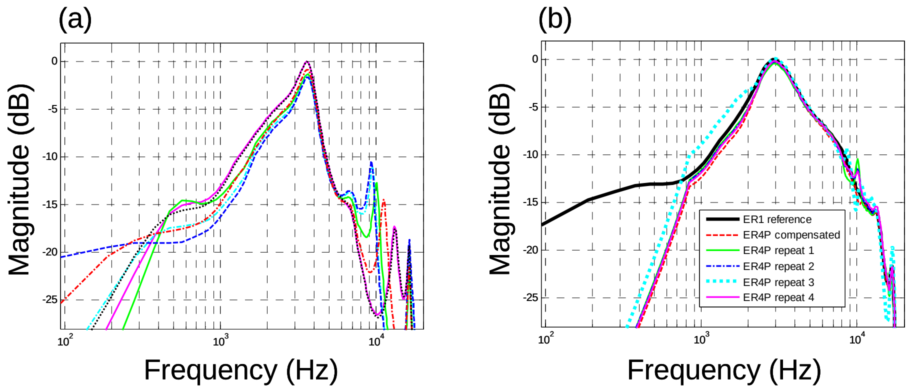

Figure 2. The magnitude frequency spectra of six repeat impulse response measurements of the ER4P-EP2 headset are shown in (a). The magnitude frequency spectra of five repeat impulse response measurements of the ER4P-EP1 headset with a compensating filter applied is shown in (b) along with the reference ER1 magnitude spectrum.

3. METHODS: SOUND LOCALISATION

3.1 Subjects, environment, and stimuli

Four subjects participated in the experiments (all male, aged 26-39) with two subjects having previous experience with auditory localisation experiments and two subjects relatively new to the testing paradigm. Localisation testing was conducted in a triple walled anechoic sound chamber. Within the chamber, the subject stood on a height-adjustable platform which was adjusted so that the subject's head was positioned in the centre of the chamber. Inside the chamber there is a robotic arm that is configured as a double hoop system that can revolve about the subject (see Figure 3). A loudspeaker is mounted on the robotic arm and delivers sound stimuli from locations on an imaginary sphere of one meter radius around the listener's head. There is also an array of LEDs to provide head-orientation feedback.

The subjects' free-field sound localization performance was tested for three different sound conditions referred to as FF control, FF EP1 and FF EP2. The FF control sound condition refers to the control sound condition in which the subject did not wear any augmented-audio headset. In the FF EP1 and FF EP2 sound conditions, the subject wore a binaural augmented-audio headset comprising the ER4P earphones and the EP1 and EP2 ear shells, respectively. The purpose of these three sound conditions was to examine the influence of the augmented-audio headset on free-field sound localization. The sound stimuli consisted of a freshly-generated 150 ms Gaussian white noise with a 15 ms raised-cosine onset and offset ramps.

Figure 3. The environment and equipment for auditory localisation testing are shown. The subject is standing on a platform and her left hand holding on to a response button. The subject is looking at the loudspeaker mounted on the semicircular robotic hoop. The subject is wearing a headband to which a head-tracking device is fastened. A support stand is shown which holds an array of LEDs for initial head alignment.

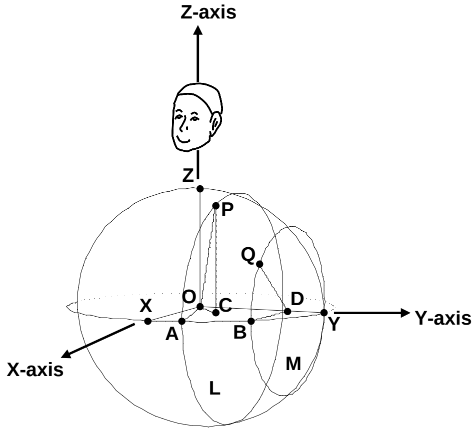

Figure 4. The lateral angle is defined with reference to the point P and is given by $\angle$ XOA and is the horizontal angle out to the vertical plane L. The polar angle is defined with reference to point Q and is given by $\angle$ BDQ and is the vertical angle from the horizontal plane.

The subjects' VAS sound localization performance was tested for two different sound conditions referred to as VAS Control and VAS EP1. The VAS Control and VAS EP1 sound conditions refer to presentation of VAS sound stimuli using the Etymōtic Research ER1 earphones and the ER4P-EP1 headset, respectively. The ER1 earphones are standard earphones used to present VAS, so the former was considered a control condition. We also chose not to test VAS localisation performance using the ER4P-EP2 headset because of the sensitivity of its transfer function to precise placement in the ear.

3.2 Localisation testing and paradigm

A full description and validation of the localisation testing paradigm is provided in [4] and a brief description is given here. In this paradigm, the subject stands in darkness in the centre of the anechoic chamber and indicates the perceived direction of a series of broadband noise bursts presented from 76 random locations evenly distributed around an imaginary sphere surrounding his/her head. The subject was required to localise the sound by turning squarely around and pointing his/her nose toward the perceived direction of the sound source. An inertial head-orientation tracker (Intersense InertiaCube3) was mounted on top of the subject's head using a headband and was used to measure the subject's head orientation and thus provide an objective measure of the perceived sound direction. The subjects performed five repeat trials of 76 localisation tests for each sound condition.

3.3 Generation of virtual auditory space

Virtual auditory space (VAS) is generated by recording the acoustic filter functions of the auditory periphery of individual listeners (the head-related transfer functions: HRTFs) and convolving these filter functions with sounds subsequently presented over earphones to create a realistic, externalized percept. HRTFs were measured using a "blocked-ear" recording technique. This approach involves embedding a small recording microphone in an earplug secured flush with the distal end of the ear canal [5]. The recordings were performed at 393 locations around the sphere in the anechoic chamber with the subject at the centre.

3.4 Data analyses

The overall localisation performance of subjects in the different experimental conditions was measured using the spherical correlation coefficient (SCC) [6]. Its use with localisation data is described in detail elsewhere [7], but in brief, it describes the degree of correlation between the target and response locations (1 = perfect correlation; 0 = no correlation).

In order to analyse the detailed pattern of localisation responses more closely, the localisation data were analysed in terms of lateral-polar angle coordinate system (see Figure 4). In addition, we calculate the percentage of cone of confusion errors made by subjects in each condition. These were defined as large polar angle errors (> 90º in magnitude) and as such include the commonly reported front-back and up-down confusions.

4. RESULTS: SOUND LOCALISATION

4.1 Free-field sound localisation

The localisation performance data are analysed in terms of lateral and polar angles. The lateral angle performance data were similar across the four subjects so they were pooled across subjects for analysis (see Figure 5). The mean size of the lateral angle error for the FF Control, FF EP1, and FF EP2 sound conditions is 7.4º, 7.8º, and 9.5º, respectively (see Figure 7a). A Kruskal-Wallis non-parametric one-way ANOVA (KW ANOVA) was performed on the group data to compare the FF Control and FF EP1 sound conditions and it revealed a slightly significant effect of condition [$\chi^2(1)=4.18$, $p=0.041$]. Two similar statistical analyses comparing the FF Control and FF EP2, and the FF EP1 and FF EP2 sound conditions revealed a highly significant effect of condition in both cases, [$\chi^2(1)=49.57$, $p<0.0001$] and [$\chi^2(1)=25.48$, $p<0.0001$], respectively.

The polar angle performance data were also similar across subjects. They were pooled across subjects and analysed as a group (see Figure 6). The test sound conditions show an increase in the number of front-back errors which were removed before analysing the mean size of the polar angle error (Figure 7b). The average size of the polar angle error for the FF Control, FF EP1, and FF EP2 sound conditions is 12.6º, 16.5º and 20.1º, respectively. A KW ANOVA was performed on the group data to compare the FF Control and FF EP1 sound conditions and it revealed a highly significant effect of condition [$\chi^2(1)=51.37$, $p<0.0001$]. Two similar statistical analyses comparing the FF Control and FF EP2, and the FF EP1 and FF EP2 sound conditions also revealed a highly significant effect of condition in both cases, [$\chi^2(1)=130.92$, $p<0.0001$] and [$\chi^2(1)=21.99$, $p<0.0001$], respectively. The SCC and percentage cone of confusion errors for each subject and each condition, as well as the average values across the population are shown in Table 1.

4.2 Virtual auditory space sound localisation

The lateral angle performance data were similar across the four subjects and thus pooled for analysis (see Figure 8). The mean size of the lateral angle error for the VAS Control and VAS EP1 sound conditions is 8.7º and 8.6º, respectively (see Figure 10a), compared to 7.4º in the free field control case. A KW ANOVA was performed on the group data to compare the VAS Control and VAS EP1 sound conditions and it revealed no significant effect of condition [$\chi^2(1)=2.19$, $p=0.1388$]. Two similar analyses comparing the VAS Control and FF Control, and the VAS EP1 and FF Control sound conditions revealed a highly significant effect of condition in both cases, [$\chi^2(1)=44.71$, $p<0.0001$] and [$\chi^2(1)=26.25$, $p<0.0001$], respectively.

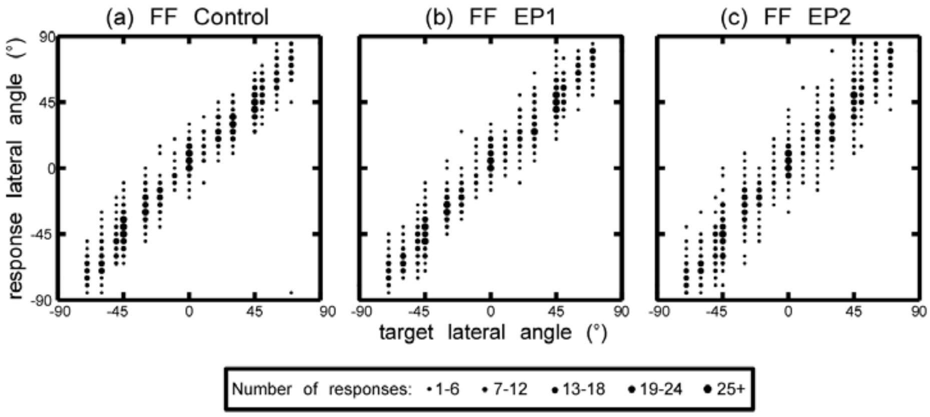

Figure 5. A scatter plot of the lateral angles associated with the free-field localisation performance data is shown. The data have been pooled across subjects. The target lateral angle is plotted against the response angle and the size of the dots represents the number of responses clustered at a point.

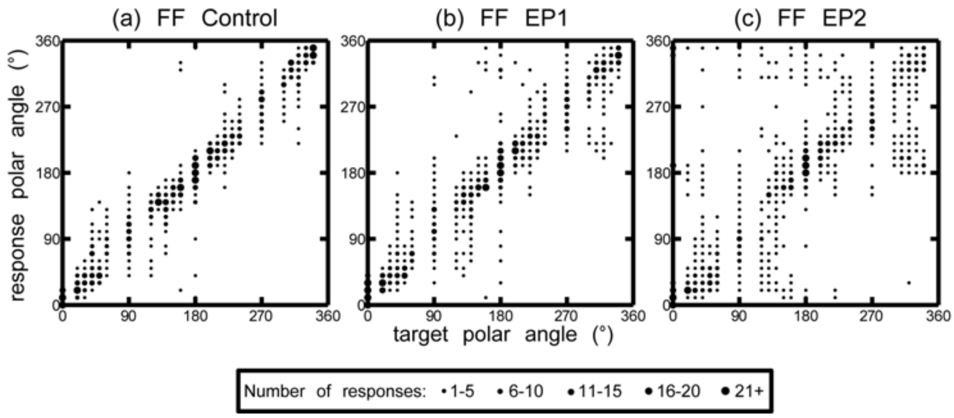

Figure 6. A scatter plot of the polar angles associated with the free-field localisation performance data is shown. The data have been pooled across subjects. The target polar angle is plotted against the response angle and the size of the dots represents the number of responses clustered at a point.

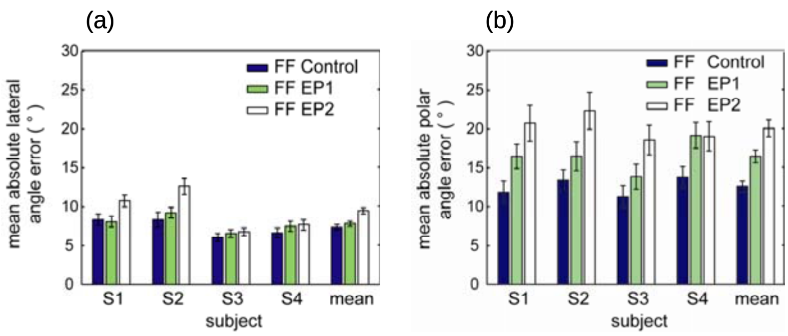

Figure 7. The mean absolute lateral angle error and its 95% confidence interval is shown for each subject and free-field sound condition in (a). The mean absolute polar angle error and its 95% confidence interval is shown for each subject and free-field sound condition in (b). The average angle error across the subject population is also shown.

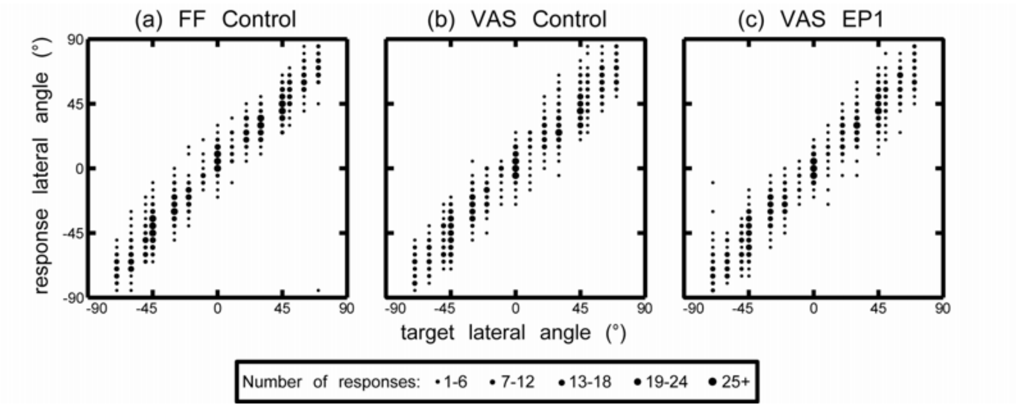

Figure 8. A scatter plot of the lateral angles associated with the VAS localisation performance data is shown. See Figure 5 for additional details.

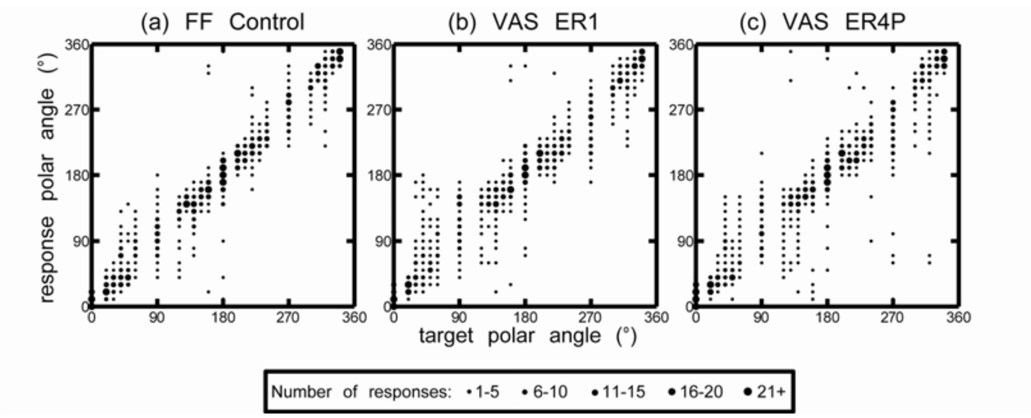

Figure 9. A scatter plot of the polar angles associated with the VAS localisation performance data is shown. See Figure 6 for additional details.

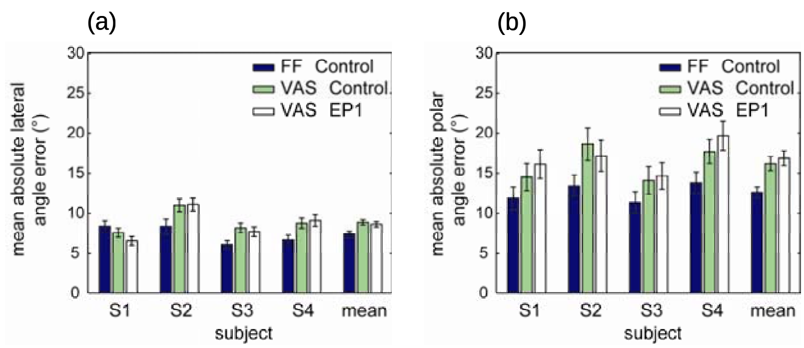

Figure 10. The mean absolute lateral angle error and polar angle error and their 95% confidence intervals are shown for each subject and VAS sound condition in (a) and (b).

Table 1. The spherical correlation coefficient and percentage of cone of confusion errors is shown for all experimental conditions for each subject as well as the mean across the subject population.

	S1	**S2**	**S3**	**S4**	**Mean**
FF Control	0.91 (2.9%)	0.91 (2.9%)	0.93 (1.8%)	0.91 (3.2%)	0.91 (2.7%)
FF EP1	0.89 (5.0%)	0.88 (4.5%)	0.87 (5.3%)	0.86 (8.2%)	0.88 (5.7%)
FF EP2	0.73 (20.5%)	0.75 (21.8%)	0.81 (9.2%)	0.82 (11.8%)	0.78 (15.9%)
VAS Control	0.90 (3.42%)	0.87 (9.2%)	0.91 (4.47%)	0.87 (5.26%)	0.88 (5.59%)
VAS EP1	0.89 (3.16%)	0.88 (5.0%)	0.89 (4.47%)	0.84 (7.37%)	0.87 (5.0%)

The polar angle performance data were similar across the four subjects and therefore pooled across subjects for analysis (see Figure 9). Front-back errors were removed before analysing the mean size of the polar angle error. The mean size of the lateral angle error for the VAS Control and VAS EP1 sound conditions is 16.2º and 16.9º, respectively, compared to 12.6º in the free field control case (see Figure 10a). A KW ANOVA was performed on the group data to compare the VAS Control and VAS EP1 sound conditions and it revealed no significant effect of condition [$\chi^2(1)=1.28$, $p=0.2586$]. Two similar analyses comparing the VAS Control and FF Control, and the VAS EP1 and FF Control conditions, revealed highly significant effects of conditions in both cases, [$\chi^2(1)=27.68$, $p<0.0001$] and [$\chi^2(1)=42.06$, $p<0.0001$], respectively. The SCC and percentage cone of confusion errors for each subject and each condition, as well as the average values across all subjects are shown in Table 1.

6. DISCUSSION AND CONCLUSIONS

The free-field sound localisation results with the ER4P-EP1 and ER4P-EP2 augmented-audio headsets demonstrate that while control levels of localisation performance were not maintained, the average lateral and polar angle error were statistically significantly smaller for the ER4P-EP1 headset compared to the ER4P-EP2 headset. The VAS sound localisation results with the ER4P-EP1 headset also showed that free-field levels of control performance were not maintained, but that the localisation performance was at levels statistically equivalent to that obtained with the standard closed ER1 earphones. There are three important differences between the ER4P-EP1 and ER4P-EP2 headsets: (i) free-field localisation performance is statistically significantly better with the ER4P-EP1 headset; (ii) the ER4P-EP2 headset is better suited to preserving the sound levels below 800 Hz in the VAS presentation (see Figure 2); (iii) the sensitivity of the ER4P-EP2 transfer function to the exact placement within the ear is likely to cause substantial variations between 6 and 10 kHz, which will impair sound localisation performance. If we take the viewpoint that the first objective of augmented audio is not to impair normal free-field hearing, then the ER4P-EP1 headset is an excellent candidate for an augmented-audio headset. There is no need to record and then playback the real acoustic environment. This eliminates the latency/leakage problems associated with a microphone-earphone augmented audio headset and also reduces the computational and energy requirements of the system.

REFERENCES

[1] J. Huopaniemi, "Future of personal audio – smart applications and immersive communication," in *Proceedings of the AES 30th International Conference Intelligent Audio Environments* Saariselkä, Finland: Audio Engineering Society, Inc., 2007.

[2] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable applicances," *Journal of the Audio Engineering Society,* vol. 52, pp. 618-639, June, 2004.

[3] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, H. Nironen, and S. Vesa, "Techniques and applications of wearable augmented reality audio," in *Proceedings of the AES 114th Convention* Amsterdam, The Netherlands: Audio Engineering Society, Inc., 2003.

[4] S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," *Hearing Research,* vol. 114, pp. 179-196, 1997.

[5] H. Moller, M. F. Sorensen, D. Hammershoi, and C. B. Jensen, "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society,* vol. 43, pp. 300-321, May 1995.

[6] N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data*. Cambridge: Cambridge University Press, 1987.

[7] P. Leong and S. Carlile, "Methods for spherical data analysis and visualization," *Journal of Neuroscience Methods,* vol. 80, pp. 191-200, 1998.