7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Sound localization and speech identification in the frontal median plane with a hear-through headset

Pablo F. Hoffmann[1], Anders Kalsgaard Møller, Flemming Christensen, Dorte Hammershøi
Acoustics, Aalborg University, Fr. Bajers Vej 7, Aalborg Ø DK-9220, Denmark

[1]Now at Huawei Technologies Duesseldorf GmbH, ERC Riesstr. 25, 80993 Munich, Germany

Summary
A hear-through headset is formed by mounting miniature microphones on small insert earphones. This type of ear-wear technology enables the user to hear the sound sources and acoustics of the surroundings as close to real life as possible, with the additional feature that computer-generated audio signals can be superimposed via earphone reproduction. An important aspect of the hear-through headset is its transparency, i.e. how close to real life can the electronically amplified sounds be perceived. Here we report experiments conducted to evaluate the auditory transparency of a hear-through headset prototype by comparing human performance in natural, hear-through, and fully occluded conditions for two spatial tasks: frontal vertical-plane sound localization and speech-on-speech spatial release from masking. Results showed that localization performance was impaired by the hear-through headset relative to the natural condition though not as much as in the fully occluded condition. Localization was affected the least when the sound source was in front of the listeners. Different from the vertical localization performance, results from the speech task suggest that normal speech-on-speech spatial release from masking is unaffected by the use of the hear-through headset. This is an encouraging result for speech communication applications.

PACS no. 43.38.Md, 43.66.Pn

1. Introduction

A hear-through option in earphones exists when the earphones have microphones mounted on their outer surface. This gives the user the option of listening to the acoustics of the surroundings, which would otherwise be attenuated by the passive attenuation of the earphones. At the same time the earphones also enable binaural rendering of 3D audio signals that can be combined seamlessly with the hear-through real life binaural sounds. This combination of real-life and virtual sounds is often referred to as spatial augmented reality audio.

One relevant design aspect of a hear-through headset is the degree of acoustical transparency it can provide. Implementing an ideal hear-through headset so that it is fully acoustically transparent is not a straightforward task. Harma et al [1] coined the term pseudo-acoustic environment to refer to the modified version of the real acoustic environment that is typically presented to the user via a semi-ideal hear-through headset. The adjective 'semi-ideal' is used to indicate that the hear-through headset is not fully acoustically transparent. A key characteristic for transparency is to preserve the spatial information available in natural conditions. Size and geometry of the earphones together with the microphone placement modify the natural acoustics of the external ears that is transmitted via the hear-through headset. Occlusion of the concha by hear-through prototypes has been reported to alter high-frequency spectral localization cues [2]. Previous studies have shown that these alterations have a negative effect on sound localization performance [3, 4, 5, 6], particularly for localization along the vertical dimension.

Here we report on an experiment conducted to evaluate the auditory transparency of a hear-through prototype by comparing human performance in a frontal vertical-plane sound localization task and a speech identification task. These two tasks are compared between a natural condition i.e. with the ears naked, and when wearing the hear-through headset. We reasoned that if performances are similar between the

natural and hear-through conditions then we could conclude that the hear-through headset is perceptually transparent. Vertical sound localization was used because elevation perception cues primarily stem from the high-frequency information provided by the pinna and concha cues. Speech identification was used because we are also interested in the extent that the hear-through headset can preserve speech communication in a multi-talker environment. It is well known that normal spatial hearing enables selective attention to the location of a target talker in the presence of other spatially separated talkers or audio distracters, i.e. the so-called "cocktail party" effect [7]. This is particularly of interest in the context of immersive communication.

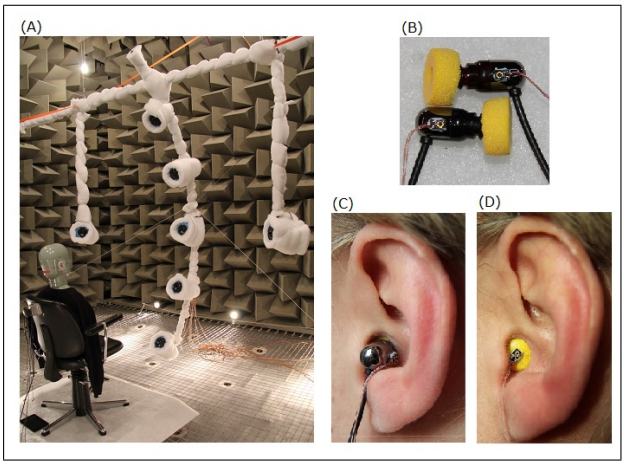

Figure 1. (A) Experimental setup with an array of 7 loudspeakers in an anechoic chamber. The frame used to hold the loudspeakers is wrapped with absorptive material that minimizes unwanted sound reflections from the setup. (B) hear-through earphone prototype combining insert earphones with mounted miniature microphones. (C) Example of a typical placement of the hear-through earphones in a human ear. The microphone is pointing towards the concha. (D) Miniature microphone placed at the entrance to the blocked ear canal. This position is considered as the ideal position to record all spatial sound information.

2. Methods

2.1. Listeners

Ten paid listeners (2 females and 8 males) took part in the experiments. Their ages ranged from 19 to 29. All listeners had absolute thresholds less than 20 dB hearing level at all audiometric frequencies (250 Hz to 8000 Hz in octave steps). Three listeners had previous experience in psychoacoustic tests, but none had experience on sound localization and speech identification experiments.

2.2. Apparatus

2.2.1. Loudspeaker array

All experiments were conducted in an anechoic chamber. As shown in Figure 1(A) the setup consisted of an array of 7 loudspeakers (Vifa M10MD-39 driver mounted in a 155-mm diameter hard-plastic ball) with 5 of them distributed along the sagittal median plane at elevations $\pm 45°$, $\pm 22.5°$, and $0°$. The remaining two loudspeakers were placed in the horizontal plane (vertical angle of $0°$) at $\pm 45°$ azimuth. The loudspeakers' frequency responses were all comparable without spectral characteristics particular to the individual loudspeaker that could have been used as unwanted cues for localization or speech identification. All digital audio signals (RME DIGI96) sent to the loudspeakers were D/A converted (RME ADI-8 DS) and amplified (ROTEL R8-976 MKII).

2.2.2. Hear-through headset

The hear-through headset was built by combining miniature MEMS microphones (Analog Devices ADM504) with insert earphones (Logitech EU700) (see Figure 1(B)). The sensitivities of the microphones were 14 mV/Pa and 12.5 mV/Pa for the left and right microphone respectively, and their frequency responses are shown in Figure 2(A). The insert earphones used balanced armature speakers. The earphones frequency response, measured in an occluded ear simulator (G.R.A.S RA0045), are characterized by a relatively flat response at low frequencies and moderate peaks at about 2 and 4.5 kHz (see upper curves in Figure 2(B)). The microphones were connected to a custom-made power supplier and amplifier (20 dB gain). The output of the amplifier was connected to the microphone input of a USB audio interface (Edirol QUAD-CAPTURE). The microphones signals were routed to the headphones output of the audio interface via custom made software (Portaudio v.19 with ASIO API). Buffering of audio samples was reduced to the smallest possible that allow for a glitch-free capturing and reproduction of sound. This resulted in a total hear-through latency of 7 ms. In the same software all necessary equalization was implemented as digital filters. This included a fourth-order infinite impulse response (IIR) digital filter to compensate for the microphone response (see Figure 2(A)), and a digital filter that reintroduced the natural acoustics of the open ear canal (see lower curves in Figure 2(B)). This filter was implemented as a cascade of five second-order IIR filters. The lower curves in Figure 2(B) show the frequency response of the hear-through earphones measured when calibrated using this filter (occluded ear simulator (G.R.A.S RA0045). The background sound pressure level (SPL) in the anechoic chamber was 30 dB(A). The measured background noise level with the hear-through on was 33 dB(A). This increment is in agreement with the 30-dB(A) self-noise specification of the MEMS microphones.

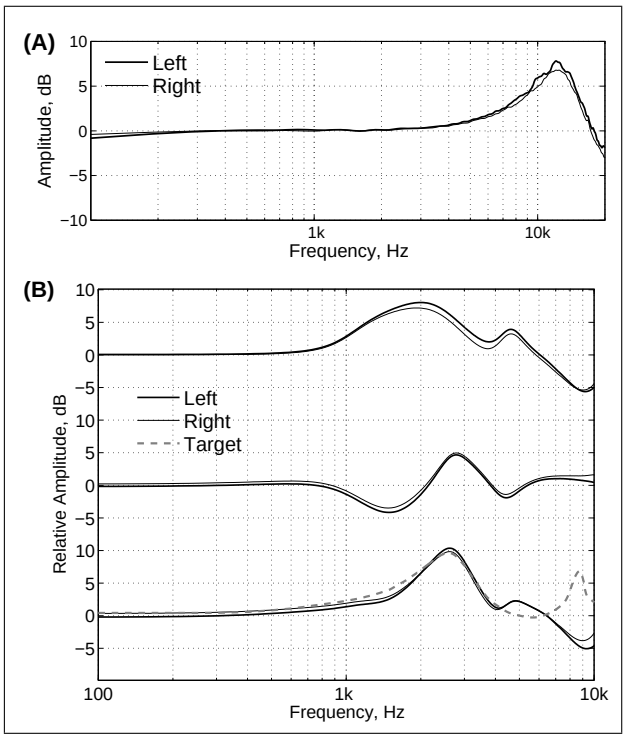

Figure 2. (A) Frequency response of the MEMS microphones. These responses were equalized so that the resulting response was flat over the range 200-20000 Hz. (B) Left (thick line) and right (thin line) earphone frequency response measured in an occluded ear simulator. The two top curves represent the normal response of the earphones. The middle curves are the hear-through equalization filters, and the two bottom curves are the resulting responses when equalized for hear-through applications. The dashed line indicates the target hear-through response corresponding to the transmission from the blocked ear canal entrance to the eardrum.

The reason for testing the psycho-acoustic transparency of the hear-through headset is because the size of our prototype does not allow positioning the microphones of the hear-through headset flush with the entrance to the blocked ear canal (see Figure 1(D)). We recognized this position as the ideal position for audio recording since all spatial information is present at the blocked ear canal entrance [8]. As seen in Figure 1(C) the hear-through earphone sticks out from the ear canal entrance, which means that the microphone is placed at a semi-ideal position, which compared to the ideal position introduces a change in the spectrum that might lead to distortions in the spatial information. With the following experiments we plan to assess the perceptual impact of these spectral changes.

In both experiments the order in which the two conditions (natural v/s hear-through) were presented was counterbalanced across subjects. In the hear-through condition the experimenter used foam eartips to couple the earphones to the ear canals (see Figure 1(B)). Once the earphones were in place the miniature microphones were carefully mounted to the earphones using putty (see Figure 1(C)). A visual mark on the earphones helped the experimenter to mount the microphones on the same position relative to the earphones.

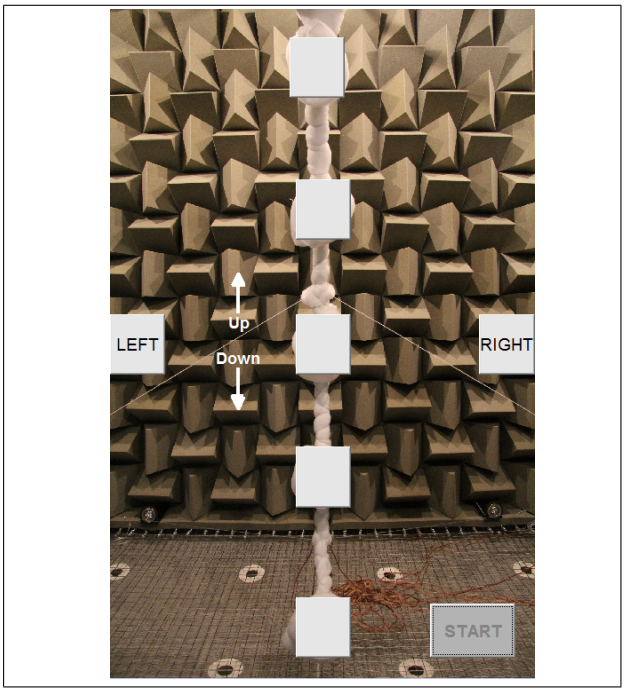

Figure 3. Graphical user interface used by listeners to input their response in the frontal-vertical plane sound localization task

3. Experiment 1 - Sound localization in the frontal vertical plane

The listener entered the anechoic chamber and sat in a chair in the center of the speaker array at a distance of 1.4 meters from the speakers (see Figure 1(A)). The height of the chair was adjusted so that the listener's ear was at 1.2 meters above floor level, which was equivalent to 0° elevation. On a given trial, a 500-ms white noise was played back over one of the 7 loudspeakers and after the offset of the sound the listener had to indicate its direction. To respond, the listener used a tablet computer (Denver, Android 4.1.1) that displayed a picture with a frontal view of the loudspeaker array (see Figure 3), and had to press on the loudspeaker that corresponded to the perceived sound direction. The listener was instructed to always look straight ahead towards the center loudspeaker after entering a response and wait for the next stimulus. All 7 directions were presented 16 times in random order within a session. All stimuli were reproduced at 68 dB(A) at the ears of the listeners in both natural and hear-through conditions (measured using an artificial head Brüel & Kjær 4158). Prior to the main experiment listeners went through a familiarization

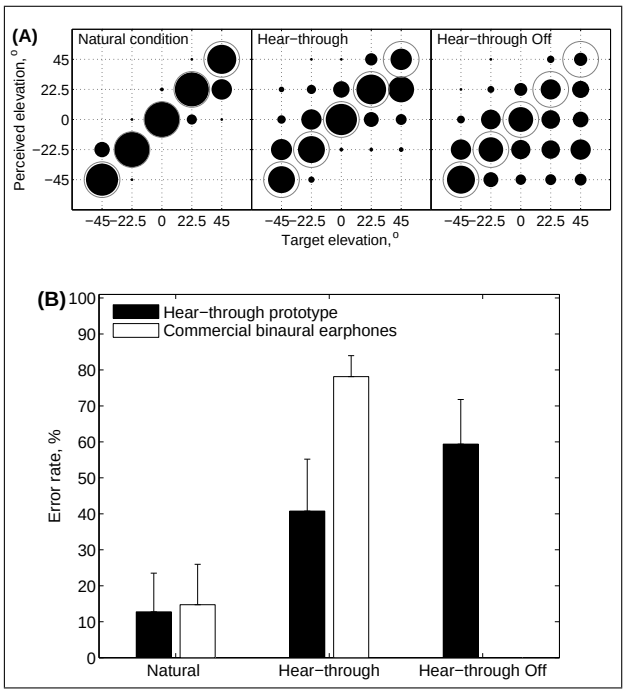

Figure 4. Results from sound localization in the frontal sagittal median plane. (A) Bubble plots indicating perceived elevation vs. target elevation. Perfect performance occurs when all diagonal open circles are completely filled in and there are no off-diagonal responses. The number of responses at a given target elevation is linearly related to the area of the solid circles. (B) Mean error rate across listeners and elevations for the tested hear-through prototype (black bars). These error rates are compared for the natural and hear-through conditions against results from a preliminary experiment using a commercially available binaural headset (white bars). Error bars indicate ±1 standard deviation across listeners.

session in which all directions were presented twice in random order.

3.1. Results

For the two directions at ±45° azimuth, localization performance was perfect for all subjects and for all three conditions. Figure 4 shows the overall performance on vertical sound localization by pooling data across all ten listeners. In Figure 4(A) a perfect performance in a given condition occurred when all the circles along the diagonal were fully filled. Not surprisingly, the closest to perfect performance was achieved in the natural condition (left panel in Figure 4(A)), though at ±45° elevation there was a tendency to compress the perceived vertical space towards the center. The use of hear-through affected performance as seen by a larger spread of the responses relative to the natural condition (center panel in Figure 3(A)). Hear-through performance decreased at all directions as reflected by the average error rates across subjects computed for each direction (see Figure 4(B)). Worst performance was observed when the hear-through device was off as seen by the large spread in the responses (right panel in Figure 4(A)) relative to the natural and hear-through conditions. Importantly, this result suggests that any leakage that may have existed in both hear-through conditions (on and off) did not seem to play a significant role in improving sound localization. Error percentages averaged across participants ranged from 1.3 to 33.8 in the natural listening condition, from 23.1% to 63.8% in the hear-through condition, and from 38.1% to 86.3% in the hear-through off condition. Best sound localization performance was at 0° elevation for the natural and hear-through condition and at −45° elevation for the hear-through off condition. Worst performance was at +45° elevation for the three listening conditions (see Figure 4(B)). A two-way ANOVA with repeated measures on listening condition (natural, hear-through, and hear-through off) and elevation (−45°, −22.5°, 0°, 22.5°, and 45°) revealed a highly significant main effect of listening condition ($F(2,18) = 43.95$, $p < 0.001$), a highly significant main effect of elevation ($F(4,36) = 9.63$, $p < 0.001$), and a significant listening condition × elevation interaction ($F(8,72) = 2.21$, $p = 0.036$). One-way ANOVAs at each level of elevation indicated that the effect of listening condition on sound localization performance was highly significant at 0°, ±22.5° and 45° elevation (all $F(2,18) > 11$, all $p \leq 0.001$). At the −45° elevation the significance of listening test on performance just approached significance ($F(2,18) = 2.87$, $p = 0.083$). Bonferroni-corrected pairwise comparisons between listening conditions showed that hear-through and hear-through off errors were significantly higher than those in the natural listening condition for locations at 0°, ±22.5° and 45° elevation (all $p \leq 0.05$). Error rates in the hear-through condition were significantly lower than in the hear-through off condition for locations at 0° elevation ($p = 0.004$) and +22.5° elevation ($p = 0.029$).

Figure 4(C) shows the mean error rates across directions and subjects (black bars). For comparison the performance of a similar preliminary experiment using a commercially available binaural headset (Roland CS-10EM) is shown. Note that error rates are comparable in the natural condition. Critically, in the hear-through condition error rates are considerably lower for the hear-prototype than for the commercially available binaural headset. Though this difference may stem from differences in the experimental procedures, we believe that it reflects for the most part that the linear distortions to the spatial information introduced by the hear-through prototype were smaller than those introduced by the commercially-available binaural headset [2].

4. Experiment 2 - Speech-on-speech masking in the median sagittal plane

An experimental procedure similar to that reported in [9] was used in the study. The Air Force Research Laboratory's publicly available coordinate response measure (CRM) speech corpus described by [10] comprised the set of stimuli. Each sentence in this corpus has the form "Ready <CALLSIGN> go to <COLOR> <DIGIT> now", where CALLSIGN can be any of a set of eight, COLOR can be any of a set of four, and DIGIT can be any of a set of eight. Considering that spectral cues were of particular interest due to their importance in vertical sound localization, and since the CRM corpus has been low-pass filtered at 8 kHz, the original unfiltered CRM recordings were used in this study.

Two sentences were presented simultaneously on all trials. The target sentence always addressed the callsign "Baron", but the color and digit it referred to and its talker varied randomly from trial to trial. The masker sentence was chosen pseudo-randomly with the constraint that it addressed a callsign other than "Baron". It also referred to a color and digit different from those referred to in the target sentence. The masker sentence was spoken by a talker different from, but of the same sex, as the target talker. The listener's task was to indicate the color and digit spoken by the target talker by pressing a button in a 4×7 button matrix displayed to the listener with a tablet computer (see Figure 5). Note that the digit 7 was excluded because it is bisyllabic whereas all other digits are represented by one-syllable words. Each participant completed one 90-trial session for each listening condition (natural vs. hear-through). Within each session, the 9 possible combinations of (target,masker) elevations (-45,-45), (-45,0), (-45,45), (0,-45), (0,0), (0,45), (45,-45), (45,0), and (45,45) were each presented 10 times in a random order. Note that in this experiment listeners were always uncertain of the location at which the target or masker would be presented from any given trial. Prior to the main experiment all listeners went through a familiarization session in which all (target,masker) combinations were presented twice in random order.

4.1. Results

Figure 6 shows the percent correct speech identification performance averaged across participants for the different combinations of target and masker locations. Mean percentages ranged from 57% to 79% for the natural listening condition and from 59% to 78% for the hear-through condition. Separate two-way ANOVAs at each level of target location with repeated measures on listening condition (natural vs. hear-through) and masker location (45°, 0°, or 45° elevation) showed a significant main effect of masker location for targets ±45° elevation (−45°: $F(2,18) = 6.73$, $p = 0.007$; +45°: $F(2,18) = 4.03$, $p = 0.036$) but not 0° elevation. These results are in agreement with those reported in [9]. Importantly, neither the effect of listening condition was significant nor its interaction with masker location. This suggests that overall hear-through performance was comparable to natural listening performance, and that the effect of masker was also comparable across the two listening conditions.

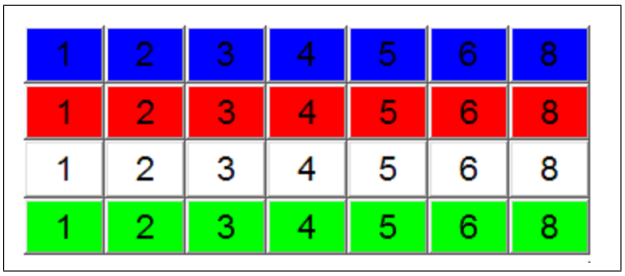

Figure 5. Graphical interface used by listeners in the speech identification task.

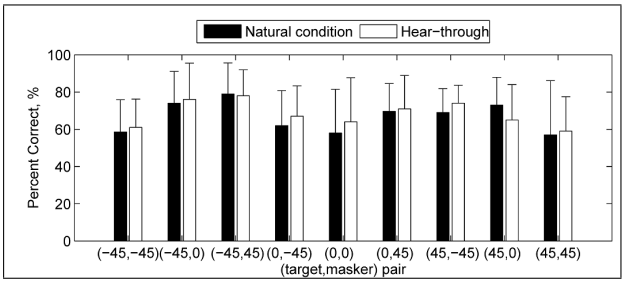

Figure 6. Mean percent correct across listeners (n=10) of identified target color-digit combination in the speech identification task. Percent correct are shown for the natural condition (black bars) and the hear-through conditions (white bars) for all combinations of target-masker directions. Error bars indicate ±1 standard deviation.

5. Discussion and Summary

The main outcome of this work is that relative to spatial auditory perception in natural conditions, hear-through sound localization performance in the frontal median plane deteriorates significantly whereas hear-through speech identification performance remains comparable.

Though the hear-through headset was selected from a set of prototypes because it introduced the least spectral difference between actual and ideal microphone position, and therefore assumably the least change to the spectral cues for localization [2, 11], this was not enough for frontal vertical localization. This is clear considering that hear-through localization at ±45° azimuth, for which interaural cues are known to be dominant, was perfect, whereas localization at ±45° elevation was worst. Further improvements in size, geometry and microphone position may help to

reduce vertical localization errors to levels comparable to those observed during natural listening. Another alternative for improving the hear-through headset performance may be the use of directional microphones which has been shown to enhance sound localization in normal listeners [12].

Now, while localization at 0° and +22.5° elevation was affected by the hear-through earphone, it was still significantly better than having the hear-through off, meaning that the benefit provided by the hear-through amplification was substantial for vertical localization.

In contrast to the sound localization results, the normal auditory processing underlying speech-on-speech spatial release from masking appears to be unaffected by the use of the hear-through earphones. This implies on the one hand that the speech identification test was not sensitive enough to reveal differences between natural and hear-through conditions. On the other hand, the result that performance between natural and hear-through conditions are comparable is encouraging at least for very simple hear-through communication applications. Further work is clearly necessary to examine more realistic scenarios of hear-through multi-talker communication that may include for example variations in sentence composition as well as changes in position and orientation of listeners and speakers.

To finish, we find important to emphasize that when a rigorous assessment of the auditory transparency of a hear-through headset is desired, a sound localization task with sounds in the vertical plane appears to be a suitable choice. Alternatively, a more challenging speech identification test (e.g. by increasing the number of maskers) may also be considered.

Acknowledgement

This study was supported by the project BEAMING funded by the European Commission under the EU FP7 ICT Work Programme. Many thanks to Russell L. Martin for kindly facilitating the original unfiltered CRM recordings. We also want to thank Claus Vestergaard Skipper for his invaluable technical help.

References

[1] A. Härma, J. Jakka, M. Tikander, M. Karjalainen, Tapio Lokki, Jarmo Hipakka, and Gaethan Lorho: Augmented reality audio for mobile and wearable appliances. J. Audio Eng. Soc. **52** (2004) 618639.

[2] P. F. Hoffmann, F. Christensen, D. Hammershøi: Quantitative assessment of spatial sound distortion by the semi-ideal recording point of a hear-through device. International Congress on Acoustics, ICA 2013, Montreal, Canada, June 2013.

[3] D.S. Brungart, A.J. Kordik, C.S. Eades, and B.D. Simpson. The effect of microphone placement on localization accuracy with electronic pass-through earplugs. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, (2003) 149–152.

[4] W. K. Vos, A. W. Bronkhorst, and J. A. Verhave. Electronic pass-through hearing protection and directional hearing restoration integrated in a helmet. J. Acoust. Soc. Am., **123** (2008) 3163.

[5] D. S. Brungart, B. W. Hobbs, J. T. Hamil. A comparsion of acoustic and psychoacoustic measurements of pass-through hearing protection devices. IEEE Workshop in Applications of Signal Processing to Audio and Acoustics, (2007) 7073.

[6] Sharon M. Abel, S. Tsang, and Stephen Boyne. Sound localization with communications headsets: Comparison of passive and active systems. Noise & Health, **9** (2007) 101–107.

[7] E. Cherry: Some experiments on the recognition of speech, with one and two ears. J. Acoust. Soc. Am. **25** (1953) 975–979.

[8] D. Hammershøi, H. Møller: Sound transmission to and within the human ear canal. J. Acoust. Soc. Am. **100** (1996) 408–427.

[9] R. L. Martin, K. I. McAnally, R. S. Bolia, G. Eberle, D. Brungart: Spatial release from speech-on-speech masking in the median sagittal plane. J. Acoust. Soc. Am. **131** (2012) 378–385.

[10] R. S. Bolia, W. T. Nelson, M. A. Ericson, B. D. Simpson: A speech corpus for multitalker communications research. J. Acoust. Soc. Am. **107** (2000) 1065–1066.

[11] P. F. Hoffmann, F. Christensen, D. Hammershøi: Insert earphone calibration for hear-through options. 51st AES International Conference on Loudspeakers and Headphones, Hesinki, Finland, August 2013.

[12] K. Chung, A. C. Neuman, M. Higgins: Effects of in-the-ear microphone directionality on sound direction identification. J. Acoust. Soc. Am., **123** (2008) 2264–2275.