# Comparison of binaural microphones for externalization of sounds

J. Cubick[1], C. Sánchez Rodríguez[1], W. Song[2], E. N. MacDonald[1]

[1] *Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark, Email: jecu@elektro.dtu.dk*

[2] *Brüel & Kjær Sound and Vibration Measurement A/S, Nærum, Denmark*

**Abstract**

Ubiquitous availability of media content through portable devices like media players and smartphones has resulted in an immensely increased popularity of headphones in recent years. However, while conventional stereo recordings usually create a good sense of space when listened to through loudspeakers, the sounds tend to be perceived inside the head (internalized) when headphones are used for listening. A more natural perception in headphone listening with sounds being perceived outside the head (externalized) can be achieved when recordings are made with dummy head microphones or with microphones placed inside the ear canals of a person. In this study, binaural room impulse responses (BRIRs) were measured with several commercially available binaural microphones, both placed inside the listeners' ears (individual BRIR) and on a head and torso simulator (generic BRIR). The degree of externalization of speech and noise stimuli was tested in a listening experiment with a multi-stimulus test. No influence was found for the stimulus signal, but the externalization scores were found to be lower for 0° incidence. With all microphones, relatively high externalization scores were achieved, and for all but one microphone, individual BRIRs resulted in slightly better externalization than generic ones.

## Introduction

In recent years, headphones have gained a lot of popularity, mainly as a side-effect of mobile devices like laptops, media players, and smartphones becoming more and more omnipresent in our daily lives. This development has given new relevance to an old topic. It has long been known that sounds presented via headphones are often perceived inside the head, i.e., internalized rather than outside the head (externalized), like they usually are in everyday listening situations. References to some early studies that describe internalization or inside-the-head locatedness can be found in [1].

A more spacious sound experience with externalized perception of the sound sources is usually desired to create a sense of immersion and reduce listening fatigue that can otherwise occur because of the 180° stereo panorama typically experienced in headphone listening when sounds are perceived internalized. In later years it was found that externalized perception of sounds can be achieved, if the signals at the two eardrums during headphone playback are identical to the signals in the corresponding natural listening situation and, specifically, if the frequency content and temporal relation of the signals at the two ears is correct [2, 3]. One way to achieve this is to use a binaural recording technique, i.e., to record sounds directly at the ears of a listener. It was shown that the full spatial information is preserved if the recording is done at any depth in the ear canal or possibly even some millimeters outside of its entrance plane [4]. Recording at the blocked entrance of the ear canal is also valid. This can result in recordings that sound very realistic, especially for the same listener. Similarly, a recording technique can be applied, where the listener is replaced with a mannequin head (and sometimes torso) that is equipped with microphones inside the ears, often referred to as a dummy head microphone or a head and torso simulator (HATS).

If a human head and torso is inserted into a sound field, reflections at the head and in the cavities of the outer ear and diffraction of sound waves around the head will generate a filter that attenuates and amplifies certain frequencies. The coloration of the sound that finally arrives at the eardrum is highly dependent on the direction of the incident sound. Apart from recording directly at the ears of a listener or dummy head, the spatial information can therefore also be described by the head related impulse response (HRIR) in an anechoic sound field or by a binaural room impulse response (BRIR), which also includes the acoustic properties of the room [1, 5, 6], when constant direction of incidence is assumed. HRIRs or BRIRs measured on a dummy head are commonly referred to as generic. Convolution of anechoic sound signals with HRIRs or BRIRs generates a playback signal that often results in a surprisingly realistic acoustic impression of an acoustical scene. Today a number of microphones for binaural recordings are available on the market, ranging from accessories for portable recorders for recording of e.g. rock concerts or soundscapes to tools for sound quality evaluation and scientific work.

Most studies that evaluated the result of binaural recording techniques focussed on localization (e.g., [7])) and they typically reported worse performance when stimuli were generated with non-individual BRIRs. For distance perception, Zahorik reported that no difference could be found between conditions with individual BRIRs and non-individual BRIRs measured on another listener's head [8], and Werner and Siegel found no influence of using individual or generic BRIRs on externalization [9]. Begault and Wenzel on the other hand found very high percentages of stimuli being perceived internalized for anechoic speech stimuli and non-individual HRIRs of a human head [10].

This study investigated the degree of externalization that could be achieved with five different commercially available binaural microphones and a dummy head using a virtual auditory space technique. In a listening experiment, eight normal-hearing listeners rated the perceived externalization of sounds presented via headphones for all microphones for four different source positions and two different types of stimuli in a multi-stimulus test paradigm.

There were four main research questions: 1) Does the stimulus material influence the perceived externalization? 2) Does the externalization percept depend on the incidence angle? And most importantly 3) Do the different microphones yield different externalization ratings? and 4) Does it make a difference whether individual or generic BRIRs are used?

# Methods

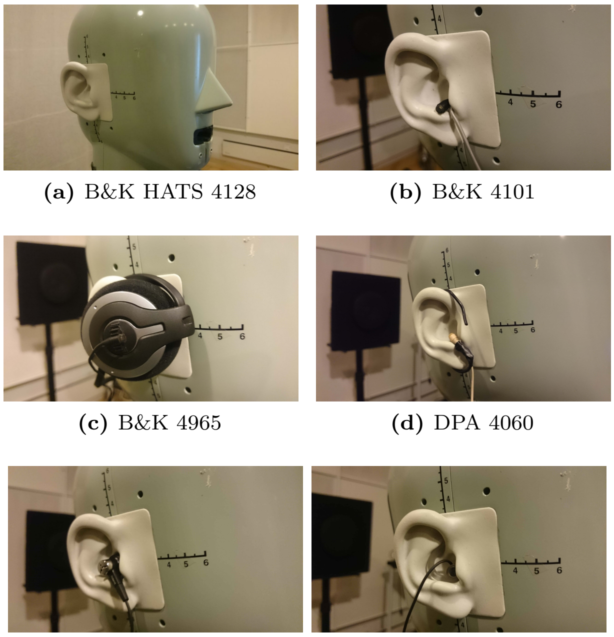
## Microphones

Five different pairs of commercially available microphones were chosen for the comparison. In addition, the internal microphones of the HATS have been used as a representative for dummy head recording techniques. An overview of the microphones and their background noise levels can be found in Table 1. All noise levels except for the HATS internal microphones was measured in an anechoic chamber at DTU, the values for the HATS were taken from the data sheet.

| Microphone | Alias | Noise level (L/R) |
|---|---|---|
| B&K HATS 4128-C-002 | HATS | 19.0/19.0 dB(A) 21.3/21.3 dB SPL |
| B&K 4101-A | 4101 | 22.4/22.6 dB(A) 28.4/28.4 dB SPL |
| B&K 4965 | 4965 | 23.3/23.1 dB(A) 29.7/29.3 dB SPL |
| DPA 4060 | DPA | 22.5/22.7 dB(A) 35.3/36.4 dB SPL |
| Roland CS-10EM | Roland | 27.4/27.3 dB(A) 30.4/30.2 dB SPL |
| Sound Professionals MS-TFB-2 | SProf | 25.0/25.3 dB(A) 31.9/31.9 dB SPL |

**Table 1:** Type, alias, and background noise level of the microphones used in this study. Note, that all given noise levels were measured except for the one of the HATS, which was taken from the data sheet.

Figure 1 shows photographs of the binaural microphones under test mounted on a HATS. All microphones were used with the mounting solution provided by the manufacturer except for the DPA 4060 (Fig. 1d), which are originally clip microphones made for stage use. These microphones were positioned on the listeners' ears by means of a wire hook that was individually adjusted to place the microphone as close as possible to the entrance of the ear canal. Note that due to the differences in construction, the position with respect to the ear canal was quite different for the respective microphones.



**(a)** B&K HATS 4128 **(b)** B&K 4101
**(c)** B&K 4965 **(d)** DPA 4060
**(e)** Roland CS-10EM **(f)** Sound Prof. MS-TFB-2

**Figure 1:** Binaural microphones used in this study mounted on a B&K HATS. Note the different positions of the microphones on the pinnae of the HATS. Especially the B&K 4965 microphones (c), but also the B&K 4101 (b) and the Roland microphones (e) are placed at a position clearly outside of the ear canal, which is less then optimal, because the transfer function from the microphone to the ear drum is not independent of direction [4].

## Listeners

The listening experiments were performed by eight normal-hearing listeners (aged 21-25, 2 female) with listening thresholds better than or equal to 20 dB HL on both ears at all of the audiometric frequencies from 125 Hz to 8 kHz. Five listeners were naïve, three listeners had participated in listening experiments before.

## BRIR measurements

Individual BRIRs were measured for each listener for all five microphone pairs in an IEC listening room [11] with an average reverberation time $T_{30}$ of about 0.3 s and a volume of about 100 $m^3$. For each set of microphones, BRIRs were measured for four loudspeakers Dynaudio BM6P at azimuth angles of 0, 25, 60, and 90° and a distance of 2.5 m using 6 repetitions of a 5 s logarithmic sine sweep and a deconvolution method according to [12]. Furthermore, generic BRIRs were measured under the same conditions on the HATS for all five microphones and the internal microphones of the HATS.

After the measurement of the BRIRs from the loudspeakers, a pair of Sennheiser HD 800 headphones was carefully placed on the head without moving the microphones and headphone impulse responses (HPIR) were measured to the respective microphones with 10 repetitions of a 2 s logarithmic sine sweep. The inverse filters for the headphone equalization were derived from the measured impulse responses using a least means squares time domain inversion method. The listeners were instructed to keep the position of head as fixed as possible.

## Stimuli

In the experiments, two different signals were used, sentences from the Danish HINT speech test corpus [13], and trains of pink noise bursts (5 bursts of 200 ms with a pause of 300 ms in between, 5 ms Hanning ramps at the beginning and end of each burst). For each experimental run, 10 stimuli were generated by convolving the signal with the individual and the generic BRIRs for the five microphones. As a control condition, the signal was also convolved with the BRIR measured with the internal microphones of the HATS. The dry signal served both as a reference (played back via loudspeaker) and an anchor (played back diotically via headphones). To avoid loudness as a cue, the reference was adjusted to subjectively match the loudness of the other signals by two of the authors. In total, 13 stimuli were used within each experimental run. The 11 signals involving BRIRs were additionally filtered with the inverse filters derived from the measured HPIRs. All auralized signals were band-limited between 50 Hz and 15 kHz with $6^{th}$ order Butterworth filters.

## Experimental procedure

**Figure 2:** Photograph of the experimental setup with a listener at the listening position inside the IEC listening room. The four loudspeakers were positioned at 0, 25, 60, and 90° at a distance of 2.5 m. The listeners controlled the experiment via a graphical user interface on a small screen using a wireless mouse.

During the experiment, the listeners were seated in the same room at the same position, where the BRIRs had been measured (see Figure 2 for a photograph of the setup.). They controlled the listening experiment via a graphical user interface in Matlab (cf. Figure 3). The procedure was a modified MUSHRA test [14]. Each of the stimuli described above was randomly assigned to one of the 13 buttons (A-M), which start the audio playback. The externalization rating for each stimulus was reported via the corresponding slider. Within each experimental run, the signal and the loudspeaker angle were kept constant. When speech stimuli were used, the same sentence was used for all stimuli within one experimental run. The angles of the loudspeakers were randomized over the experimental runs

The listeners were instructed to judge the degree of externalization on a scale from 0 to 100, where 0 means that the sound was perceived inside the head and 100 that the sound was perceived at the position of the loudspeaker. They were instructed to rate the hidden reference as 100 (if found). To help the judgement, a five-point scale similar to [15, 16] was supplied ranging from “Inside my head” (0), “Near my head” (25), “Close to me” (50), “Close to the loudspeaker” (75), and “At the loudspeaker” (100). The listeners could listen to the stimuli as often as needed in order to make a judgement. Once they rated all stimuli, hitting “Continue” started the next run.

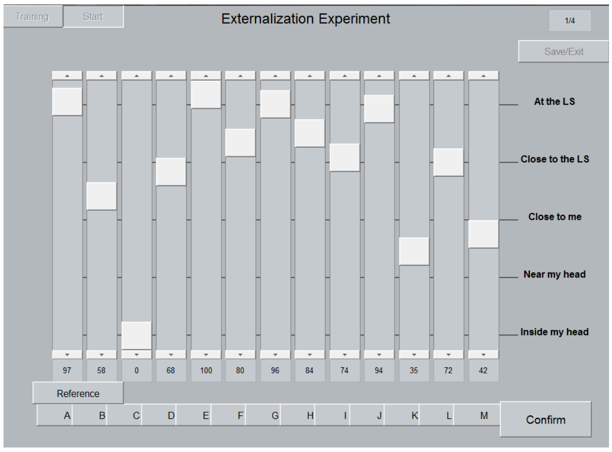



**Figure 3:** Graphical user interface for the listening experiment. The 13 buttons (A-M) allow for playing back the stimuli (5 individual and 5 generic BRIRs for the microphones under test, the internal microphones of the HATS, the hidden reference, and the anchor in random order). The externalization rating is entered via the corresponding slider.

Before the experiment, the listeners performed two training runs with stimuli presented from 0° and 60°. The actual experiment consisted of eight runs (4 angles, 2 stimuli). The whole experimental session took about 40 minutes per listener.

## Statistics

To test the results, a repeated measures Analysis of Variance (rANOVA) was carried out with “Angle”, “Microphone”, and “Stimulus” as within-subject factors. Post-hoc pairwise t-tests were carried out for all factors that showed a significant effect in the rANOVA.

# Results

## Influence of the stimulus signal

Fig. 4 shows the externalization rating averaged over all listeners, microphones and loudspeaker positions for the noise bursts (left) and the speech stimuli (right). To increase readability, the plot only shows the upper half of the response scale. The ratings for the reference and the anchor were excluded. The average rating for the noise signal was 66.3, the rating for the speech signal was 67.2 or slightly below “Close to the loudspeaker”. The choice of the stimulus signal thus did not seem to have an influence on the perceived externalization, which was confirmed in the rANOVA, where the main factor “Stimulus” showed no significant effect [$F(1, 7) = 0.069$, $p = 0.8$].

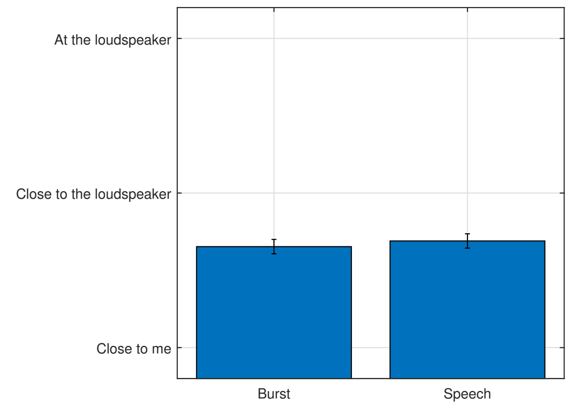



**Figure 4:** Average externalization rating for noise (left) and speech (right) stimuli. Error bars indicate ± one standard error. Note that only the upper half of the scale is shown.

## Influence of the loudspeaker angle

Fig. 5 shows the average externalization rating of all listeners for all microphones over the four loudspeaker angles 0, 25, 60, and 90°. The ratings increase with angle from 61.6 at 0° over 65.8 at 25°, 69.7 at 60° to 70 at 90°. The rANOVA showed a significant effect of the factor "Angle" on the externalization rating [$F(3, 21) = 3.228$, $p = 0.043$], the post-hoc analysis revealed that the only significant differences are found between the rating for 0° and the ratings for 60° and 90°. This was expected, because front-back confusions and internalization were reported to be most common for directions close to the median plane (e.g., [10]), where the differences between the ear signals are small. A recent study, however, did not find a significant difference on externalization when presenting virtual stimuli from 0, 90, or 180° [19].

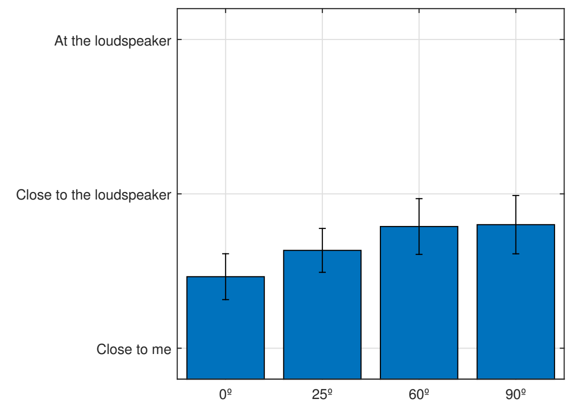



**Figure 5:** Average distance rating for the externalization for the four different loudspeaker angles. The error bars indicate ± one standard error. Again, only the upper half of the scale is shown.

## Influence of the microphone type

Among the generic BRIRs, the highest externalization scores were obtained with the HATS internal microphones with an average value of 75.5, closely followed by the DPA microphones (74.5). Generic BRIRs measured with all other microphones resulted in significantly lower average externalization ratings, as confirmed by the post-hoc analysis.

All stimuli that were generated using individually recorded BRIRs were on average judged as fairly well externalized (with ratings of 68.5 for the B&K 4101, 66.8 for the B&K 4965, 69.1 for the DPA 4060, 69.1 for the Roland, and 69.9 for the Sound Professionals). The ratings were thus just below the "Close to the LS" category. The post-hoc analysis showed that none of the microphones yielded significantly different externalization scores when the BRIRs were measured individually. For the individual BRIRs, there is therefore no statistical evidence that one of the microphones yields better results than the others. For a full overview over the results of the post-hoc analysis, see Table 2.

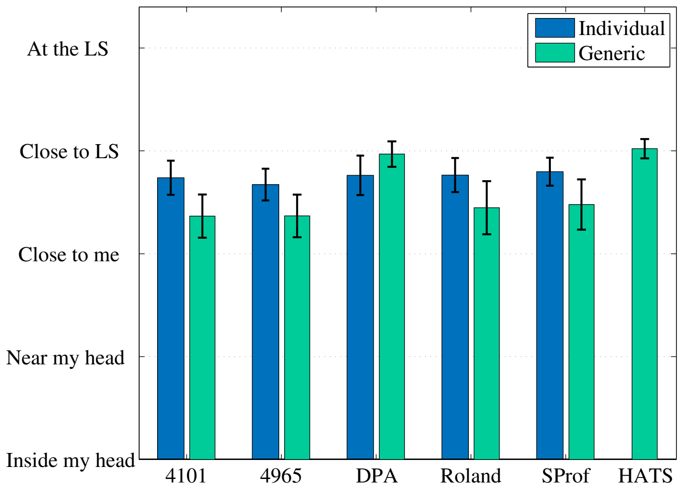



**Figure 6:** Average externalization rating for five different pairs of binaural microphones, each for individual and generic BRIRs.

## Individual vs. generic BRIRs

Only for the 4101 and the 4965 a significant difference in the externalization ratings was found between the individual and the generic BRIR for the same microphone. In both cases, the individual BRIR yielded higher ratings.

Another rANOVA was carried out to further analyze the effect of the individual versus generic BRIRs. The HATS was excluded from the calculation and the within-subject factor "Individual/Generic" was added. The results showed again that the angle has a significant effect, whereas the stimulus signal does not. The main effects of "Microphone" and "Individual/Generic" were not significant but they did show a trend. Furthermore, the interaction between "Microphone" and "Individual/Generic" was found to be significant [$F(4,28) = 3.305$, $p = 0.024$]. A look at the data reveals, that this interaction occurred because the DPA 4060 microphones yield higher externalization ratings for the generic than for the individual BRIRs, whereas for all other microphones the individual BRIR yielded higher externalization ratings (ca. 8% on average). When the DPA microphones were excluded from the statistical analysis, the main effect of "Individual/Generic" was significant [$F(1,7) = 8.151$, $p =$

| | | individual BRIR | | | | | generic BRIR | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mic | 4101 | 4965 | DPA | Roland | SProf | 4101 | 4965 | DPA | Roland | SProf | HATS |
| individual | 4101 | | | | | | ✓ | | ✓ | | | ✓ |
| | 4965 | | | | | | ✓ | ✓ | | | | ✓ |
| | DPA | | | | | | | | | | | |
| | Roland | | | | | | ✓ | | | | | |
| | SProf | | | | | | ✓ | ✓ | ✓ | | | ✓ |
| generic | 4101 | ✓ | ✓ | | ✓ | ✓ | | | ✓ | | | ✓ |
| | 4965 | | ✓ | | | ✓ | | | ✓ | | | ✓ |
| | DPA | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| | Roland | | | | | | | | ✓ | | | ✓ |
| | SProf | | | | | | | | ✓ | | | ✓ |
| | HATS | ✓ | ✓ | | | ✓ | ✓ | ✓ | | ✓ | ✓ | |

**Table 2:** Results of the post-hoc analysis. The checkmarks indicate pairs for which a significant difference was found ($\alpha = 0.05$)

0.025]. This could be explained by the fact that it was quite easy to accidentally move the DPA microphones during the measurement of the individual BRIRs due to the way they were attached to the ear, which could lead to less precise BRIR measurements and potentially incorrect equalization filters. It could be suspected that with a more optimal and stable placement of the microphones on the ears, the individual BRIRs would lead to higher externalization scores for these microphones as well.

## Discussion

The externalization ratings found in this study were generally quite high with most of the ratings occurring somewhere between "Close to me" and "Close to the loudspeaker" (grand average: 66.8), indicating that the auralization technique used here works well. This corresponds well with the subjective impression, where most sources were clearly externalized and it was difficult to make out a clear difference between the stimuli. It seemed a bit surprising that no bigger difference was found between generic and individual BRIRs, even though the individual BRIRs yielded higher average externalization ratings for all but one microphone. What might have helped in the current study, was the fact that the experiments were performed in the same room as the BRIR measurements. Some recent work has pointed out that the auditory image is usually perceived most externalized when the playback room and the recording room are identical [17, 18].

Note that the basic assumption of binaural technology has been violated in some of the measurements. The basic assumption is that the transfer path of the sound from a sound source to the eardrum can be divided into a directional-dependent and a directional-independent part and that the perception of an acoustic scene simulated via binaural technology will correspond to the one in the real scene, if the sound pressure is reproduced correctly at the eardrum or at a point at the ear, where the frequency response is independent of direction [4]. This is the case inside the ear canal, but not (far) outside it. Equalizing the headphones relative to a microphone position outside the ear canal therefore very likely introduces sound coloration, a disturbed localization, and might also cause a reduced externalization percept. Especially for the B&K 4965 Microphones, but also for the B&K 4101 and the Roland microphones, this was expected to be problematic, because the microphones are positioned rather far outside the ear canal. Interestingly, this "wrong" equalization did not seem to have a big impact on the externalization rating, since the ratings were not significantly different from the ones for the other microphones. It might, however, be one of the reasons why both the 4101 and the 4965 scored lower average externalization ratings than the DPA 4060, even though all three microphones are based on the same microphone capsules. One area where the "wrong" placement of the microphones might have an influence are attributes of sound quality. Especially in some conditions with noise stimuli, timbral differences between the microphones were quite obvious. In future investigations, this and other perceptual attributes like compactness or localization of the auditory image should be considered, because that would allow for a more complete understanding and clearer ranking of the binaural microphones.

Looking at the results it should also be considered that the microphones under test will most likely be used in very different ways in practice. Someone, who invested in a very expensive HATS, will most likely be aware of the necessity to equalize the headphones, the amateur who occasionally records a concert of a local rock band on a cheap portable recorder will most likely not and just listen to the recording as it is through whatever headphones available. If these different approaches had been considered in the listening experiments, some larger differences might have been found in the externalization ratings between the microphones.

Considering that most of the stimuli were perceived well externalized, using an omnidirectional room impulse response for the anchor signal might have resulted in a wider range of judgements, whereas the anechoic signal used in this study, being very different from the other stimuli, might have limited the range of responses that has been used by the listeners.

## Conclusion

Five commercially available types of binaural microphones have been evaluated with respect to the achieved amount of externalization. In a listening experiment with eight listeners, the average externalization scores were relatively high (just below "Close to the LS"). With the exception of the DPA 4060, individual BRIRs resulted in higher ratings than generic BRIRs. However, the differences were surprisingly small. This indicates that, if only externalization is considered, BRIRs measured on dummy heads might well be sufficient in many situations to generate a more natural sound experience with sources perceived well outside the head. This argument is supported by the fact that the stimuli that used BRIRs measured using the internal microphones of the HATS consistently yielded the highest average externalization scores. Using either speech or pulsed noise stimuli did not change the overall judgement. As found by others before, good externalization seems most difficult to achieve for frontal directions, which is reflected in the lower externalization scores measured for 0° incidence.

Externalization scores are only one aspect in judging the performance of binaural microphones. As a next step, other outcome measures should be considered as well. It seems especially crucial that the microphones do not introduce coloration, that they allow for natural localization, and that the auditory image is compact.

## References

[1] Blauert, J.: Spatial hearing: the psychophysics of human sound localization. The MIT Press. 1997

[2] Laws, P.: Entfernungshören und das Problem der Im-Kopf-Lokalisiertheit von Hörereignissen. Acustica **29**(5) (1973), 243–259

[3] Wightman, F.L., and Kistler, D.J.: Headphone simulation of free-field listening. I: Stimulus synthesis. The Journal of the Acoustical Society of America **85**(2) (1989), 858–867

[4] 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

[5] Møller, H.: Fundamentals of binaural technology. Applied acoustics **36**(3) (1992), 171–218

[6] Vorländer, M., and Summers, J.E.: Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms, and Acoustic Virtual Reality. Springer-Verlag, Berlin. 2008

[7] Møller, H., Hammershøi, D., Jensen, C.B., and Sørensen, M.F.: Evaluation of artificial heads in listening tests. Journal of the Audio Engineering Society, **47**(3) (1999), 83–100.

[8] Zahorik, P: Auditory display of sound source distance. In Proceedings of the International Conference on Auditory Display. 2002

[9] Werner, S. and Siegel, A.: Effects of binaural auralization via headphones on the perception of acoustic scenes. in Proc. of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. 2012

[10] Begault, D.R., and Wenzel, E.M.: Headphone localization of speech. Human Factors: The Journal of the Human Factors and Ergonomics Society, **35**(2), 361–376 (1993)

[11] IEC 268-13. Sound system equipment - part 13: Listening tests on loudspeakers. *(International Electrotechnical Commission, Geneva, Switzerland)*, 1985

[12] Müller, S. and Massarani, P.: Transfer-Function Measurement with Sweeps. Journal of the Audio Engineering Society **49**(6) (2001), 443–471

[13] Nielsen, J.B. and Dau, T.: The Danish hearing in noise test. International Journal of Audiology **50**(3) (2011), 202–208

[14] ITU-R BS.1534-2 Method for the subjective assessment of intermediate quality level of audio systems. *(International Telecommunication Union – Radiocommunication Sector)*, 2014

[15] Catic, J., Santurette, S. Buchholz, J. M., Gran, F., and Dau, T.: The effect of interaural-level-difference fluctuations on the externalization of sound. The Journal of the Acoustical Society of America **134**(2) (2013), 1232–1241

[16] Boyd, A.W., Whitmer, W.M., Soraghan, J.J., and Akeroyd, M.A.: Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. The Journal of the Acoustical Society of America **131**(3) (2012), EL268–EL274

[17] Udesen, J. and Piechowiak, T. and Gran, F.: Vision Affects Sound Externalization. In: Audio Engineering Society Conference: 55th International Conference: Spatial Audio, 2014

[18] Gil Carvajal, J.C.: The influence of visual cues on sound externalization. Master's thesis at Hearing Systems group, Department of Electrical Engineering, Technical University of Denmark. 2015

[19] Udesen, J., Piechowiak, T., and Gran, F.: The Effect of Vision on Psychoacoustic Testing with Headphone-Based Virtual Sound. Journal of the Audio Engineering Society **63**(7/8) (2015), 552–561

[20] Begault, D.R., Wenzel, E.M., and Anderson, M.R.: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. Journal of the Audio Engineering Society **49**(10) (2001), 904–916