Audio Engineering Society

Conference Paper 24

Presented at the AES International Conference on Headphone Technology
2025 August 27–29, Espoo, Finland

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Comparison of Passive Transparent Headphones for Augmented Reality Audio

Alexander Mülleder[1], Pedro Lladó[2], and Nils Meyer-Kahlen[3]

[1]*University of Music and Performing Arts Graz, Graz, Austria*
[2]*Institute of Sound Recording (IoSR), Department of Music and Media, University of Surrey, Guildford, United Kingdom*
[3]*Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Espoo, Finland*

Correspondence should be addressed to Alexander Mülleder (alex.muelleder@gmx.at)

ABSTRACT

In Augmented Reality (AR) applications, transparent headphones allow users to perceive their surrounding environment while integrating virtual sound sources. This requires two independent qualities of the headphones: high acoustic transparency and suitability for high-quality binaural rendering. In this study, we assess these two properties for eight headphones commonly used in AR research. To this end, head-related transfer functions and headphone transfer functions were measured for ten participants and a head-and-torso simulator using microphones at the blocked ear canal. The measurements were repeated fifteen times after repositioning each pair of headphones to capture not only inter- but also intra-individual variability of both measurements. These acoustic measurements were then used to estimate headphone-induced coloration and localization errors by means of auditory models. Our results show the variability of transparency estimates for the first time and highlight the importance of low headphone transfer function variability for the suitability of headphones for equalized binaural reproduction. All data are made publicly available to support reproducible AR audio research.

1 Introduction

In Augmented Reality (AR) applications, the headphones required to render virtual sound sources inevitably obstruct the path of external sound to the listener's ears. Therefore, the suitability of headphones for sound in AR can be considered in two almost independent aspects, which depend on their specific hardware design. The first one is preserving the sound from the surrounding real environment with as little impairment as possible, i.e., the transparency. The second one is achieving a well-controlled reproduction. High transparency can be achieved through either active hear-through or by using passive devices that are specifically designed for this purpose [1, 2, 3]. Some of these designs, with drivers hovering a few centimeters away from the ears, are referred to as *extra-aural*. Interestingly, solutions based on active hear-through have been the prime topic of earlier research in AR audio [4, 5], but researchers studying requirements of sound for AR

in recent years have more frequently used passive transparent headphones, see e.g., [6, 7, 8, 9, 10, 11] for the most recent studies from various different labs, most likely because they are easier to characterize. This study compares passive transparent headphones that are most commonly used in recent research papers.

Several existing studies have assessed the transparency of such headphones already. Transparency measurements that include such passive transparent headphones were presented, for example, in [12, 13, 14, 15, 16, 17]. Most of these papers also include listening experiments evaluating the effects of reduced transparency.

In [15], both measurements and listening test results of passive transparent headphones were compared to assess the ability of auditory models to predict perceptual transparency scores regarding coloration and static vertical localization. They were shown to correlate well. Moreover, such auditory models have been shown to estimate the impairment caused by headphones in a more complex auditory-aided visual search task, particularly relevant in AR settings [16, 18].

There are, however, three important aspects that previous studies did not formally address. First, none of the studies considered the variability of the transparency measurements and the model-based evaluation results derived from them. The results from [19] showed a first indication that model predictions of vertical localization impairments varied considerably between five measurements conducted on a head and torso simulator (HATS). Second, existing studies on transparent headphones have often focused solely on transparency and do not address the headphone transfer functions (HpTFs) and their variability across the same devices. We expect this variability to be high for some of the more unusual, extra-aural designs. Third, some of the headphones most commonly used in research have not yet been compared directly using the same measurement procedures and analysis methods.

Here we compare eight headphones that are commonly used in research regarding both predicted perceived transparency and potential for high-quality binaural playback. They were measured on ten human participants and on an HATS using microphones at the blocked ear canal, including 15 repositionings of each pair of headphones. This allowed variability between participants (*inter*-individual) and within participants (*intra*-individual) to be assessed.

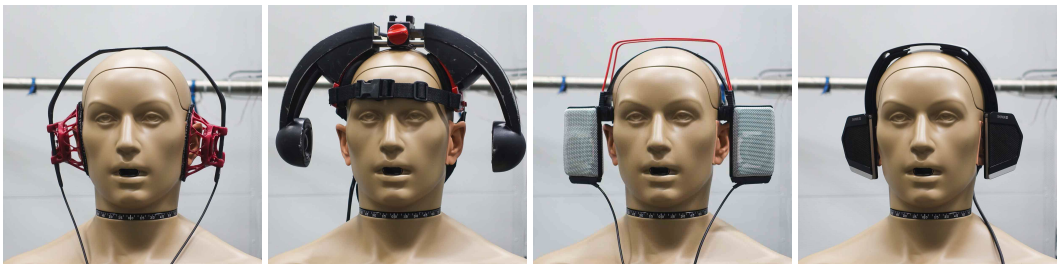
For the evaluation regarding perceptual transparency, the models presented in [15] are employed, wherein one model is used for predicting perceived coloration, and one for vertical localization impairments caused by wearing the headphones. Due to the repeated measurements with several subjects, the variability of the perceptual transparency model outputs can be assessed for the first time. Moreover, the same headphones are compared regarding reproduction quality. We argue that for research use, the reproduction quality mainly depends on inter- and intra-individual variability of the HpTF, which limits the efficacy of headphone equalization methods. Previous studies, e.g., [12, 20, 21], have investigated inter- and intra-individual variability in HpTF measurements by showing measures of dispersion like standard deviation or percentiles. We extend this by analyzing the effect of variability on the predicted coloration following equalization.

In summary, we divide our research questions in two blocks. First, in terms of transparency, our main research questions are: (1) How do the selected headphones differ in perceptual transparency? (2) What is the intra-personal and inter-personal variability of their perceptual transparency? Second, regarding reproduction, the main research question is: How large are the effects of intra-personal and inter-personal variation in HpTF on equalized reproduction? The following section describes the headphones tested and details the measurement procedure. Then, Sections 3 and 4 explain the models and present the results regarding transparency and reproduction quality, respectively. Section 5 concludes the results.

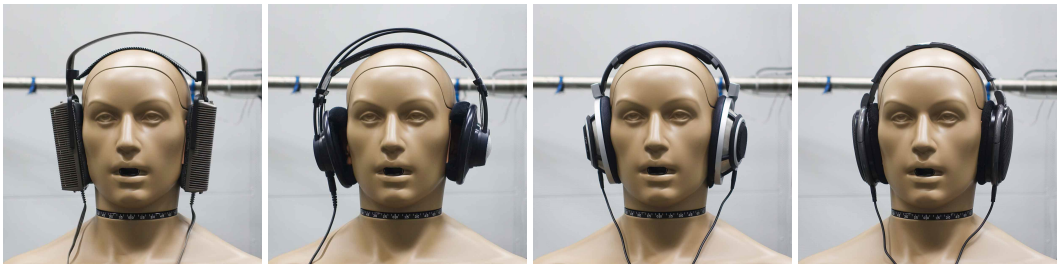
2 Headphones and Measurement Procedure

The different headphone models are shown in Fig. 1. The selection comprises most choices of passive transparent headphones that have been used in research in recent years. The tested models are:

1. MushRoom [3]. Designed for AR and used for recent AR experiments [6, 10] and several experiments containing real and virtual sound sources [22, 23].
2. BK211. A custom, extra-aural headphone designed for binaural experiments [1]. Used, e.g., in [24] and evaluated regarding transparency in [14].



(a) MushRoom **(b)** BK 211 **(c)** K1000 (opened) **(d)** Mysphere 3.2 (opened)



(e) STAX SR-303 **(f)** AKG K702 (modified) **(g)** Sennheiser HD 800 **(h)** Sennheiser HD 650

Fig. 1: All tested models on the KEMAR HATS.

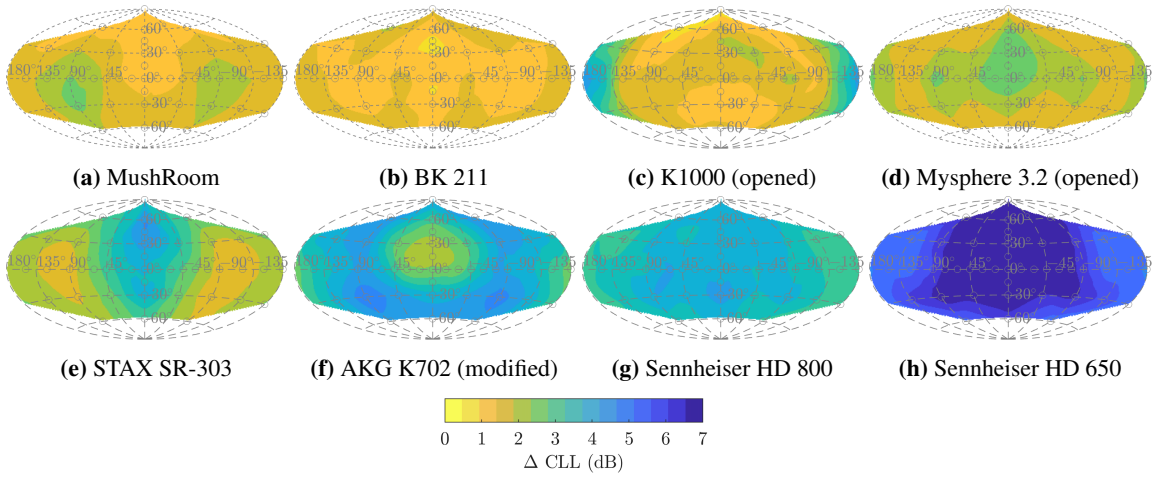
3. AKG K1000. A pair of extra-aural headphones originally designed for high-fidelity listening. They have been used in numerous AR-related studies, most recently [8, 11, 9, 7]. They were compared to the BK211 and other models in [14]. The measurements were performed with the movable frames in open position.
4. Myshphere 3.2. Headphones with an extra-aural design used in listening experiments regarding VR and AR, like [25, 26]. Measured in [15]. Also its frames were in the open position.
5. STAX SR-303. An open headphone with electrostatic drivers. A STAX SR-202 was evaluated regarding localization impairments in [12].
6. Modified AKG 702, as introduced in [2] and used for real time virtual acoustics for music practicing [27], and AR experiments [28]. Previously measured in [15].
7. Sennheiser HD 800. High-quality open-back headphones for comparison.
8. Sennheiser HD 650. Open-back headphones that are very commonly used for psychoacoustic experiments.

All measurements were conducted in the anechoic chamber "Wilska" at the Aalto Acoustics Lab, Finland. The chamber contains an array of 45 surrounding Genelec 8331A loudspeakers. Their directions are shown as circles in Fig. 2.

Ten adults (three women and seven men, aged between 23 and 31) volunteered to participate in the measurement campaign. The data of the first participant had to be excluded because the headphone and microphone gains had not been adjusted properly yet. HRTFs without headphones, and HRTFs and HpTFs with the eight headphones were measured using the "blocked-ear-canal" method with the PIRATE in-ear microphones [29]. An RME MadiFace UCX II in combination with the RME ADI-6432 was used for measurements. In-ear microphones and the headphones were directly plugged into the MadiFace unit. The exact same setup using the PIRATE microphones was used for the KEMAR HATS[1].

Both HpTF and HRTF measurements were repeated 15 times. Before the first repetition, the participants were fitted with the headphones by the experimenter. They were then instructed to take them off and put them back on after each measurement. They were asked not to put on the headphones in a different way intentionally,

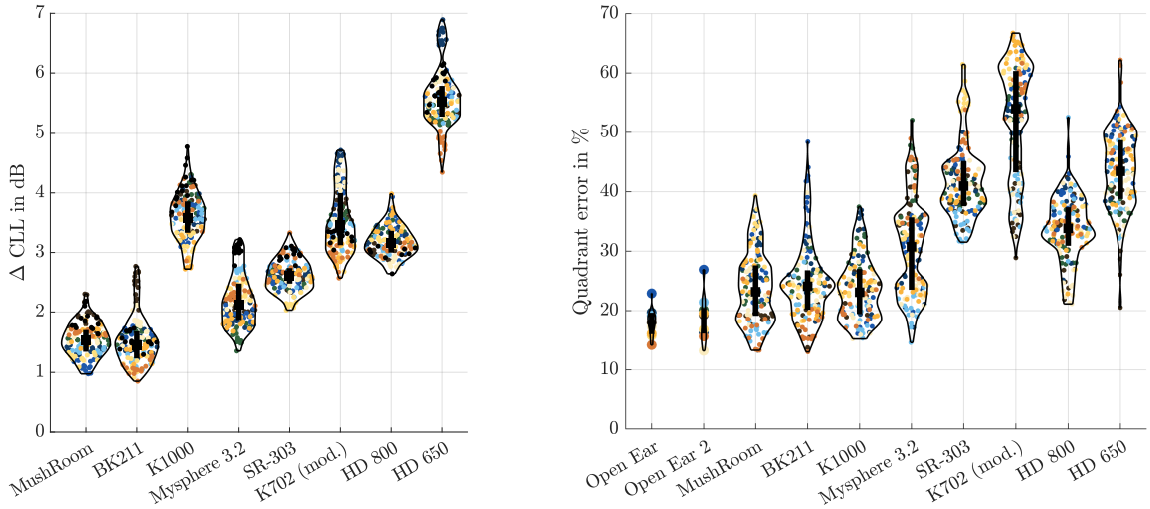
[1] Data for the internal ear simulator of the HATS, as well as high-resolution HRTF data of open ears were also recorded, but are not analyzed here.



(a) MushRoom **(b)** BK 211 **(c)** K1000 (opened) **(d)** Mysphere 3.2 (opened)

(e) STAX SR-303 **(f)** AKG K702 (modified) **(g)** Sennheiser HD 800 **(h)** Sennheiser HD 650

Fig. 2: Model-based predictions of the perceived coloration of real-world sounds introduced by insufficient transparency when wearing the headphones. The CLL was computed in relation to the first open ear HRTF measurement and averaged over frequency bands.



(a) Predicted coloration. ΔCLL averaged over frequency bands and directions.

(b) Predicted quadrant error, estimated using the `Baumgartner2014` model.

Fig. 3: Violin plots of the predicted perceptual transparency. Black squares indicate median values and bars show interquartile ranges. Different colors belong to different subjects. Black dots belong to HATS data.

but were not instructed further during these replacements. Upon each repetition, the HpTF was measured first, followed by the HRTF measurements using the 45 loudspeakers. Interleaved exponential sine sweeps with a duration of 2 s and an interval of 0.1 s between the start of sweeps were used. HRTFs without wearing headphones were measured once before and once after all other measurements to check if the microphones were still fitted properly. All measurement data, along with plots of raw HRTFs and HpTFs, are available online[2].

3 Transparency

This section describes the models and the results of the evaluation regarding acoustic transparency. Reproduction quality is discussed in the next section.

[2] `https://doi.org/10.5281/zenodo.16055513`

3.1 Method

The perceptual transparency can be determined from the HRTF measurements of each participant with and without wearing the headphones. Instead of simply taking the ratio of the two, the models presented in [15] are employed. The first one predicts the influence of the device on coloration by means of the composite loudness level (CLL) [30]. The second one predicts vertical localization impairments caused by wearing the headphones by employing the `Baumgartner2014` model [31], as implemented in the Auditory Modeling Toolbox [32]. We focus on the predicted quadrant error (QE). For both models, the reference was the first open ear HRTF measurement. ΔCLL was computed as the difference of the CLL values for the HRTF measurements while wearing each headphone and the CLL values of the reference, as in [15]. We analyze averages over frequency bands, knowing that more differences occur in the high-frequency range. For QE, the reference was used as "template" and each headphone measurement as "target" for the `Baumgartner2014` model. The model parameters were fixed, as in [15]. They were $S = 0.35$, $\varepsilon = 27$, and $\Gamma = 17$.

3.2 Results and Discussion

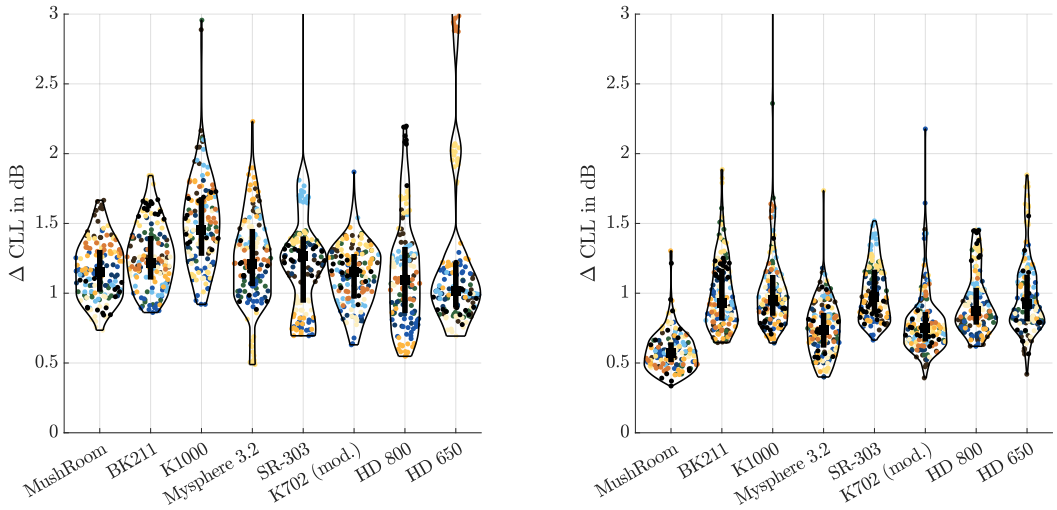
The results of the perceptual transparency evaluation are shown in Fig. 2 and 3. In the former, the direction dependent coloration is shown as the mean over participants. As seen in previous studies, very open headphones like the MushRoom show lower coloration impairments than more conventional open headphones like the Sennheiser HD 650 or HD 800. When comparing to HATS measurements done earlier [19], the mean ΔCLLs based on the measurements using human subjects confirm the high transparency of the MushRoom, even when compared to K1000 and Mysphere 3.2. In direct comparison, it becomes clear that the BK211 induces even lower coloration for the lateral directions. When comparing the very commonly used K1000 and the Mysphere 3.2, which are designed based on similar principles, the plots show that the K1000 tends to color the sound from behind more than from the front, whereas the opposite is true for the Mysphere.

Figure 3a allows for checking the variability of the predicted coloration, averaged over all tested directions. The MushRoom and the BK211 result in the lowest values. Also, the interquartile range (IQR) is below 0.5 dB for them and for SR-303 and HD 800, indicating low inter- and intra-individual variability of the predicted coloration effect. The higher IQR of the BK211 compared to the MushRoom is mainly due to the measurements on the KEMAR (the black dots on top), showing that it might be more difficult to place the more unusual device on an artificial head. Some headphones show higher variability, such as the modified K702 and the HD650. For many headphones, the dots of the same color are clustered together, showing that the differences are mostly inter-personal. This can be quantified by the ratio of the mean of the variance per subject to the total variance. It is between 9% for the Mysphere, indicating that the differences are mostly inter-personal, and 45% for the K1000, indicating comparatively larger intra-individual variability.

Fig. 3b shows the predicted QE at the model output. Note that even the "Open Ear" and "Open Ear 2" measurements, taken without headphones, lead to QEs around 20%, which may exceed the range expected from localization tests without impairments. These absolute estimates depend on the model parameters, which could be tuned for each subject using listening experiment data. However, the relative differences are expected to remain the same across conditions even if the parameters are fixed. Yet, the similarity of the two Open Ear values indicates that the setup remains consistent throughout the measurement period. As expected, very open headphones like the MushRoom and the BK211 perform the best. Most noticeably, the K1000, which caused a large coloration impairment, retains relatively low median predicted QE. Moreover, the variability of the QE predictions is generally larger than that of the CLL predictions. Also, the ratio of intra-individual variance to total variance is larger, now between 16% for the K702 and 69% for the HD 800. Thus, in the case of predicted QE, variability depends more on replacement than in the case of predicted coloration.

4 Reproduction

When reproducing sound using a pair of headphones, the spectral characteristics are determined by the HpTF. In most cases of consumer media playback, headphones are used without equalization, making the HpTF an important quality aspect that determines user preference [33]. In contrast, during binaural listening experiments in the lab, the headphones are most often equalized. Even if the equalization is individual, a



(a) Non-individual EQ.

(b) Individual EQ.

Fig. 4: Violin plots of the reproduction CLL predicted by the model. Black squares indicate medians and bars indicate IQRs. Different colors belong to different subjects. Black dots belong to the KEMAR HATS measurements. Individual measurements caused outliers.

headphone equalizer (EQ) usually is not created based on an HpTF measurement conducted on a subject immediately before presenting an experiment, without moving the headphones at all (see [24] for an exception). Thus, we argue that the most important characteristic is not a specific frequency response, but rather the inter- and intra-individual variability of it. In case a non-individual EQ is used, both sources of variability influence the result. If the headphones are equalized separately for each subject, only the intra-individual variability determines the quality of the final rendering.

4.1 Method

To assess the effect of inter- and intra-individual variability on equalized binaural rendering, we show the distribution of ΔCLL values obtained in reproduction with non-individual and individual equalization. Here, the reference was a flat response. Individualized EQs were created by averaging the HpTFs for each ear separately over repositioning, normalizing at 1 kHz, applying 3rd-octave smoothing, and inverting the result with a maximum gain of 10 dB. For the non-individual EQs, averaging was also performed over subjects.

4.2 Results and Discussion

Plots of HpTFs using the percentiles are found in the online repository. The model-based evaluation of the reproduction is shown in Figure 4a and 4b. On the left side, the coloration of the reproduction is shown when using a non-individual EQ. It is determined by inter- and intra-individual variability. On the right side, CLL errors are shown after employing an individual EQ that compensates for inter-individual variability, leaving only the intra-individual effect.

As expected, individualization helps in reducing coloration error. Whereas the median CLL differences were all above 1 dB when applying a non-individual EQ, they are below 1 dB after individual equalization. Interestingly, some headphones benefit more from individualization than others. Despite the median errors in the non-individual case being lowest for the HD 650 and HD 800, which would suggest a trade-off between high reproduction quality and their low transparency, large errors are seen for some individuals. These errors are reduced through individual equalization. Moreover, individual EQ results are similar for K1000 and BK211; the Mysphere receives lower scores, and the MushRoom scored the lowest in CLL difference.

5 Conclusion

We presented data and evaluation of passive transparent headphones for AR audio research based on data of nine human subjects and one HATS.

We stressed that transparency and reproduction quality are two distinct characteristics that depend on the headphone design. We showed that predicted perceptual transparency depends on inter- and intra-individual variability, where predicted quadrant error induced by wearing the devices varied more and depended more on intra-individual variability (i.e, differences between measurements made by the same subject) than CLL. Regarding reproduction, we showed that different headphones lead to varying degrees of playback coloration in combination with individual and non-individual equalization. In future work, different equalization strategies should be used for this kind of evaluation.

6 Acknowledgment

The authors would like to thank Frank Schultz (University of Rostock) and Fabian Brinkmann (TU Berlin) for lending us the BK211 headphones, and Vincent Ederle (TU Graz) for providing the K1000 headphones. PL's contribution was funded by EP/X032914/1, "Challenges in Immersive Audio Technology".

References

[1] Erbes, V., Schultz, F., Lindau, A., and Weinzierl, S., "An extraaural headphone system for optimized binaural reproduction," in *DAGA*, pp. 313–314, Darmstadt, Germany, 2012.

[2] Meyer-Kahlen, N., Rudrich, D., Wirler, S. A., Winther, S., and Frank, M., "DIY Modifications for Transparent Headphones," in *148th Audio Engineering Society Convention*, 2020.

[3] Mülleder, A. and Zotter, F., "Ultralight circumaural open headphones," in *154th Audio Engineering Society Convention*, Espoo, Finland, 2023.

[4] 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), pp. 618–639, 2004.

[5] Rämö, J. and Välimäki, V., "Digital Augmented Reality Audio Headset," *Journal of Electrical and Computer Engineering*, 2012.

[6] Meyer-Kahlen, N., Amengual Garí, S. V., Schlecht, S. J., and Lokki, T., "Testing Auditory Illusions in Augmented Reality: Plausibility, Transfer-Plausibility and Authenticity," *J. Audio Eng. Soc.*, 72(11), pp. 797–812, 2024.

[7] Martin, V., Meyer, J., and Picinali, L., "Exploring Dynamic Binaural Rendering Approaches in 6DoF Acoustic Augmented Reality," in *DAGA*, Copenhagen, Denmark, 2025.

[8] Lübeck, T., Weber, T., and Pörschmann, C., "A Perceptual Study on the Plausibility of Binaural Rendering in Anechoic Environments," *J. Audio Eng. Soc.*, 72(12), pp. 846–859, 2024.

[9] Stärz, F., Van De Par, S., Roßkopf, S., Kroczek, L. O. H., Mühlberger, A., and Blau, M., "Comparison of binaural auralisations to a real loudspeaker in an audiovisual virtual classroom scenario: effect of room acoustic simulation, HRTF dataset, and head-mounted display on room acoustic perception," *Acta Acust.*, 9, 2025.

[10] Rendle, G., Immohr, F., Kehling, C., Lammert, A., Kreskowski, A., Brandenburg, K., Raake, A., and Froehlich, B., "Influence of Audiovisual Realism on Communication Behaviour in Group-to-Group Telepresence," in *IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, Saint Malo, France, 2025.

[11] Schneiderwind, C. and Neidhardt, A., "Modified late reverberation in an audio augmented reality scenario," in *Proceedings of the International Conference on Digital Audio Effects (DAFx 2024)*, Guildford, UK, 2024.

[12] Satongar, D., Pike, C., Lam, Y., and al, e., "The Influence of Headphones on the Localization of External Loudspeaker Sources," *J. Audio Eng. Soc.*, 63(10), pp. 799–810, 2015.

[13] Pörschmann, C., Arend, J. M., and Gillioz, R., "How wearing headgear affects measured head-related transfer functions," in *EAA Spatial Audio Signal Processing Symposium*, Paris, France, 2019.

[14] Schneiderwind, C., Neidhardt, A., and Meyer, D., "Comparing the effect of different open headphone models on the perception of a real sound

source," in *150th Audio Engineering Society Convention*, 2021.

[15] Lladó, P., McKenzie, T., Meyer-Kahlen, N., and Schlecht, S. J., "Predicting perceptual transparency of head-worn devices," *J. Audio Eng. Soc.*, 70, pp. 585–600, 2022.

[16] Lladó, P., Hyvärinen, P., and Pulkki, V., "The impact of head-worn devices in an auditory-aided visual search task," *J. Acoust. Soc. Am.*, 155(4), pp. 2460–2469, 2024.

[17] Watanabe, Y., Chiba, H., Noguchi, K., Itou, H., and Kako, T., "Effects of Acoustic Transparency of Wearable Audio Devices on Audio AR," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 2025, ISBN 9798400713941.

[18] Lladó, P., Barumerli, R., Baumgartner, R., and Majdak, P., "Predicting the effect of headphones on the time to localize a target in an auditory-guided visual search task," *Frontiers in Virtual Reality*, 5, 2024.

[19] Mülleder, A., Romanov, M., Meyer-Kahlen, N., and Zotter, F., "Do-it-yourself headphones and development platform for augmented-reality audio," in *AES Conference on Spatial and Immersive Audio*, Huddersfield, UK, 2023.

[20] Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F., "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, 43(4), 1995.

[21] Völk, F., "Inter- and Intra-Individual Variability in the Blocked Auditory Canal Transfer Functions of Three Circum-Aural Headphones," *J. Audio Eng. Soc.*, 62(5), pp. 315–323, 2014.

[22] Hofmann, A., Meyer-Kahlen, N., Schlecht, S. J., and Lokki, T., "Audiovisual Congruence and Localization Performance in Virtual Reality: 3D Loudspeaker Model vs. Human Avatar," *J. Audio Eng. Soc.*, 72(10), pp. 679–690, 2024.

[23] Riedel, S., Frank, M., and Zotter, F., "Effect of HRTFs and head motion on auditory-visual localization in real and virtual studio environments," *Acta Acust.*, 9, 2025.

[24] Brinkmann, F., Lindau, A., and Weinzierl, S., "On the authenticity of individual dynamic binaural synthesis," *J. Acoust. Soc. Am*, 142(4), pp. 1784–1795, 2017.

[25] McKenzie, T., Schlecht, S., and Pulkki, V., "Auralisation of the Transition between Coupled Rooms," in *International Conference on 3D and Immersive Audio*, online, 2021.

[26] Meyer-Kahlen, N., Amengual Garí, S. V., McKenzie, T., Schlecht, S. J., and Lokki, T., "Transfer-plausibility of binaural rendering with different real-world references," in *DAGA*, Hannover, Germany, 2022.

[27] Klanjscek, N., David, L., and Frank, M., "Evaluation of an E-Learning Tool for Augmented Acoustics in Music Education," *Music & Science*, 4, 2021.

[28] Wirler, S., Meyer-Kahlen, N., and Schlecht, S. J., "Towards transfer-plausibility for evaluating mixed reality audio in complex scenes," in *AES International Conference on Audio for Virtual and Augmented Reality*, online, 2020.

[29] Denk, F., Brinkmann, F., Stirnemann, A., and Kollmeier, B., "The PIRATE: an anthropometric earPlug with exchangeable microphones for Individual Reliable Acquisition of Transfer functions at the Ear canal entrance," in *DAGA*, Rostock, Germany, 2019.

[30] Ono, K., Pulkki, V., and Karjalainen, M., "Binaural Modeling of Multiple Sound Source Perception: Coloration of Wideband Sound," in *112th AES Convention*, Munich, Germany, 2002.

[31] Baumgartner, R., Majdak, P., and Laback, B., "Modeling sound-source localization in sagittal planes for human listeners," *J. Audio Eng. Soc.*, 136(2), pp. 791–802, 2014.

[32] Majdak, P., Hollomey, C., and Baumgartner, R., "AMT 1.x: A toolbox for reproducible research in auditory modeling," *Acta Acust.*, 6, 2022.

[33] Olive, S. E., Welti, T., and McMullin, E., "The Influence of Listeners' Experience, Age, and Culture on Headphone Sound Quality Preferences," in *137th AES Convention*, Los Angeles, USA, 2014.