forum acusticum 2026
8-12 September · Graz, Austria

EVALUATING THE IMPACT OF HEAD-WORN DEVICE CONFIGURATIONS ON SPATIAL HEARING FOR AUGMENTED REALITY APPLICATIONS

Julie Meyer[1,2], Fulvio Missoni[3]

[1] *Department of Engineering, King's College London, England*
[2] *Dyson School of Design Engineering, Imperial College London, England*
[3] *DIBRIS, University of Genoa, Italy*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

With the recent surge in augmented reality applications, devices such as head-mounted displays and headphones are increasingly used as multimodal interfaces that combine visual and auditory information to create immersive experiences. However, to date, studies have focused on evaluating the impact of head-worn devices on head-related transfer functions (HRTFs) considering either different headphones or head-mounted displays without considering configurations where two such devices are used simultaneously, as sometimes can be in augmented reality. This study aims at bridging this gap and presents a perceptually-based numerical evaluation of the impact of head-worn device configurations on HRTFs. To this end, HRTFs were measured on a KEMAR mannequin wearing four configurations of head-worn devices, including two pairs of headphones (MushRoom and AKG K1000) measured with and without a head-mounted display (Meta Quest 3S). Results are presented in terms of interaural cues, estimates of perceived localisation-, and colouration-induced error metrics.

Keywords: *head-related transfer functions, head-worn devices, augmented reality, auditory models*

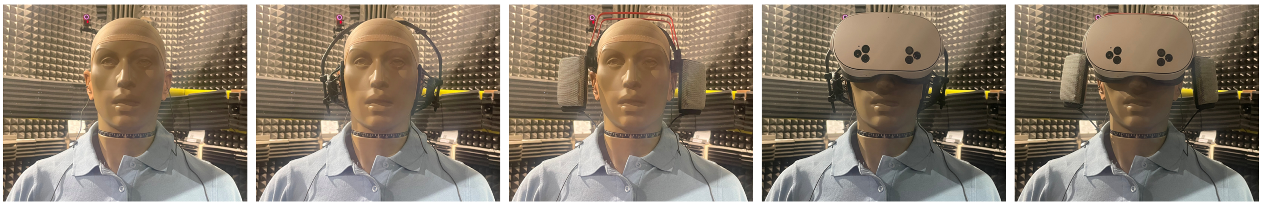
1. Introduction

In audio augmented reality (AAR) settings, virtual sounds aim at seamlessly blend with real ones from the physical world. This is normally achieved by rendering the virtual sounds binaurally through headphones. Binaural rendering integrates head-related transfer functions (HRTFs) such that a virtual sound source can be perceived as coming from specific locations in 3D around the listener. In addition to headphones, head-mounted displays (HMDs) have been also recently increasingly employed for research purposes in AAR contexts (e.g., [1]). However, head-worn devices (HWDs) such as HMDs and headphones alter the acoustic cues, captured by HRTFs, that listeners normally use to localise sounds in the free field. In an AAR setting, this can cause a mismatch between the perceived locations of virtual and real sound sources, weakening the illusion of a seamless blend.

A large number of studies have measured [2], [3] and evaluated the effect of different HWDs on HRTFs. Two main approaches can be identified in the literature: a signal-based perspective, relying on numerical analyses [4], [5], and a perceptual perspective, examining how HWDs affect perception. The latter has been addressed either through listening experiments or through auditory-based models predicting the colouration and localisation impairments introduced by HWDs, as in [6]. Most of those studies combined both numerical quantification and listening experiments as in [7], [8], [9]. Others have specifically investigated the impact of HWDs on localisation abilities in static [10], [11] and dynamic tasks [12].

Despite this body of literature, to the best of the authors' knowledge, no prior research has numerically examined the acoustic effect of wearing a pair of headphones and an HMD simultaneously, as may occur in AAR contexts such as in [1]. The only related work is the authors' own recent study [13], which addressed this combined setup from a purely perceptual standpoint, using a modified visual search task with eye-tracking. As such, this study bridges the gap and evaluates the effect of wearing two pairs of headphones, MushRoom [14] and AKG K1000 (AKG, Austria), with and without a Meta Quest 3S (Meta Corporation, USA) on HRTFs. The effect of HWDs on HRTFs is evaluated using perceptually-based metrics computed numerically, including interaural cues,

Corresponding author: *julie.meyer@kcl.ac.uk.*



(a) (b) (c) (d) (e)

Figure 1: HWD configurations during HRTF measurements: (a) reference, (b) MushRoom, (c) AKG (with the driver panels fully open), (d) MushRoom and Quest, and (e) AKG (with the driver panels fully open) and Quest.

predicted localisation behaviour in both horizontal and vertical dimensions, and colouration.

Although several studies have investigated the impact of HWDs on HRTFs, to the best of the authors' knowledge, no prior research (apart from a recent related publication by the same authors [13]) has examined a setup in which both a pair of headphones and an HMD are worn simultaneously, as may occur in AAR contexts such as in [1]. As such, this study bridges the gap and evaluates the effect of wearing two pairs of headphones, MushRoom [14] and AKG K1000 (AKG, Austria), with and without a Meta Quest 3S (Meta Corporation, USA) on HRTFs. The effect of HWDs on HRTFs is evaluated using perceptually-based metrics computed numerically such as interaural cues, predicted localisation behaviour in both horizontal and vertical dimensions as well as colouration.

2. Methodology

2.1. HWD selection

The MushRoom headphones [15] were selected for this study as they are becoming increasingly used in AR research (e.g., [16], [17]). They were developed as a simple and low-cost alternative to high-quality commercial open-design headphones such as the AKG K1000, which is also why the AKG K1000 headphones were selected for this study. In addition, prior work has shown that the MushRoom headphones are highly transparent compared to other passive headphone models, including the AKG K1000 [6]. As for the Meta Quest 3S, it was selected as a representative contemporary AR headset since it has recently been adopted in a growing number of AR research studies, including investigations of AAR navigation and immersive user experiences [18]. For brevity, the AKG K1000 headphones and Meta Quest 3S will be hereafter referred to as AKG and Quest, respectively.

2.2. HRTF measurements

To enable the numerical analysis, HRTFs were measured on a KEMAR mannequin (GRAS Sound & Vibration, Denmark), as well as on the KEMAR wearing four different combinations of HWDs[1]. The HRTF measurements were conducted in a semi-anechoic facility following the same protocol, hardware configuration, and signal processing steps detailed in [19]. More specifically, 793 head-related impulse responses were measured per HWD configuration using the multiple exponential sweep method [20] for source positions located 1.5 m from the KEMAR's head with azimuth angles sampled at 5° intervals and elevation angles ranging from -45° to 90° (sampled every 10° between -30° and 30°, and every 15° below and above that).

The configurations differed in HWDs present during the HRTF measurements: (a) no device (i.e., KEMAR only), hereafter referred to as the reference, (b) MushRoom, (c) AKG, (d) MushRoom and Quest, and (e) AKG and Quest. Figure 1 shows the five configurations for which HRTFs were measured. In the subsequent analysis, free-field compensated HRTFs were used.

2.3. Numerical Analysis

The numerical analysis of the measured HRTFs was conducted in MATLAB™ (MathWorks) using the Auditory Modelling Toolbox (version 1.6) [21]. The analyses on the interaural cues and on localisation were performed using methods and auditory models provided in the toolbox, whereas the colouration analysis used the model provided in [8].

3. Results and discussion

3.1. Interaural cues

Human sound source localisation relies on interaural time and level differences (ITD and ILD) for horizontal localisation, and on monaural spectral cues for elevation and front–back resolution. To quantify the scattering effect of HWDs on the interaural cues, ITD (via the Threshold method [22] with a 3 kHz low-pass filter applied to the head-related impulse responses (HRIRs) and a -10 dB threshold level), ILD (estimated as in [23]), and the interaural cross-correlation (IACC) (from the 3 kHz low-pass filtered HRIRs) were computed across the horizontal plane for all configurations and compared against the unobstructed reference. Results are shown in Figure 2.

ITD deviations were limited across all conditions, with the median ΔITD falling within the anechoic JND range of 20–30 μs reported by [24], indicating perceptually negligible changes. The exception was the MushRoom configuration, where the median value was higher (42 μs), suggesting a perceptible effect.

[1] The HRTFs are publicly available on Zenodo at https://doi.org/10.5281/zenodo.19223828.

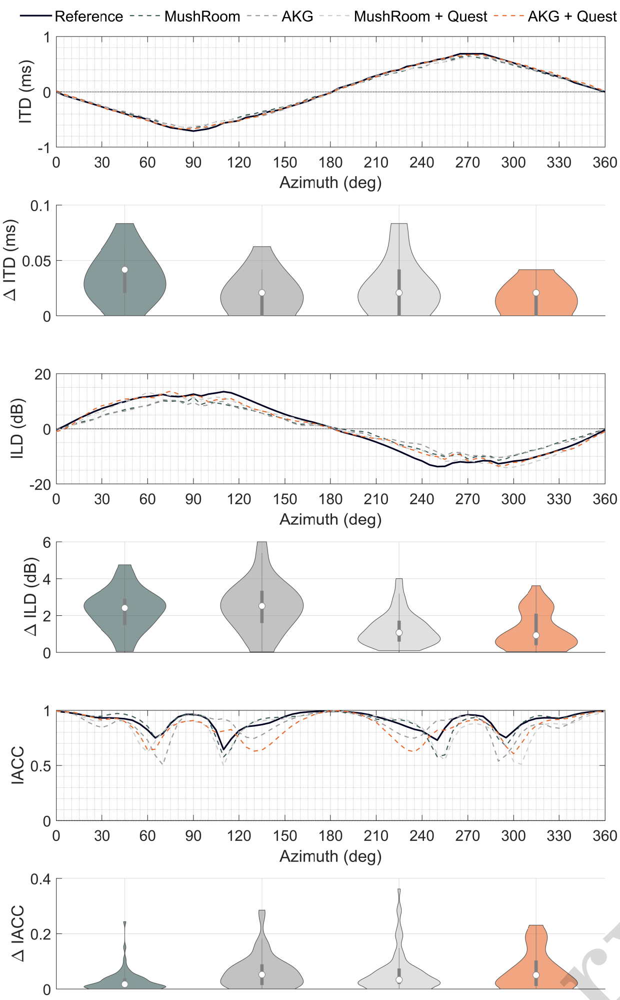


Figure 2: Interaural cues as a function of azimuth angle and difference distributions relative to the reference for each head-worn condition. All metrics were computed for sources in the horizontal plane only.

ILD deviations were generally larger without the HMD (MushRoom: median ΔILD = 2.41 dB, AKG: median ΔILD = 2.52 dB) than with it (MushRoom + Quest: median ΔILD = 1.07 dB, AKG + Quest: median ΔILD = 0.93 dB). On average, deviations without the HMD exceeded the 1 dB anechoic JND [25], particularly for lateral directions, whereas those with the HMD did not — indicating a benefit of wearing the HMD. No significant differences in ITD or ILD deviations were found between HWD configurations.

IACC analysis showed a limited effect of HWDs relative to the reference condition, with no significant differences observed between conditions (see Figure 2, last panel). The effect was negligible in the front sector across all configurations. Maximum decorrelation occurred for diagonal directions, for both the MushRoom and AKG configurations (MushRoom: ΔIACC = 0.036; AKG: ΔIACC = 0.056), and was further increased by the addition of the HMD (MushRoom + Quest: ΔIACC = 0.127; AKG + Quest: ΔIACC = 0.094). Wearing the HMD also increased lateral decorrelation for both HWDs (MushRoom: 0.003 → 0.062; AKG: 0.016 → 0.061), even though no significant differences were reported.

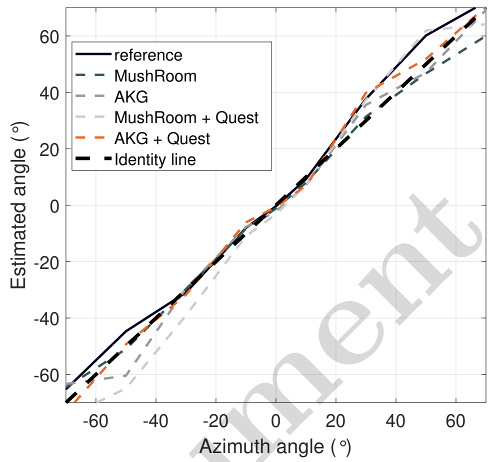


Figure 3: Estimated perceived lateral angle as a function of target lateral angle for each HWD configuration. The black dashed line is the identity line, representing an ideal reference.

3.2. Auditory models

3.2.1. Localisation

In order to interpret the computed interaural cues on perceived localisation, the auditory-based model proposed in [26] was employed. The model trained a neural network with data collected in a subjective listening experiment, where horizontal plane localisation was assessed while wearing different HWDs. More specifically, the model estimates the perceived direction and the localisation blur when wearing a HWD. The method has proven to be suitable to predict the degradation of localisation ability for unseen HWDs in the frontal horizontal plane in [12].

First, the estimated perceived lateral angle was computed as a function of target lateral angles (see Figure 3). Furthermore and similarly to [12], the median absolute deviation (MAD) from the reference was computed for each HWD configuration, h, as follows

$$\mathrm{MAD}^h = \mathrm{median}(|e_\theta^h - e_\theta^{\mathrm{reference}}|)\,, \qquad (1)$$

where the median is computed over target lateral angles θ and corresponding estimates of perceived lateral angle, e_{lat}, and localisation blur, e_{blur}. Target lateral angles considered were 0°, ± 10°, ± 30°, ± 50° and ± 70° in the horizontal plane (i.e., at elevation 0°). The results, reported in Table 1, indicate a limited impact on lateral localisation accuracy and precision.

To evaluate the perceptual relevance of the spectral variations in the HRTF magnitudes for the dif-

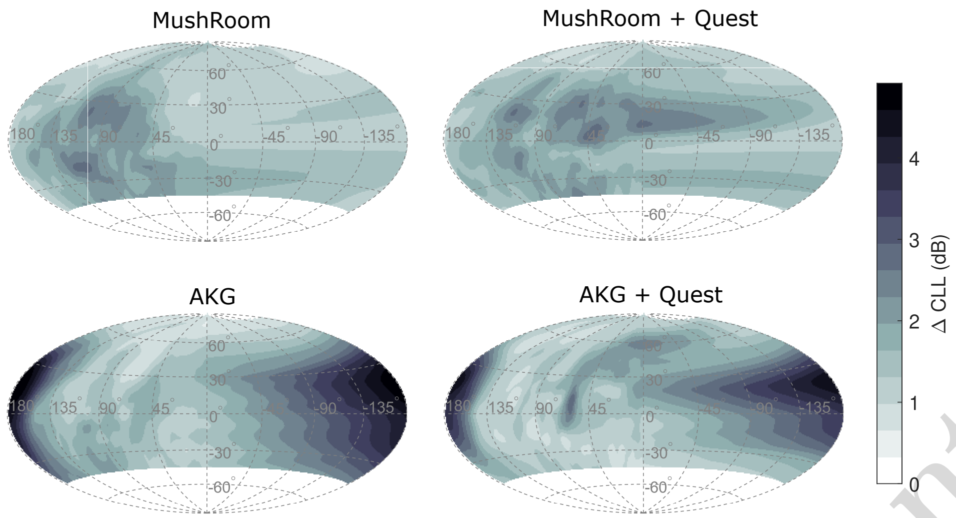


Figure 4: Predicted colouration calculated with respect to the reference.

HWD configuration	MAD_{lat} (°)	MAD_{blur} (°)
MushRoom	1.3	1.1
AKG	2.1	1.1
MushRoom + Quest	2.9	0.7
AKG + Quest	2.2	0.7

Table 1: Difference in the estimated lateral localisation metrics of each HWD compared to the reference.

ferent HWD configurations, specifically looking at sound source localisation in the vertical dimension, the Bayesian computational model described in [27] was employed. The model takes a "template" and a "target" set of HRTFs as inputs and predicts the localisation response of a subject (whose set of HRTFs is the template) for a given target HRTF set. The model first spatially interpolates the two provided input sets of HRTFs at uniformly distributed directions (to account for non-uniform HRTF acquisition grids), then computes the following features on the interpolated grid of directions for both inputs: the ITD and ILD cues, as well as the spectral amplitudes of the directional transfer functions. From these calculated features, the model, which is based on a template-matching procedure, outputs the predicted localisation performance of the subject whose set of HRTFs was provided as the template for the given target set of HRTFs. Here, the template HRTFs was selected as the reference HRTFs and the target HRTFs were the HWD HRTFs configurations.

Following the predicted localisation responses from the computational model, two error metrics were calculated. The quadrant error (QE, in %), defined as absolute polar errors between actual source positions and predicted response positions exceeding 90°, was used to quantify the non-local performance in the polar dimension. As for quantifying the local performance in the polar dimension, the local polar root-mean-square error (PE, in °) was calculated as the root-mean-square average of polar errors that were less than 90° in magnitude. For QE and PE, only responses within the lateral range of ± 30° were considered, as in [28]. The results from the Bayesian model in terms of QE and PE are reported in Table 2. While PE is similar across configurations, QE seems to differ between the headphones configurations regardless of the presence or absence of the HMD, which indicates that the effect of the HMD on QE is not significant.

Considering that PE is relatively high, the Bayesian computational model was also run with both template and target HRTFs corresponding to that of the reference, as a sanity check. From this sanity check, QE = 8.3% and PE = 36.5°, which confirms that the model worked as expected as those values are comparable to results from previous literature (e.g., see Fig. 7 (top row) from [19]).

HWD configuration	QE (%)	PE (°)
MushRoom	18.4	41.4
AKG	32.6	42.6
MushRoom + Quest	21.3	43.4
AKG + Quest	30.8	42.1

Table 2: Difference in the estimated polar localisation metrics of each HWD compared to the reference.

3.2.2. Colouration

While localisation errors and predicted responses provide insight on how the localisation cues are altered, this does not necessarily inform on the perceived colouration induced by HWDs. To evaluate the impact of the HWDs on colouration, the model proposed in [8] was used. The results from the model, displayed in Figure 4, showed the predicted colouration calculated as the difference in composite loudness level (ΔCLL) between the reference and the HWD configurations.

The predicted colouration deviated the most from the reference for the HWDs configurations that included the AKG, especially for the rear source positions (i.e., at around ± 180° azimuth). The HWD configurations with the Quest seemed to marginally impact ΔCLL compared to the configurations where only headphones were present and for the frontal source positions only.

4. Conclusion

This study presented a numerical evaluation of the impact of HWDs on HRTFs using perceptually motivated metrics. HRTFs were measured from a KEMAR mannequin fitted with four HWD configurations — two pairs of headphones (MushRoom and AKG K1000), each with and without an HMD (Meta Quest 3S) — and analysed in terms of interaural cues, localisation, and colouration.

Across the horizontal plane, the binaural cues were only moderately affected. ITD deviations remained within the reported JND [24], except for the MushRoom configuration, where they were moderately higher. ILD deviations were somewhat larger, exceeding the anechoic JND [25] in the headphones-only conditions. A plausible interpretation is that adding the HMD restores a smoother, more head-like geometry around the ears, partially compensating for the scattering introduced by the headphones alone. Interaural coherence was likewise contained: negligible in the frontal sector across all devices, with the strongest decorrelation confined to diagonal directions and increasing further once the HMD was added. At the perceptual level, the auditory models indicated a limited effect on lateral localisation, whereas front–back confusion was significantly increased as predicted colouration, both most affected for the AKG K1000 and broadened toward the rear.

Taken together, the AKG K1000 emerged as the most disruptive device across metrics, albeit with a limited effect overall, whereas adding the Meta Quest 3S produced comparatively minor further changes. Whether these acoustic alterations are perceptually consequential remains an open question. As a preliminary numerical evaluation, our results suggest that their impact is likely limited, affecting mainly front–back confusion. In recent work, the same authors employed an auditory-aided visual search task in which listeners were tested with disrupted HRTFs [13]; there too the effect was limited, leaving overall search performance largely unaffected and appearing primarily as a small increase in eye-movements searching, while head movements were only marginally affected. This is consistent with the results from [10], which report that wearing an HMD increased localisation error only slightly, on the order of a degree, small enough that an HMD can reasonably be treated as an HMD-free condition outside the most demanding localisation paradigms. In [11], the HMD-related localisation errors were similarly attributed to the induced ITD and ILD changes, while showing that the amount of visual information available strongly modulates the outcome. The role of listener auditory experience in mediating these effects remains, however, an open avenue worth pursuing. Future work will include listening experiments based on the transfer-plausibility paradigm [29] to test whether HWDs affect auditory illusions: that is, whether a listener believes a virtual sound source to be real. This question could have significant implications for how listening experiments are designed: if such an effect were found, experimenters would need to compensate for the scattering introduced by HWDs.

5. Acknowledgments

The authors thank Lorenzo Picinali for providing access to the facility and equipment for the HRTF measurements, and Pedro Lladó for providing useful insight on the auditory model from [26]. The author FM thanks Andrea Canessa and Silvio P. Sabatini for financial support and for providing the opportunity to conduct this research.

References

[1] A. Hofmann, N. Meyer-Kahlen, S. Schlecht, and T. Lokki, "Audiovisual congruence and localization performance in virtual reality: 3D loudspeaker model vs. human avatar," *Journal of the Audio Engineering Society*, vol. 72, no. 10, pp. 679–690, 2024.

[2] A. Genovese and A. Roginska, "HMDiR: An HRTF dataset measured on a mannequin wearing XR devices," in *AES 2019 International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[3] R. Schlieper, S. Preihs, and J. Peissig, "An open dataset of measured HRTFs perturbed by headphones," in *Audio Engineering Society Convention*, vol. 152, 2022, pp. 593–598.

[4] R. Gupta, R. Ranjan, J. He, and W.-S. Gan, "Investigation of effect of VR/AR headgear on Head related transfer functions for natural listening," in *AES 2018 International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[5] C. Pörschmann, J. M. Arend, and R. Gillioz, "How wearing headgear affects measured head-related transfer functions," in *EAA Spatial Audio Signal Processing Symposium*, 2019, pp. 49–54.

[6] A. Mülleder, P. L. Gonzalez, and N. Meyer-Kahlen, "Comparison of Passive Transparent Headphones for Augmented Reality Audio," in *AES International Conference on Headphone Technology*, Acoustical Society of America, 2025.

[7] C. Schneiderwind and A. Neidhardt, "Comparing the effect of different open headphone models on the perception of a real sound source," in *150th Audio Engineering Society Convention 2021 proceedings*, Audio Engineering Society (AES), 2021.

[8] P. Lladó, T. Mckenzie, N. Meyer-Kahlen, and S. Schlecht, "Predicting perceptual transparency of head-worn devices," *Journal of the Audio Engineering Society*, vol. 70, no. 7/8, pp. 585–600, 2022.

[9] Y. Watanabe, H. Chiba, K. Noguchi, H. Itou, and T. Kako, "Effects of acoustic transparency of wearable audio devices on audio AR," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025, pp. 1–15.

[10] D. Poirier-Quinot and M. S. Lawless, "Impact of wearing a head-mounted display on localization accuracy of real sound sources," *Acta Acustica*, vol. 7, p. 3, 2023.

[11] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau, "Sound source localization with varying amount of visual information in virtual reality," *PloS one*, vol. 14, no. 3, e0214603, 2019.

[12] P. Lladó, P. Hyvärinen, and V. Pulkki, "The impact of head-worn devices in an auditory-aided visual search task," *The Journal of the Acoustical Society of America*, vol. 155, no. 4, pp. 2460–2469, 2024.

[13] F. Missoni, J. Meyer, and A. Canessa, "Investigating the perceptual impact of head-worn devices for augmented reality using a dynamic task with continuous head–eye tracking," in *6th International Conference on Audio for Virtual and Augmented Reality and Immersive Games*, Audio Engineering Society (AES), 2026.

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

[15] A. Mülleder, M. Romanov, N. Myer-Kahlen, and F. Zotter, "Do-it-yourself headphones and development platform for augmented-reality audio," in *AES 2023 International Conference on Spatial and Immersive Audio*, Audio Engineering Society, 2023.

[16] S. Riedel, M. Frank, and F. Zotter, "Effect of HRTFs and head motion on auditory-visual localization in real and virtual studio environments," *Acta Acustica*, vol. 9, p. 21, 2025.

[17] C. Schneiderwind, E. De Sena, and A. Neidhardt, "Perceptual effects of modified late reverberation and reverberation time in auditory augmented reality in two rooms," *Acta Acustica*, vol. 10, p. 17, 2026.

[18] S. Hinzmann et al., "Finding my way: Influence of different audio augmented reality navigation cues on user experience and subjective usefulness," in *2025 17th International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, 2025, pp. 1–4.

[19] I. Engel et al., "The SONICOM HRTF dataset," *Journal of the Audio Engineering Society*, vol. 71, no. 5, pp. 241–253, 2023.

[20] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, p. 623, 2007.

[21] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1. x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, vol. 6, p. 19, 2022.

[22] A. Andreopoulou and B. F. Katz, "Identification of perceptually relevant methods of inter-aural time difference estimation," *The Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 588–598, 2017.

[23] T. McKenzie, D. Murphy, and G. Kearney, "Interaural level difference optimisation of first-order binaural Ambisonic rendering," in *2019 AES International Conference on Immersive and Interactive Audio: Creating the Next Dimension of Sound Experience*, 2019.

[24] J. E. Mossop and J. F. Culling, "Lateralization of large interaural delays," *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1574–1579, 1998.

[25] S. Klockgether and S. van de Par, "Just noticeable differences of spatial cues in echoic and anechoic acoustical environments," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, EL352–EL357, 2016.

[26] P. Lladó, P. Hyvärinen, and V. Pulkki, "Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation," *Acta Acustica*, vol. 6, p. 1, 2022.

[27] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A Bayesian model for human directional localization of broadband static sound sources," *Acta Acustica*, vol. 7, p. 12, 2023.

[28] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.

[29] N. Meyer-Kahlen, S. Schlecht, S. V. A. Garí, and T. Lokki, "Testing auditory illusions in augmented reality: Plausibility, transfer-plausibility, and authenticity," *Journal of the Audio Engineering Society*, vol. 72, no. 11, pp. 797–812, 2024.