# Predicting perceptual transparency of head-worn devices

OPEN ACCESS Freely available online
P. Lladó, T. Mckenzie, N. Meyer-Kahlen, and S. J. Schlecht,, "Predicting Perceptual Transparency of Head-Worn Devices"
*J. Audio Eng. Soc.*, vol. 70, no. 7/8, pp. 585–600, (2022 July/August).
DOI: https://doi.org/10.17743/jaes.2022.0024

# Predicting Perceptual Transparency of Head-Worn Devices

**PEDRO LLADÓ,**[1] ***THOMAS MCKENZIE,**[1] **NILS MEYER-KAHLEN,**[1] *AES Student Member* **AND**
(pedro.llado@aalto.fi) (thomas.mckenzie@aalto.fi) (nils.meyer-kahlen@aalto.fi)
**SEBASTIAN J. SCHLECHT,**[1, 2] *AES Associate Member*
(sebastian.schlecht@aalto.fi)

[1]*Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland.*
[2]*Media Lab, Department of Art and Media, Aalto University, Espoo, Finland.*

Acoustically transparent head-worn devices are a key component of auditory augmented reality systems, in which both real and virtual sound sources are presented to a listener simultaneously. Head-worn devices can exhibit a high transparency simply through their physical design but in practice will always obstruct the sound field to some extent. In this study, a method for predicting the perceptual transparency of head-worn devices is presented using numerical analysis of device measurements, testing both coloration and localization in the horizontal and median plane. Firstly, listening experiments are conducted to assess perceived coloration and localization impairments. Secondly, head-related transfer functions of a dummy head wearing the head-worn devices are measured, and auditory models are used to numerically quantify the introduced perceptual effects. The results show that the tested auditory models are capable of predicting perceptual transparency and are therefore robust in applications that they were not initially designed for.

## 0 INTRODUCTION

In auditory augmented reality (AR) applications, the real world is enhanced with virtual sounds. In a high-quality rendering system, virtual sound sources should be believed to be real by listeners. In that case, the rendering can be called plausible [1, 2]. In practice, virtual sound sources should also be accepted as real, if they occur alongside real sources, a concept that we refer to as transfer-plausibility [3]. The delivery of virtual sound requires acoustic transducers, such as headphones or loudspeakers. In this study, the focus is on analyzing the perceptual effects caused by transducers mounted on head-worn devices (HWDs).

All HWDs have some effect on the external sound waves arriving at the eardrums, which bears several perceptual consequences such as coloration and reduced accuracy of sound-source localization. This may cause poorer speech intelligibility, unnecessary fatigue when carrying out simple tasks, and a worsened overall AR experience. Thus, a robust technique for assessing the acoustic transparency of HWDs is required. In this paper, perceptual transparency is defined as the degree to which an HWD impairs external sounds.

Wearing HWDs has multiple consequences for the perception of external sounds. Even though several studies aimed at assessing the quality of HWDs have been presented in the past [4–7], they often only focused on one perceptual aspect, such as coloration [4, 7] or horizontal localization [6]. Furthermore, the relationship between perceptual listening and objective measurements has not yet been investigated in detail.

Measurements with head and torso simulators wearing several HWDs designed for AR applications have previously been conducted to assess their introduced acoustic impairments objectively. Whereas some studies have focused on analyzing the head-related transfer functions (HRTFs) in anechoic conditions [5, 8–10], other studies concentrated on binaural room impulse responses (BRIRs) [4, 6]. In principle, these measurements contain the necessary information to characterize the HWD-induced effect. However, evaluations have mostly focused on calculating the frequency-averaged magnitude ratio between the transfer functions of the dummy head with and without the HWD fitted on it [4, 5, 8–11] and calculating the changes to interaural time and level differences as a measure of horizontal

*To whom correspondence should be addressed, e-mail: pedro.llado@aalto.fi

localization impairments [5, 8, 10, 11]. Perceptual study findings and objective metrics should be related to each other to estimate the impairments introduced by HWDs using computational models. Such successful models would help to avoid the complexity of formal perceptual testing for selecting a suitable device for the user's needs.

This paper aims at predicting the perceptual impairments caused by HWDs in static scenarios using acoustic measurements and pre-existing auditory models. Firstly, listening experiments are conducted on coloration and localization impairments caused by a set of HWDs that are relevant for AR research. Subsequently, auditory models are evaluated for their capabilities in predicting the perceptual impairments caused by HWDs and, because this is an application the auditory models were not initially designed for, their suitability in new applications.

This paper is laid out as follows. Sec. 1 discusses previous research in the area of auditory AR, HWD transparency, and auditory model usage in related fields. In Sec. 2, a dataset of HRTFs of a dummy head wearing four HWDs is measured. Sec. 3 then presents an evaluation of the coloration induced by the HWDs, both examining a perceptual study and using an auditory model. Sec. 4 presents an evaluation of the localization impairments caused by the HWDs, again both perceptually and using auditory models. Finally, the findings of the paper are concluded, along with proposed further work, in Sec. 5.

## 1 BACKGROUND

For auditory AR applications, there are two options for achieving acoustic transparency in HWDs. The simpler alternative, on which this study focuses, is to design the HWDs such that they disturb the natural sound field as little as possible, a term often defined as free-air equivalent coupling [12]. This is achieved through an open construction, for example, by placing transducers relatively far from the ears and using a supra-aural or extra-aural design. Such designs [13] have been shown as capable of producing authentic binaural synthesis [14] for certain stimuli types and are the focus of the experiments presented here. The other option, which is not discussed in this paper, is to use headphones with active transparency that rely on microphones, circuitry, and adaptive digital signal processing algorithms to play back external sound through the headphones. Audible artefacts, such as comb filtering and noise, can easily occur in active designs [11], and it appears that transparency has not yet been achieved [7] with such designs.

In terms of passive AR-related HWDs, several authors have published measurements of HRTFs with head and torso simulators wearing different headphones [8, 9, 15] and head mounted displays [8, 10, 5]. BRIRs were measured in [4, 6] in a similar manner, taking into account the effects of the room. So far, in most investigations, a measure of transparency has been calculated as the ratio between the smoothed magnitude response of a real or artificial listener when wearing the headphones and listening with open ears [4, 8–10, 5]. It is clear that such measures only serve as a rough indication of coloration and do not comprehensively reveal the induced perceptual impairments. As a first step toward predicting perceptual transparency that includes all the relevant dimensions, perceptual tests are therefore required.

Perceptual tests with AR-related HWDs have been conducted in [4], in which the similarity of seven different circumaural and extra-aural models to an open-ear condition was tested at different sound-source angles and distances. The results clearly showed the strong impairments caused by some circumaural models and smaller, but still perceivable, effects of extra-aural models. Overall timbral and spatial differences were investigated in [5] by synthesizing sound sources using measured HRTFs wearing head-mounted displays, in which small but significant effects were found in both tests. In [6], a horizontal plane localization test was conducted to compare open ears to wearing STAX SC-202 headphones. Localization was only slightly affected by the device, but the time taken by participants to locate the source was higher. This may be because of the need for additional time to resolve front-back confusions.

Apart from those experiments, the effects of more strongly impairing HWDs, such as hearing protection devices and protection helmets, on sound localization have been studied more extensively in the past. In terms of horizontal localization, the error increased in perceptual experiments when hearing protection [16] or protective headgear [17] were worn. In the vertical plane, HWDs increased the probability of having front-back confusions [17, 18] and up-down confusions [18]. HWDs also had an implication on perceptual tests where a sound source had to be found, affecting the search time and the head movement patterns [19]. However, the introduced effect varied from mild to severe depending on the HWDs and their characteristics [18]. Altogether, previous perceptual studies have focused on coloration, horizontal and vertical localization. Hence, these dimensions shall be tested in the following experiments as well, before comparing the results to auditory model predictions.

Models for predicting coloration typically attempt to approximate the summation of loudness across frequency [20]. Some include modeling of the middle ear [21]. The Composite Loudness Level (CLL) used in this study is a perceptual loudness model appropriate for predicting coloration of binaural signals, which uses middle-ear modeling, perceptual loudness weighting, and non-linear frequency weighting [22]. Some implementations of CLL use the adaptation network of [23] (included in [23]), such as in [24–26], which is a multilayer feedforward neural network to acquire mappings and properties from the front-end. In this study, the feasibility of the CLL model to predict HWD-induced coloration is assessed. Although recent models have been shown to outperform the CLL [27], the CLL is preferred in this study because it was the simplest model that was able to predict the data as well as other models. The implementation used in this study is without Karjalainen et al.'s model [23].

Models for predicting sound localization usually process the acoustic signals that reach the ears to estimate direction

of arrival of the sound based on binaural or monaural cues, depending on the scope of the model. May et al.'s horizontal localization model [28] has been proven to be robust in reverberant conditions and multiple sources. In the model, interaural time difference and interaural level difference estimates are computed with an inter-aural cross-correlation approach, after a Gammatone filter-bank [29] process with channel-dependent gains and half-wave rectification. The final azimuth estimation is made by a Gaussian Mixture Model trained under multiple sources and reverberation conditions. Even though the model was not trained for listening conditions in which HWDs were involved, it may be suitable to predict horizontal localization under such situations. In this study, the feasibility of May et al.'s model [28] to predict HWD-induced impairments in localization on the horizontal plane is assessed.

Baumgartner et al.'s model [30] analyzes monaural cues to estimate the direction of arrival in a given sagittal plane. The model procedure is template-based, whereby the target sound and template HRTFs (known directions in the given sagittal plane) are processed by a peripheral auditory model that consists of a Gammatone filter-bank [29] and the extraction of its spectral profile. The positive spectral gradient profile of the target sound is then compared with those of the template HRTFs. These comparisons produce a probabilistic prediction of the listener's distribution of responses across the polar dimension for the target sound.

This model has been proven in the past to predict effects caused by vector-based amplitude panning [31] and to predict complex effects outside its scope, such as sensorineural hearing loss with minor modifications [32]. In this study, the ability of Baumgartner et al.'s model [30] to predict the localization impairment caused by HWDs on the polar dimension is assessed. Moreover, the capability of the model to predict a group of listeners' mean response using generic HRTFs is evaluated, although this model was originally designed to be used to predict individual listeners' responses using their own HRTFs.

## 2 STUDIED HEAD-WORN DEVICES

To aid in the assessment of perceptual transparency of head-worn devices, a dataset of HRTFs was measured using the G.R.A.S. KEMAR 45BC head and torso simulator wearing different HWDs. Measurements were conducted in the multichannel anechoic chamber "Wilska" at Aalto University Acoustics Lab, Finland. In this study, HRTFs in anechoic conditions were measured to avoid any room reflections. Reverberation would complicate the study of direction-dependent effects caused by the HWDs. Because the impact of the room on coloration and localization can vary substantially, an anechoic environment was decided to be the best option to assess the effect caused by the HWDs in isolation.

HWDs were selected according to their relevance for AR and in order to facilitate comparison with previous measurements. Additionally, the selection was chosen across a range from small to large impairments. An Oculus Quest 2 [Fig. 1(a)] virtual reality head-mounted display was included, which is highly relevant for studies that involve external loudspeaker reproduction [33, 34]. As an extra-aural model, the Mysphere 3.2 [Figs. 1(b) and 1(c)] headphones, marketed as having a fully open design, were used, which have been shown to cause minimal disturbances to measured HRTFs [33] and are similar to the discontinued AKG K1000, which have been employed in several past studies involving real and virtual sounds. Because the Mysphere 3.2 headphones feature a variable transducer frame position, two different configurations were tested: one with the frames in an open position [Fig. 1(b)] and one with the frames closed, close to the ears [Fig. 1(c)]. A modified AKG K702 [Fig. 1(d)] was included as an inexpensive pair of headphones with increased transparency [9]. Finally, the Sennheiser HD650 [Fig. 1(e)] headphones were included as an HWD with relatively low transparency, despite its open back design.

In total, six HWDs configurations were measured. These are summarized as follows:

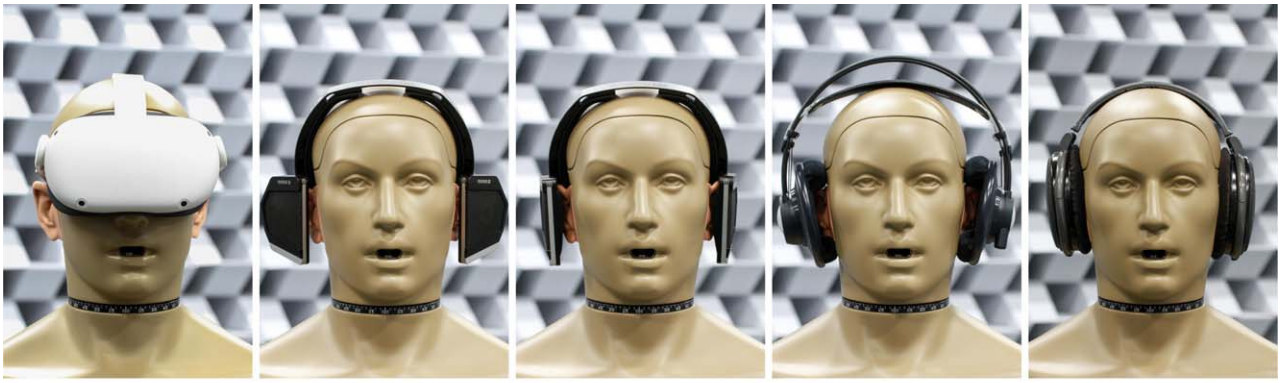
C1: Open ear (no HWD),
C2: Oculus Quest 2,
C3: Mysphere 3.2 with open frame position,
C4: Mysphere 3.2 with closed frame position,
C5: AKG K702 (with transparency modification [9]), and
C6: Sennheiser HD650.

For each HWD configuration, 45 HRTFs were measured (positions indicated in Fig. 4). Before measuring the impulse responses, the KEMAR dummy head was laser-aligned in the center of the loudspeaker array and care was taken to ensure that the fitting of headphones did not alter the alignment of the KEMAR. Impulse response measurements used exponential sine sweeps with a length of 1 s and a sampling rate of 48 kHz [35].

Expecting to observe different levels of impairment between the studied HWD configurations, three hypothesis are tested with regard to the subjective experiments. Firstly, (i) all HWDs will cause a difference with respect to open ears. Furthermore, (ii) the open and the closed configurations of the Mysphere 3.2 will be different, which is relevant for future AR experiments using this device. Lastly, (iii) the AKG K702 headphones (modified to improve transparency) will differ from the Mysphere 3.2 in its closed condition, which is of interest because it shows whether a simple modification of some inexpensive headphones can be comparable to specifically designed and engineered transparent headphones.

## 3 PREDICTION OF HWD-INDUCED COLORATION

The effect of HWDs on coloration was evaluated first through perceptual listening tests and then numerically through model-based evaluations using the measurements from Sec. 2. Finally, the perceptual and numerical results were compared to evaluate the applicability of replacing listening tests with numerical evaluation and measurements.



(a) C2 (b) C3 (c) C4 (d) C5 (e) C6

Fig. 1. KEMAR head and torso simulator wearing the measured head-worn devices. (a) Oculus Quest 2 virtual reality head-mounted display, (b) Mysphere 3.2 headphones with open frame position, (c) Mysphere 3.2 headphones with closed frame position, (d) AKG K702 headphones with transparency modification, and (e) Sennheiser HD650 headphones.

### 3.1 Listening Test Methodology

To assess the coloration of HWDs, a listening test was conducted that followed the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) paradigm, ITU-R BS.1534-3 [36], using webMUSHRA[1] [37]. The reference was C1, and the anchor was the reference low-passed at a cut-off frequency of 3.5 kHz. The test conditions consisted of convolutions between the base stimulus and measured HRTFs of the KEMAR wearing different HWDs. In each trial, the participant directly compared signals corresponding to all the measured HWD configurations at one sound-source direction and rated them in terms of perceived similarity to the reference, in which reduced similarity indicates increased coloration.

Three base stimuli were used: 1 s of monophonic pink noise, a 3-s anechoic male speech recording [38], and a 1-s diffuse rainfall simulation. All audio used a sample rate of 48 kHz, windowed by onset and offset half-Hanning ramps of 5 ms. To assess the coloration at specific directions, the pink noise and anechoic speech were convolved with five of the measured HRTFs at the following directions: $(\theta, \phi) = (0^\circ, 0^\circ)$, $(45^\circ, 30^\circ)$, $(90^\circ, 0^\circ)$, $(180^\circ, 0^\circ)$, and $(0^\circ, 90^\circ)$, where $(\theta, \phi)$ denotes azimuth and elevation in the boundaries $-180^\circ < \theta < 180^\circ$ and $-90^\circ < \phi < 90^\circ$, respectively. To assess coloration of diffuse sound, 45 separate 1-s excerpts of a monophonic rainfall recording were convolved with HRTFs at all 45 measured directions, and the results were summed for each ear.

Stimuli and conditions were randomized and presented double anonymous. Fifteen participants (average age: 28.5 years, three female, and 12 male) took part in the experiment, with self-reported normal hearing and prior critical listening experience (such as education or employment in audio or music engineering).

[1]https://github.com/audiolabs/webMUSHRA.

### 3.2 Listening Test Results

The results of the listening test are presented as violin plots [39] in Fig. 2 for the pink noise and speech stimuli at specific directions and Fig. 3 for the diffuse rain stimulus. Violin plots display the density trace and box plot together, which better illustrates the structure of the data than traditional box plots.

The data was first tested for normality using the Shapiro-Wilk test. Even excluding the reference condition, not all data was normally distributed. Therefore, non-parametric testing was used to assess which differences can be considered significant. Performing Friedman tests that compared the answers for each sound-source direction indicated highly significant differences (all tests $p < 0.001$).

Interestingly, when comparing results of different directions within one HWD, a significant effect ($p < 0.05$) of the direction was found for each model and both the pink noise and speech stimuli types, except for C4 under the noise condition ($p = 0.21$). Although the comparison between different models at each sound-source direction is the focus of the analysis, this result indicates the importance of testing several directions.

To further analyze the differences between the HWDs at each of the sound-source directions, Wilcoxon signed rank tests were performed as post-hoc tests. Reported $p$ values were corrected using the Bonferroni-Holm [40] procedure by considering all comparisons that were needed to test the three hypotheses at each direction and for each signal (70 paired comparisons in total).

The first point of analysis was to compare all HWDs against C1. Apart from C2, all tested HWD configurations exhibited statistically significant differences between the open-ear reference for both tested stimuli types. As expected, C2 induced the least coloration. However, some statistically significant differences were present.

For the pink noise stimulus, a significant difference between the open-ear conditions, which had a median (Mdn) = 100, at all directions, was observed for the frontal direction $(0^\circ, 0^\circ)$, Mdn = 72: $Z = 120, p = 0.004$; at $(45^\circ, 30^\circ)$,

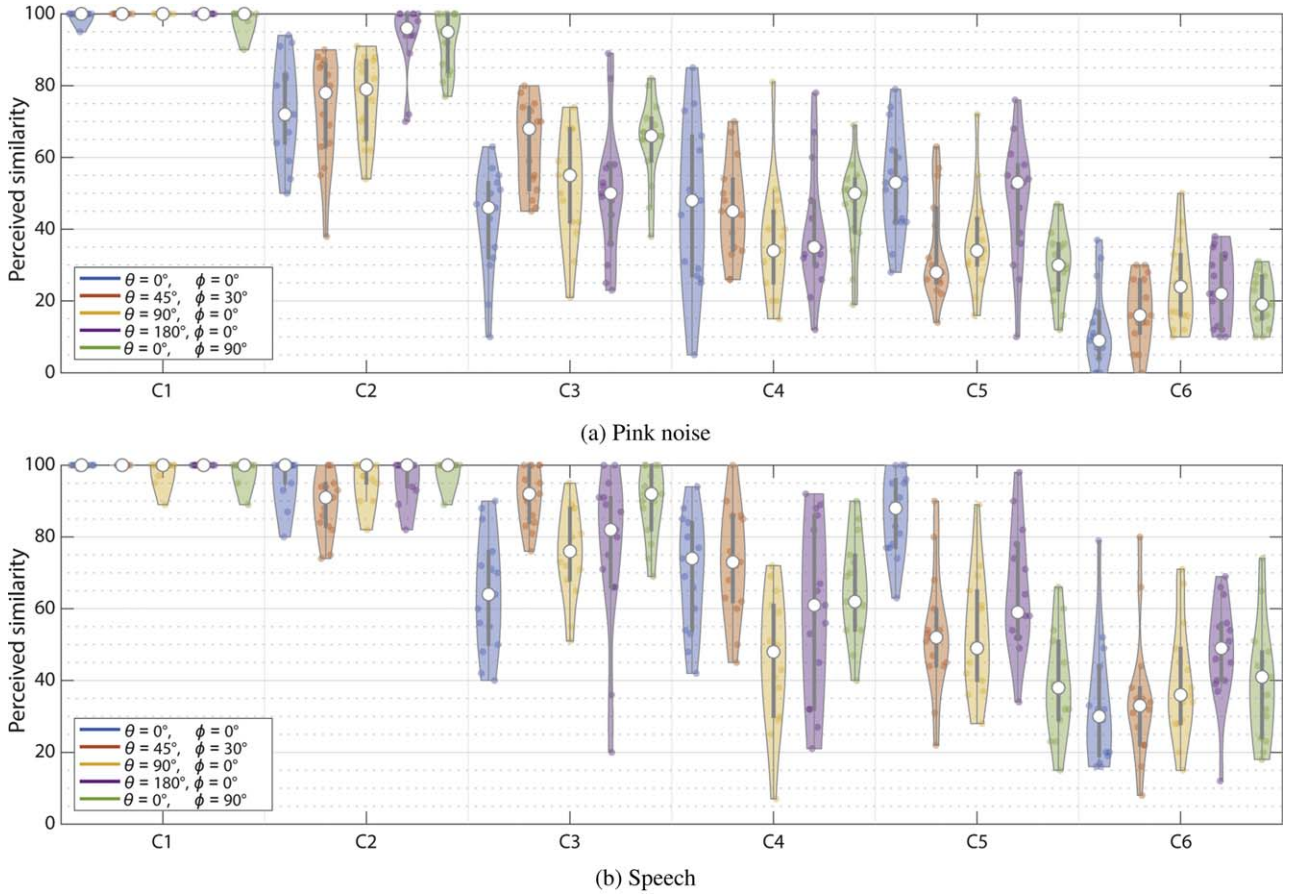



Fig. 2. Violin plots of the coloration listening test for all tested head-worn–device configurations at specific directions. Individual results are displayed as colored points, and the width of the violin indicates data density. The median values are presented as a white point, interquartile ranges are marked using a thick gray line, and ranges between the lower and upper adjacent values are marked using a thin gray line. C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650. (a) Pink noise and (b) speech stimuli types.

Mdn $= 78: Z = 120, p = 0.004$; and the lateral direction (90°, 0°), Mdn $= 79: Z = 120, p = 0.004$. In contrast, the rear direction (180°, 0°), Mdn $= 96: Z = 45, p = 0.06$, and the above direction (0°, 90°), Mdn $= 95: Z = 49, p = 0.3$, were not statistically significantly different at a confidence interval of 95%. For the speech stimulus, the perceived similarity was generally higher than for pink noise, and only for the direction (45°, 30°) was a significant difference observed between C1, Mdn $= 100$, and C2, Mdn $= 91: Z = 78, p = 0.01$.

The Mysphere 3.2 headphones introduced more coloration. It was expected that there would be a difference between the open and closed frame position, which was only true for some stimulus and direction combinations. For pink noise at the lateral direction (90°, 0°), a significant difference was found between C3, Mdn $= 55$, and C4, Mdn $= 34: Z = 114.5, p = 0.01$. For the speech signal, the lateral direction and (45°, 30°) showed significant differences. This is likely explained by the closed frames blocking high-frequency sound arriving from the sides (see again Fig. 1). However, at the frontal direction, the perceived coloration was comparable between the two configurations, and the difference between C3, Mdn $= 46$, and C4, Mdn $= 48$, was small for the pink noise stimulus: $Z = 48.5, p = 1$. For the

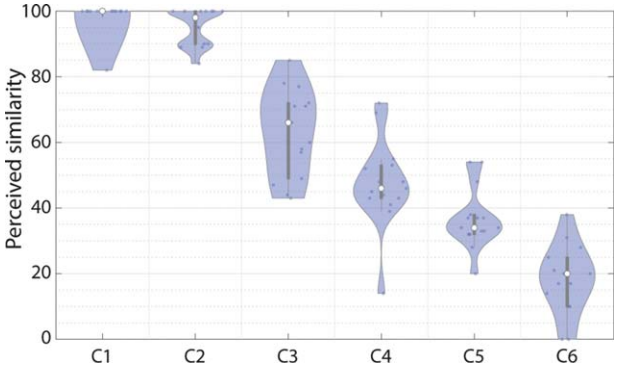



Fig. 3. Violin plots of the coloration listening test for all tested head-worn device configurations using the diffuse rain stimulus. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

speech stimulus at the front, a slightly stronger impairment was observed for C3, Mdn = 64, compared with C4, Mdn = 74, though this was not statistically significant: $Z = 31, p = 1$.

Lastly, the results of C5 were compared with C4. The impairment caused by C5 was significantly larger at the above direction (0°, 90°), Mdn = 30, compared with C4, Mdn = 50: $Z = 120, p = 0.02$, for pink noise. This was significant for the speech stimulus as well. However, no significant differences were observed between the two configurations in the lateral direction. For the frontal direction and speech stimulus, C5, Mdn = 88, even performed significantly better than C4, Mdn = 74: $Z = 0, p = 0.005$. However, this effect was not found in the case of the noise stimulus.

The HD650 was the most colored, indicating that circumaural "open" headphones still produce strong coloration of external sounds. It is worth noting that for the above and rear directions with the speech stimulus, the differences between the modified K702 and HD650 were the smallest.

A more integral view is offered by the results of the diffuse rain simulation, in which the diffuse nature of the stimulus was intended to weigh all directions evenly. A Friedman test showed statistical significant differences between the conditions: $\chi^2(5) = 72.55$, $p < 0.001$. Again, Wilcoxon signed rank tests using the Holm correction ($m = 7$) were used as a post-hoc test. Pairwise comparisons between open ears and the HWD conditions showed significant effects with respect to the reference, except for the difference between C1, Mdn = 100, and C2, Mdn = 98: $Z = 36, p = 0.12$. In this experiment, the difference between C3, Mdn = 66, and C4, Mdn = 46: $Z = 117, p < 0.001$, was significant. Furthermore, C5, Mdn = 46, performed worse than C4, Mdn = 34: $Z = 102, p = 0.028$. This result leads to a clear ranking of the devices.

### 3.3 HWD-Induced Coloration

As expected, the tested HWD configurations produced varying levels of coloration, which changed with stimulus and sound-source direction. The pink noise stimulus produced more critical results than speech, with overall lower levels of perceived similarity observed. This is likely because of the fact that pink noise has energy at all frequencies and higher energy levels at high frequencies. The diffuse rain stimulus showed a general view of the induced coloration of the HWD configurations, with C2 producing the least coloration and C6 producing the most.

It appears the perceived coloration could be predicted from visual observation of the design and fit of the HWDs on the head (see again Fig. 1). In general, C2 produced a low amount of coloration, but it was the strongest for (0°, 0°) and slightly increased at (45°, 30°) and (90°, 0°). This is as expected, considering the HWD shape protrudes from the front. In the case of the Mysphere 3.2, the difference between open and closed configuration was large for the lateral directions, and small for the frontal direction, which is explained by how the different positions of the frames alter which directions are most blocked. Furthermore, for the modified K702, the frontal impairment was similar to that of the Mysphere 3.2, but sounds from above were more impaired. Observing the design offers a possible explanation: the pads are cut away at the front but not above and below. The HD650 headphones produced the greatest impairments, which is as expected considering they cover the ears the most.

### 3.4 Model-Based Prediction

To numerically assess the device-induced effects on coloration of the different HWD configurations, the difference between the HRTF measurements detailed in Sec. 2 of the dummy head with open ears and of the dummy head wearing the HWDs was calculated using the CLL model [22]. The MATLAB implementation of the model was used,[2] which utilizes equivalent rectangular bandwidth weightings to account for linear Fast Fourier Transform frequency sampling and a Phon calculation for perceptual loudness.

For a binaural input signal, the left and right signals are first processed into 42 equivalent rectangular bandwidth frequency bands [41]. The frequency band signals are then rectified, whereby any negative values are set to 0 and low-pass filtered at a cutoff frequency of 800 Hz. The loudness values for each band are calculated [22] as

$$L = \sqrt[4]{\frac{1}{N}\sum_{n} x^2(n)} \quad , \tag{1}$$

where $L$ denotes the loudness value, $x$ the rectified frequency band signal, and $N$ the total number of samples in the signal. This is repeated for the 42 frequency bands and for both left and right signals. The CLL of each frequency band is then calculated in Phons [22] as

$$\mathrm{CLL} = 10\log_2(L_l + L_r) + 40 \quad , \tag{2}$$

where $l$ and $r$ denote left and right, respectively. In this study, the CLL difference between the open ears and each tested HWD configuration for each frequency band was calculated as

$$\Delta\mathrm{CLL}_{\mathrm{CX}}(\theta, \phi) = \mathrm{CLL}_{\mathrm{C1}}(\theta, \phi) - \mathrm{CLL}_{\mathrm{CX}}(\theta, \phi) \quad , \tag{3}$$

where C1 refers to the open ears condition and CX refers to the studied HWD configuration (see again Sec. 2), and a single value of ΔCLL was calculated as the mean of the 42 ΔCLL values for each frequency band. Projected maps of the predicted coloration are presented in Fig. 4, including a mean average of the CLL calculations for all measurement directions ($\overline{\Delta\mathrm{CLL}}$).

### 3.5 Accuracy of Coloration Prediction

The numerical results in Fig. 4 appear to be closely related to the perceptual results in Fig. 2. Upon first look, they show that the predicted coloration is greatest for C6 and least for C2. They also represent the stronger coloration of the lateral angles when comparing C3 and C4, which was observed in the listening test.

To analyze the relationship between the perceived similarity and numerical coloration results in greater detail,

[2] www.acoustics.hut.fi/-ville/software/auditorymodel/.

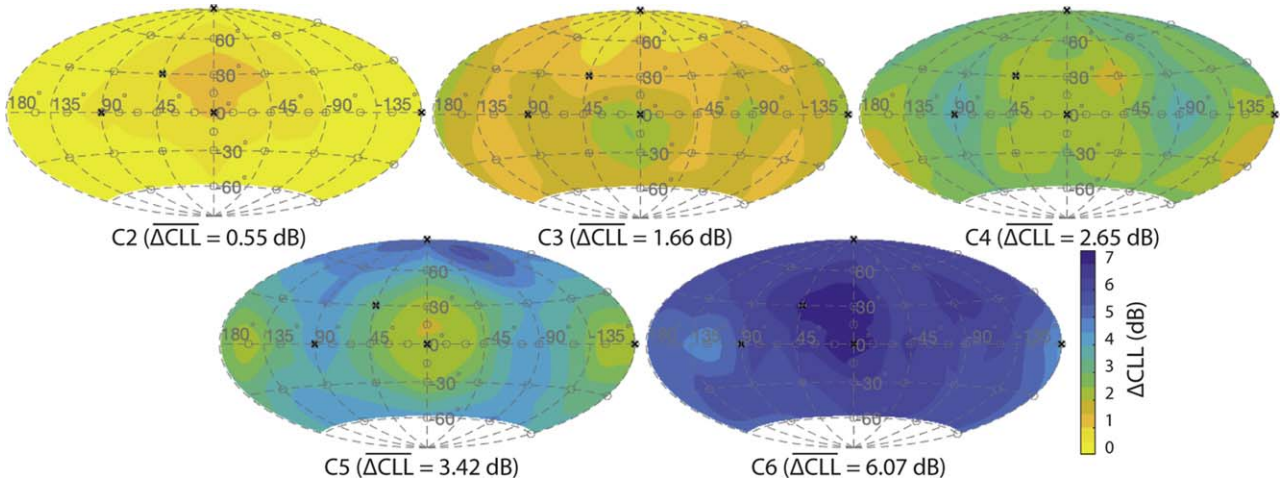



Fig. 4. Hammer-Aitoff projection plots of the predicted coloration between different head-worn-device configurations and open ears (no head worn device), calculated as the difference in composite loudness level (ΔCLL). Circles indicate the positions of the measurement loudspeakers; positions used in the listening test are marked with an x. C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

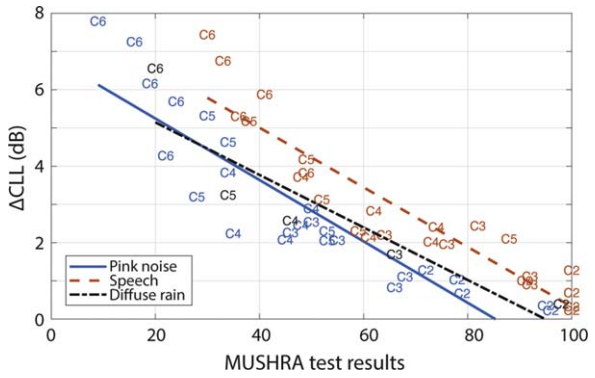



Fig. 5. Comparing the median Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test results on perceived coloration to the difference in composite loudness level (ΔCLL) calculations between the reference and test stimuli. The straight lines denote linear regressions. C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

Table 1. Pearson's correlation coefficients and linear regression coefficients for $y = ax + b$, comparing the median Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test results to the difference in composite loudness level calculations for the three tested stimuli types. Model inputs were the references and the test signals (top) and measured head-related transfer functions (HRTFs) (bottom).

| | | Correlation | | Regression | |
|---|---|---|---|---|---|
| Stimuli type | Model input | $r$ | $p$ | $a$ | $b$ |
| Pink noise | Signal | −0.902 | < 0.001 | 6.85 | −0.0803 |
| Speech | Signal | −0.927 | < 0.001 | 8.13 | −0.0780 |
| Diffuse rain | Signal | −0.903 | 0.036 | 6.52 | −0.0687 |
| Pink noise | HRTFs | −0.903 | < 0.001 | 6.89 | −0.0792 |
| Speech | HRTFs | −0.912 | < 0.001 | 8.55 | −0.0811 |
| Diffuse rain | HRTFs | −0.926 | 0.024 | 4.91 | −0.0504 |

Fig. 5 presents the correlation of the median MUSHRA test results to the ΔCLL calculations between the test and reference stimuli, with lines to denote linear regressions, for all three stimuli types. Table 1 presents the Pearson's correlation coefficients and linear regression coefficients for the three tested stimuli types. The values presented were calculated both using the convolved signals used in the listening test (top) and simply using the measured HRTFs from Sec. 2 (bottom). Similar results were found using the McKenzie et al. binaural coloration model [27], so CLL is preferred in this paper because of its simplicity.

The high correlation values further reinforce the hypothesis that the coloration model can successfully predict the perceptual transparency of the tested HWDs and suggest that this relationship would still exist for other HWDs. Furthermore, the predictions of the diffuse stimuli suggest that it would generalize reasonably well to non-anechoic conditions and BRIRs.

The regression lines for the different stimuli types have very similar slopes, as shown by the similar values of $b$, with the main difference being the intercept $a$. This simple shift between the results for the pink noise and speech stimuli is evident in the listening test results and can be observed in Fig. 4, because the results for the speech and pink noise stimuli were highly comparable, except for the lower overall similarity levels observed for the pink noise stimuli. This suggests that it may be possible to approximately predict the results for speech from the results obtained for noise. For modeling, this observation further suggests that signal dependency is not needed to rank devices in terms of the coloration they introduce. Although a linear fit is presented in Fig. 5, the distribution of data suggests that a curve may be a more appropriate fit, which tends toward $\Delta\text{CLL} = 0$ for MUSHRA test results at 100.

This theory is further supported by the correlation and regression values presented in Table 1 that compare the perceptual results to ΔCLL values obtained using just the measured HRTFs. The correlation with the listening test data is equally as high as with the ΔCLL values obtained from the signals used in the listening test, and the linear regression coefficients are highly comparable.

Note that in this study, only static binaural rendering was considered, which is one particular case of the evaluation of HWD-induced coloration. The static experiment allowed for analyzing each direction separately so that the results could be compared to the predictions of the auditory model. Dynamic models would be required for predicting coloration introduced by HWDs in more realistic scenarios and should be studied in the future. However, because additional consideration to interpolation artefacts would be required, dynamic conditions for coloration were considered beyond the scope of this work.

## 4 PREDICTION OF HWD-INDUCED IMPAIRMENT ON LOCALIZATION

The effect of HWDs on localization was first evaluated perceptually through listening tests and then numerically through model-based evaluations using the measurements from Sec. 2. Finally, the perceptual and numerical results were compared to evaluate the applicability of replacing listening tests with numerical evaluation and measurements.

### 4.1 Listening Test Methodologies

Two listening tests were conducted to assess the effect of HWDs on localization. Both listening tests took place in the multichannel anechoic chamber "Wilska" at Aalto University Acoustics Lab, Finland. The first listening test assessed the effect of HWDs on horizontal plane localization. Eighteen Genelec 8331A loudspeakers were located on the horizontal plane at azimuth angles $\theta = 0°$, $\pm 15°$, $\pm 30°$, $\pm 45°$, $\pm 60°$, $\pm 75°$, $\pm 90°$, $\pm 120°$, $\pm 150°$, and $180°$. The density of loudspeakers was higher in the frontal plane to assess the perceived effect of HWDs on interaural cues with greater precision.

The second listening test assessed the effect of HWDs on median plane localization. Fourteen Genelec 8331A loudspeakers were located on the median plane at polar angles $\phi' = -60°$, $-30°$, $-15°$, $0°$, $+15°$, $+30°$, $+45°$, $+60°$, $+90°$, $+120°$, $+150°$, $+180°$, $+210°$, and $+240°$. The loudspeaker density was higher in the frontal plane, which is a region where human localization is more accurate [42]. Note that the interaural polar coordinates system [43] is used here to determine angles in the median plane, as in [30]. In this system, the range of the lateral angle is $-90° < \theta' < 90°$ in contrast to the range of the azimuth angles, $-180° < \theta < 180°$, and the polar range is within $-90° < \phi' < 270°$ as opposed to the elevation angle, which ranges from $-90° < \phi < 90°$. Thus, a region on the sphere that shares the same interaural time differences and interaural level differences (cone of confusion) also shares the same lateral angle.

For both listening tests, participants sat on a chair with a fixed rotation in the center of the loudspeaker array. The chair was fitted with an adjustable headrest in order to position the head consistently among trials, and the height of the chair was set for every participant to ensure that their ears were aligned with the center of the loudspeaker array. The task was to identify the sound source. Loudspeakers were numbered, and perceived direction was measured by selecting the closest loudspeaker number. A tablet computer was used to record responses. This loudspeaker identification method has been used in the past to measure the effect of HWDs on localization [17], which appears to be a reasonable task to assess localization impairments with the AR application scenario in mind.

The stimulus was a 250-ms pink noise burst of 65-dB sound pressure level A-weighted at the listeners position, windowed by onset and offset half-Hanning ramps of 5 ms. The duration of the stimulus was selected to prevent the subjects from using dynamic cues [44]. At the start of each trial, participants were instructed to face the loudspeaker at (0°, 0°). After a 1-s delay, the stimulus was then played back from a randomized loudspeaker. After the sound finished, the participant's task was to report the perceived sound source among the numbered loudspeakers. At this point, head movements were allowed in order to read the loudspeaker numbers. The stimulus was only played once.

Before the start of the tests, participants first familiarized themselves with the task and user interface in a training round. They could run as many trials as they wanted in the open ears condition to make sure that they understood the test and how to use the user interface.

Both listening tests consisted of six rounds, one for each studied HWD configuration. In the horizontal plane test, each round consisted of three repetitions for each loudspeaker, resulting in 54 trials per round. Fifteen participants (average age: 29.6 years, three female, and 12 male) took part in the experiment, with self-reported normal hearing and prior critical listening experience (such as education or employment in audio or music engineering).

In the median plane listening test, the procedure was the same as in the horizontal plane listening test but with different test sound directions, resulting in 42 trials for each round. Again, 15 participants (average age: 28.8 years, two female, and 13 male) took part in the experiment and also had normal hearing and prior experience.

### 4.2 Listening Test Results

The results of the horizontal plane localization test are presented in Fig. 6. The dot plots show the responses for all participants together, and front-back confusions are plotted in red. The collected data was analyzed to assess the introduced effect on (a) front-back confusion rate (FB%) and (b) front azimuth error (FAE).

The front-back confusions were computed using the loudspeakers evenly distributed on the horizontal plane: $\theta = 0°$, $\pm 30°$, $\pm 60°$, $\pm 120°$, $\pm 150°$, and $180°$. If a sound was emitted from a frontal hemiplane loudspeaker but perceived from the rear, or vice versa, the response was considered a

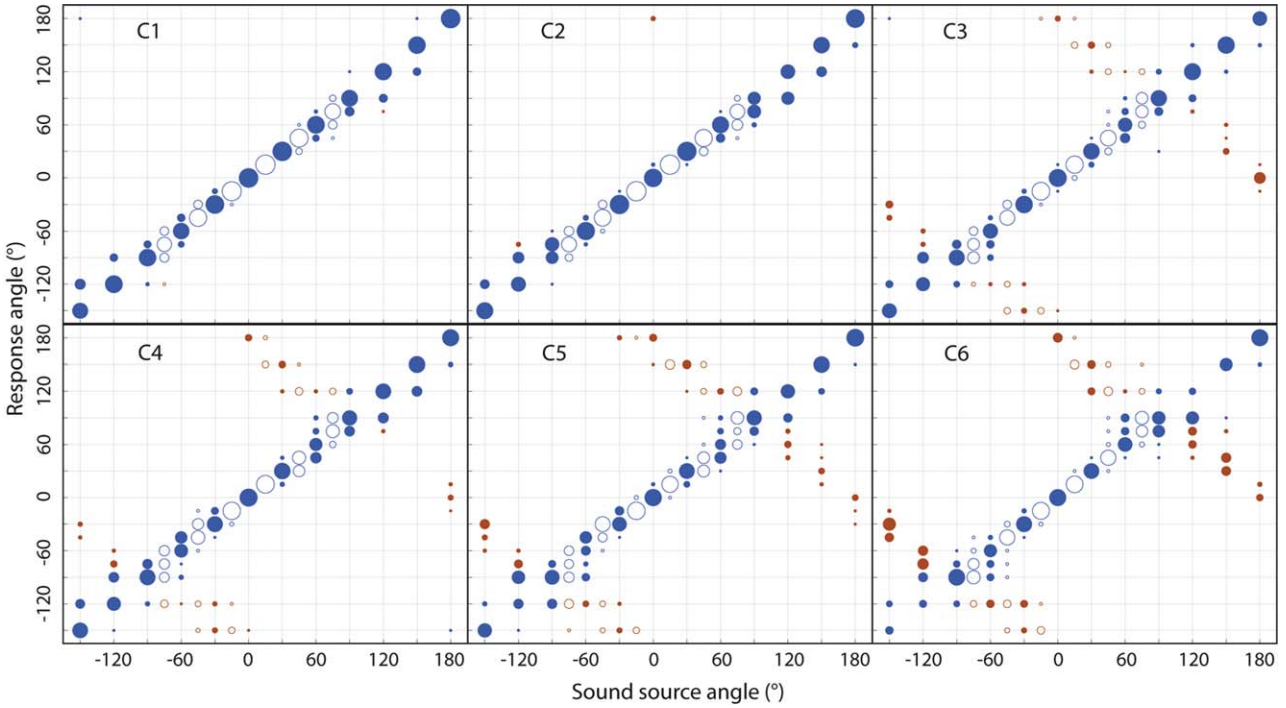



Fig. 6. Listening test responses for the horizontal plane localization experiment. The size of the dots indicates the amount of responses for that particular point in the plot. The angles used to compute the front-back confusion rate (FB%) are presented in filled dots. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

Table 2. Results overview of the localization listening tests regarding front-back confusion rate (FB%), front azimuth error (FAE), quadrant error rate (QE%), and front polar error (FPE).

| | Horizontal plane | | Median plane | |
|---|---|---|---|---|
| | FB% Mean (median) | FAE RMS | QE% Mean (median) | FPE RMS |
| C1 | 0.22 (0) | 4.69° | 4.65 (0) | 10.47° |
| C2 | 1.33 (0) | 5.19° | 5.45 (3.03) | 12.19° |
| C3 | 14.89 (0) | 6.02° | 16.77 (15.15) | 28.03° |
| C4 | 10.44 (0) | 7.38° | 21.62 (18.18) | 30.44° |
| C5 | 20.89 (0) | 9.33° | 19.19 (18.18) | 38.30° |
| C6 | 35.33 (13.3) | 4.60° | 23.23 (24.24) | 37.59° |

C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

front-back confusion. For each participant and HWD configuration, the FB% was computed dividing the number of front-back confusions by the number of trials of the analyzed loudspeakers. Table 2 shows the mean and median confusion rate between participants.

The FAE was assessed by analyzing the responses when the sound source was in the frontal horizontal hemiplane: $\theta = 0°$, $\pm 15°$, $\pm 30°$, $\pm 45°$, $\pm 60°$, $\pm 75°$, and $\pm 90°$. The FAE was computed as the RMS azimuth error of all the trials with sounds from the frontal horizontal hemiplane, after discarding the front-back confusions.

The results of the median plane localization test are presented in Fig. 7 as dot plots, which show the responses for all participants together. Quadrant errors are defined as any angular absolute error in the median plane greater than 90° and are plotted in red. An overview of the median plane localization results is presented in Table 2. For each participant and listening condition, the quadrant error rate (QE%) was computed by dividing the number of quadrant errors by the number of trials. To characterize each condition, a QE% was computed by averaging the results over all the participants. For the quadrant errors analysis, only the evenly distributed loudspeakers on the median plane were used: $\phi' = -60°, -30°, 0°, +30°, +60°, +90°, +120°, +150°, +180°, +210°$, and $+240°$. To analyze the front polar error (FPE), only the loudspeakers in the frontal median plane with a 15° spacing were used: $\phi' = -30°, -15°, 0°, +15°, +30°, +45°$, and $+60°$. The FPE for each condition was computed as the RMS error of all the trials in that given condition, excluding quadrant errors.

The induced FAE was quite contained (FAE $< 10°$). However, the differences in front-back confusions were much larger among conditions. The confusion rate in C1 was very low (FB% $< 1\%$). A Friedman test was conducted to compare the effect of each HWD on front-back confusions between participants. The effect of the HWDs reached statistical significance ($\chi^2(5) = 22.22, p < 0.001$). For the horizontal localization experiment, Wilcoxon signed rank tests were conducted to test the hypotheses (i), (ii), and (iii), as stated in Sec. 2. Only the difference between the open ears and HD650 was found to be significant.

In the median plane localization test, the FPE varied from 10.47° in C1 to 38.30° in C5. These errors show how

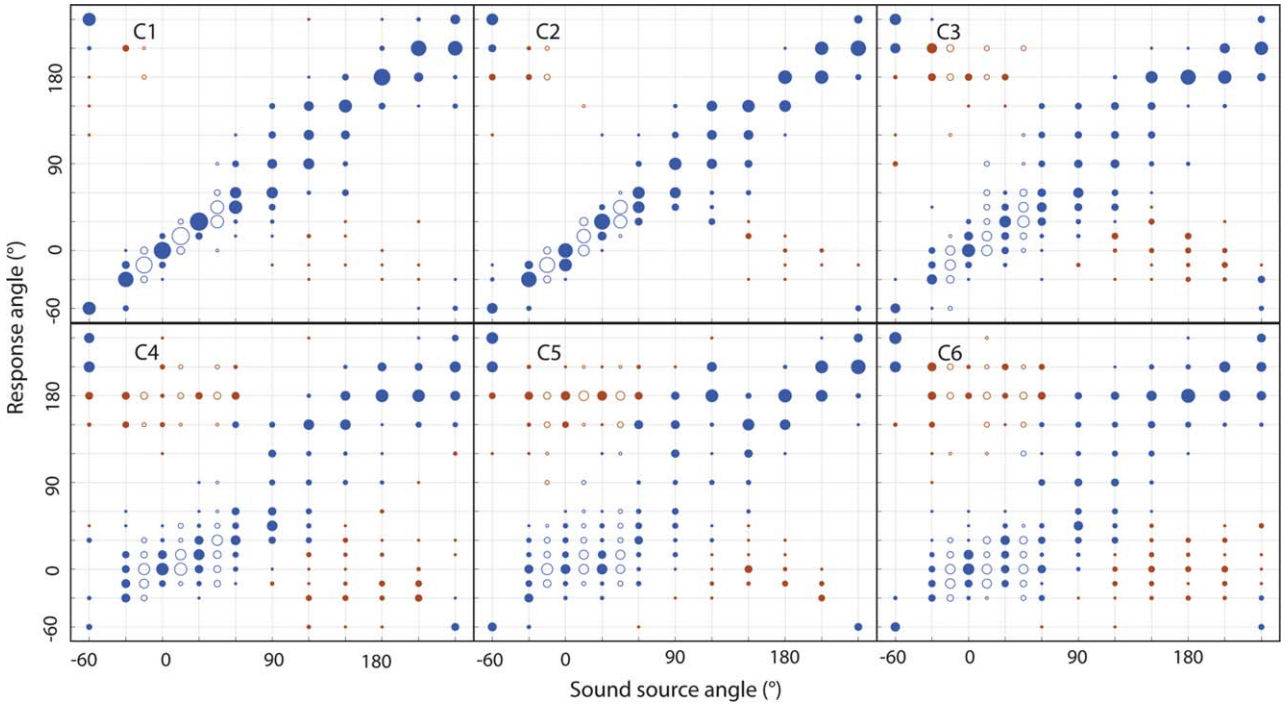



Fig. 7. Listening test responses for the median plane localization experiment. The size of the dots indicates the amount of responses for that particular point in the plot. The angles used to compute the quadrant error rate (QE%) are presented in filled dots. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

the HWDs impair localization ability in the median plane. However, the perceptual implications of an increased QE% (percentage of errors larger than 90°) is more critical [45]. The QE% in C1 was lowest (QE% < 5%). A Friedman test showed a significant effect between the results, $\chi^2(5) = 45.52, p < 0.001$, and Wilcoxon signed rank tests revealed that the differences between the open-ear condition and all other conditions, except for C2, were significant ($p < 0.01$). However, neither differences between C3 and C4 nor differences between C4 and C5 were statistically significant.

### 4.3 HWD-Induced Impairment on Localization

When examining the results, the FAE was relatively low, showing that front-back confusions were the predominant problem in the horizontal test, as previously reported in [6, 17], which may be an important factor when choosing an HWD for AR applications. In the median plane test, an increase in QE% and FPEs was caused by the studied HWDs. These results are also in accordance with previous studies that tested more occlusive HWDs [16–18] and with the observation of [6, 10] that tested AR HWDs.

It is important to note that the tests were not specifically designed to measure fine FAEs or FPEs. The loudspeaker density was more sparse than the just-noticeable differences (especially in the horizontal plane), but it is an approximation of a real environment where multiple possible sources are present. In the front horizontal plane, the results of a setup with hidden loudspeakers and denser response grid can be found in [46], showing that the introduced effect is rather small. On the other hand, the FPE results show that even in C1 the subjects had problems finding the correct loudspeaker. It is expected that the results in a denser loudspeaker array setup would be similar in this case. Nonetheless, the conducted perceptual tests are not a substitute for more traditional localization tests methodologies (i.e., measuring just-noticeable differences).

When comparing HWDs in both tests, C2 produced similar localization errors to C1, and therefore, it cannot be considered to affect localization. The Mysphere 3.2 caused an increase in front-back confusions and quadrant errors, in which C3 induced more errors than C4 for both. The modified AKG K702 induced front-back confusions were higher than for the Mysphere 3.2 but performed better than the Mysphere 3.2 in C4 in median plane quadrant error. The Sennheiser HD650 performed worse than the other devices in both localization tests. The FPE followed a similar trend as front-back confusions and quadrant errors, which was expected because of the importance of monaural cues to the polar dimension.

Even though a clear trend was observed of how these devices affect FB% and QE%, it is important to note that the introduced effect was highly individual. The studied open headphones, C3, C4, and C5, only affected some participants, generating a great number of confusions and errors, whereas others performed as well as in the open ears condition. For both types of confusion, this can be explained by the fact that quadrant errors and front-back confusions are affected by distorted spectral cues, which are known to be highly individual [47].

The differences in individual performance are revealed by the mean and median confusion rates over all partici-

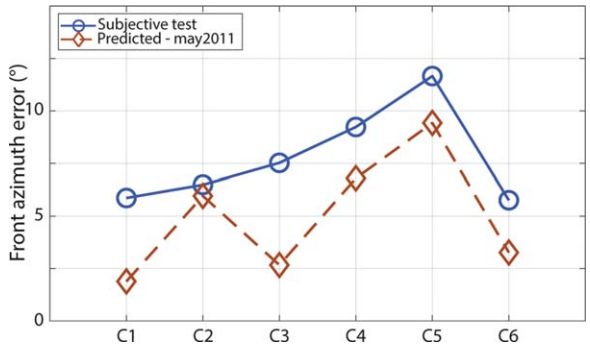



Fig. 8. Front azimuth error (degrees) comparison between the horizontal plane listening test and auditory model output for the studied conditions. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

pants, as shown in Table 2. The median is zero in many conditions, indicating that at least half of the participants had no front-back confusions at all, whereas others had many. Despite these inter-subject differences, the mean confusion rates over participants was used for further analysis in this study. This value was considered to represent how problematic a certain HWD may be for a group of participants. In an attempt to explain these individual differences, two possible alternatives seemed equally valid. In the first case, these individual differences could be explained by acoustic factors only. For this to be the case, the introduced acoustic effect could be explained by measurements at each participant's eardrum, strongly modifying the acoustic properties to some subjects but not the others.

Another alternative would be that these acoustic factors are not enough to explain the inter-subject differences. This could be explained by the importance of non-acoustic factors, such as individual sensitivity [48], or by the fact that different subjects may use different frequency ranges of the spectral cues to resolve the location of the source [49]. Further studies are needed to estimate the individual effect of a particular HWD on a particular participant, which could involve individual HRTF measurements and more detailed perceptual studies.

### 4.4 Model-Based Prediction

For the FAE, May et al.'s model [28] was evaluated to predict the mean of the subjective collected data using the KEMAR HRTF measurements. In previous studies, this model was shown to be robust against reverberation and multiple simultaneous sounds, and it may be appropriate for predicting the HWD's effect. Fig. 8 compares the predicted FAE to the perceived FAE from the horizontal plane listening test.

Baumgartner et al.'s [30] sagittal plane localization model was used to predict performance in the localization tests related to the polar dimension (front-back confusions, quadrant errors, and polar errors). Instead of trying to predict individual results (which is the original goal of the model in [30]), the model here was used to predict the mean of the collected subjective data using the HRTF measurements detailed in Sec. 2.

In this study, it is assumed that there was no adaptation to the monaural cues during the listening test because the duration of sound exposure was short, no feedback was given, and no specific training was conducted [50]. For that reason, in this template-based model (explained in Sec. 1), the template was the measurement set of HRTFs with the KEMAR with C1. For each evaluated condition, the target HRTFs corresponded to the measurements of the tested HWD configurations from Sec. 2.

The FB% estimation was conducted to assess whether it could mimic the horizontal plane listening test results. The model parameters (i) degree of selectivity and (ii) motoric response scatter were set to default ($\Gamma = 6$ dB$^{-1}$; $\varepsilon = 17°$). The (iii) listener-specific sensitivity (S) was optimized for the open ears condition, treating the KEMAR as a specific participant, following the procedure described in [30]. The model was optimized under the open ears condition only to be consistent with the assumption of no adaptation during the listening test. The performance of the model showed a monotonic function, such that error decreased as the S was lowered, thus the lowest value in [30] was selected ($S = 0.21$).

In the description of Baumgartner et al.'s model [30], directional transfer functions (DTFs) are used to focus on the direction-dependent cues only. However, it is expected that in the open ears condition, DTFs and HRTFs may provide very similar results. In this study, it was considered appropriate to use HRTFs because HWDs may introduce directional-independent distortions that may contribute to errors in localization, which has also been used in the past [32].

Even though it is possible to input the stimulus to the auditory model, using pink noise did not allow the model to achieve errors of the order of the perceptual data, so impulse responses were used to get the best possible performance of the model. Moreover, the described default parameters were computed for human participant DTFs, and the model parameters were optimized so that the model predicted quadrant and polar errors were as close as possible to the median plane listening test data in the open ears condition ($\Gamma = 17$ dB$^{-1}$; $\varepsilon = 27°$; $S = 0.35$). The estimated FB% for each tested condition compared to the horizontal plane listening test data are shown in Fig. 9.

A comparison between QE% in the median plane localization test and those predicted by the model is presented in Fig. 10. Again, the model was evaluated using both the default parameterization and optimization for the open ears condition. Note that the optimized parameters are the same as for the front-back confusions prediction. Baumgartner et al.'s model [30] was used again with the same parameterizations to predict FPE (see Fig. 11).

### 4.5 Accuracy of Localization Impairment Predictions

The results suggest that the presented auditory model approach is appropriate for assessing the most important ef-

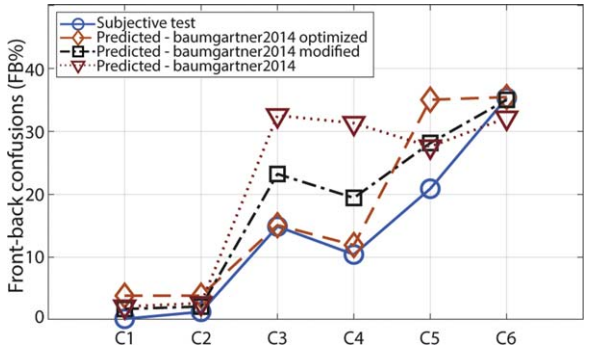



Fig. 9. Front-back confusion rate (FB%) comparison between the listening test and auditory model output for the studied conditions. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

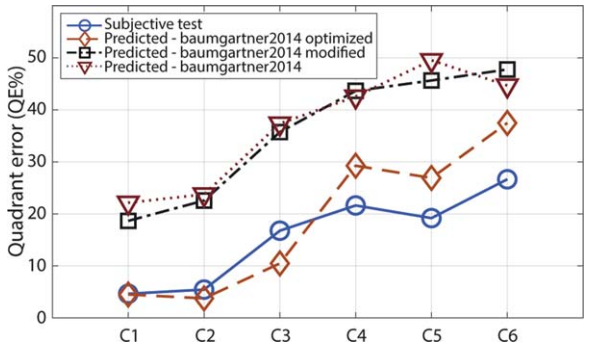



Fig. 10. Quadrant error rate (QE%) comparison between the listening test and auditory model output for the studied conditions. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

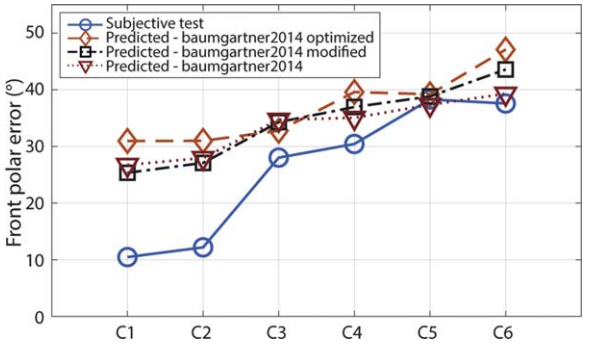



Fig. 11. Front polar error (degrees) comparison between the median plane listening test and auditory model output for the studied conditions. C1, open ear; C2, Oculus Quest 2; C3, Mysphere 3.2 with open frame position; C4, Mysphere 3.2 with closed frame position; C5, AKG K702 (with transparency modification [9]); C6, Sennheiser HD650.

Table 3. Pearson's correlation coefficients comparing the mean localization test results to the auditory model prediction (*). The modified version uses head-related transfer functions instead of directional transfer functions, and an impulse is used as input instead of the actual subjective test stimulus (**). The optimized version is as (*) but with optimized model parameters Γ and ε as described in [30]. Detailed description of (*) and (**) can be found in Sec. 4.4.

| | Auditory model | Correlation | |
|---|---|---|---|
| | | *r* | *p* |
| FAE | [28] | 0.978 | <0.001 |
| FB% | [30] optimized (**) | 0.930 | 0.007 |
| | [30] modified (*) | 0.952 | 0.003 |
| | [30] | 0.768 | 0.074 |
| QE% | [30] optimized (**) | 0.940 | 0.005 |
| | [30] modified (*) | 0.973 | 0.001 |
| | [30] | 0.915 | 0.011 |
| FPE | [30] optimized (**) | 0.834 | 0.039 |
| | [30] modified (*) | 0.969 | 0.001 |
| | [30] | 0.989 | <0.001 |

FAE, front azimuth error; FB%, front-back confusion rate; FPE, front polar error; QE%, quadrant error rate.

fects of HWDs on localization tasks. May et al.'s model [28] and Baumgartner et al.'s model [30] were able to predict subjective data accurately. This indicates that the principles underlying the models are applicable to different scenarios, showing good performance even for new listening conditions that were not initially taken into account in the design of the models. The prediction by May et al.'s model [28] overestimated participants' accuracy for C1 and underestimated it for C5. However, it seemed to generalize the global trend well (see Table 2).

The results of Baumgartner et al.'s model [30] have shown that the predicted values for FB% and QE% are related to the perceptual results (see Table 3). The model results show that it is possible to establish the relative differences caused by the HWDs using the default parameters defined in [30]. Using HRTFs instead of DTFs improves the prediction of relative errors among HWDs, which could be explained by the non–directional-dependent distortions introduced by the HWDs. The default parameters failed at estimating the absolute values of QE%, which may be because of the use of non-individualized HRTFs and average responses of a group of subjects, instead of separately for each individual. The predicted absolute values for QE% estimation improved when the model parameters (Γ, ε, and S) were optimized, which also improved the performance for front-back confusions. It is noteworthy to reiterate that the parameters of the model were optimized for C1 only, so the localization impairments while wearing HWDs (C2–C6) were always predicted in terms of differences to C1.

The correlation between the output of the models and subjective data is shown in Table 3. The high correlation values show that the used models performed well with predicting subjective data. However, there are other approaches, such as data-driven methods, that could perform better if higher correlation is sought. The presented approach, because of its physiological nature, may be more appropriate for generalizing new listening conditions. This is supported by the presented results, because conditions

including HWDs were never included in the model design. The model predictions and optimization used either the subjective test stimulus or an impulse signal. Instead, multiple randomized input signals may further improve generalization of the model predictions.

To estimate the effect of HWDs in non-anechoic environments, BRIRs should be used instead of HRTFs as input to the models. May et al.'s model [28] is expected to maintain its performance because it has been shown to be robust under reverberant conditions. Baumgartner et al.'s model [30] has not yet been tested under reverberant conditions to the knowledge of the authors, though it is expected that some modifications in the structure of the model may be necessary. These modifications could include some time-dependent stages, such as a more complex model of the basilar membrane or dynamic comparison between the target and template.

As in the coloration evaluation, subjective data on dynamic conditions was not collected. Future work should extend the evaluation in order to understand the effect caused by HWDs when the source or the listener can move. However, when trying to predict the HWDs effects in such conditions, the models should include a time-dependent stage for dynamic cue processing.

## 5 CONCLUSION AND FUTURE WORK

This paper has investigated the feasibility of using pre-existing auditory models to predict the perceptual transparency of HWDs, which is an important quality measure for auditory AR applications in which real-world sounds must not be degraded. Firstly, perceptual tests assessing coloration and localization error showed that the tested HWDs produce varying but significant impairments. In terms of coloration, non-negligible effects were produced by all the studied HWDs. These varied based on sound-source direction and stimuli type. In terms of localization, the main impairing effect in the horizontal plane was an increase in FB%, which changed depending on the HWD. However, median plane localization was highly affected depending on the HWD, most notably through an increase in QE%.

Auditory models were then evaluated to predict the perceptual effects introduced by the studied HWDs, using HRTF measurements of a dummy head wearing different configurations of the HWDs. The models proved accurate at predicting the perceived coloration (CLL), azimuth localization [28], and front-back confusions and median plane quadrant errors [30] produced by wearing the tested HWDs, which is an application they were not initially designed for. Therefore, the results presented in this paper reinforce the principles of physiologically accurate auditory models.

Although this paper has presented methods for predicting the perceived transparency of HWDs, the accuracy of binaural reproduction of virtual sounds is also an important factor. Therefore, future work on assessing the suitability of HWDs for AR applications should look into transfer-plausibility tests, in which real and virtual sounds are presented simultaneously. This will evaluate not only the transparency of the HWDs but also the quality of complete rendering systems. Measuring coloration separately for left and right ears could give an insight into the interaural contributions to coloration. Furthermore, externalization was not analyzed in this study, which could also be related to the HWD introduced effect on monaural cues distortion, because several participants reported a lack of externalization in some of the studied conditions.

The effect of the devices in more realistic tasks should also be analyzed in the future, such as the HWD's effect on localization when dynamic and visual cues are available or in circumstances where real and virtual sources are presented together. However, it is expected that localization will improve when dynamic cues are available, especially regarding sagittal plane localization.[3]

## 6 ACKNOWLEDGMENT

This research was supported by the Human Optimised XR Project and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 812719.

## 7 REFERENCES

[1] A. Lindau and S. Weinzierl, "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acust. united Acust.*, vol. 98, no. 5, pp. 804–810 (2012 Sep.). https://doi.org/10.3813/AAA.918562.

[2] A. Neidhardt and A. M. Zerlik, "The Availability of a Hidden Real Reference Affects the Plausibility of Position-Dynamic Auditory AR," *Front. Virtual Real.*, vol. 2, paper 678875 (2021 Sep.). https://doi.org/10.3389/frvir.2021.678875.

[3] S. A. Wirler, N. Meyer-Kahlen, and S. J. Schlecht, "Towards Transfer-Plausibility for Evaluating Mixed Reality Audio in Complex Scenes," in *Proceedings of the AES Conference on Audio for Virtual and Augmented Reality* (2020 Aug.), paper 3-4.

[4] C. Schneiderwind, A. Neidhardt, and D. Meyer, "Comparing the Effect of Different Open Headphone Models on the Perception of a Real Sound Source," presented at the *150th Convention of the Audio Engineering Society* (2021 May), paper 10489.

[5] 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 *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P3-9.

[6] D. Satongar, C. Pike, Y. W. Lam, A. I. Tew, "The Influence of Headphones on the Localization of External Loudspeaker Sources," *J. Audio Eng. Soc.*, vol. 63, no. 10, pp. 799–810 (2015 Oct.). https://doi.org/10.17743/jaes.2015.0072.

[3] Acoustic measurements, anonymized listening test data, and MATLAB code to illustrate the methods used in this paper are available to download on https://github.com/lladopedro/TransparencyEvaluation.

[7] H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, "Acoustic Transparency in Hearables—Perceptual Sound Quality Evaluations," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 495–507 (2020 Jul.). https://doi.org/10.17743/jaes.2020.0045.

[8] C. Pörschmann, J. M. Arend, and R. Gillioz, "How Wearing Headgear Affects Measured Head-Related Transfer Functions," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*, pp. 49–54 (Paris, France) (2019 Sep.). https://doi.org/10.25836/SASP.2019.27.

[9] N. Meyer-Kahlen, D. Rudrich, M. Brandner, et al., "DIY Modifications for Acoustically Transparent Headphones," presented at the *148th Convention of the Audio Engineering Society* (2020 May.), e-Brief 603.

[10] A. Genovese, G. Zalles, G. Reardon, and A. Roginska, "Acoustic Perturbations in HRTFs Measured on Mixed Reality Headsets," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P8-4.

[11] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, "Acoustic Transparency in Hearables—Technical Evaluation," *J. Audio Eng. Soc.*, vol. 68, no. 7/8, pp. 508–521 (2020 Jul.). https://doi.org/10.17743/jaes.2020.0042.

[12] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, vol. 43, no. 4, pp. 203–217 (1995 Apr.).

[13] V. Erbes, F. Schultz, A. Lindau, and S. Weinzierl, "An Extraaural Headphone System for Optimized Binaural Reproduction," in *Proceedings of the Fortschritte der Akustik (DAGA)*, pp. 313–314 (Darmstadt, Germany) (2012 Jul.).

[14] F. Brinkmann, A. Lindau, and S. Weinzierl, "On the Authenticity of Individual Dynamic Binaural Synthesis," *J. Acoust. Soc. Am.*, vol. 142, no. 4, pp. 1784–1795 (2017 Oct.). https://doi.org/10.1121/1.5005606.

[15] E. H. A. Langendijk and A. W. Bronkhorst, "Fidelity of Three-Dimensional-Sound Reproduction Using a Virtual Auditory Display," *J. Acoust. Soc. Am.*, vol. 107, no. 1, pp. 528–537 (2000 Jan.). https://doi.org/10.1121/1.428321.

[16] R. S. Bolia, W. R. D'Angelo, P. J. Mishler, and L. J. Morris, "Effects of Hearing Protectors on Auditory Localization in Azimuth and Elevation," *Hum. Factors*, vol. 43, no. 1, pp. 122–128 (2001 Mar.). https://doi.org/10.1518/001872001775992499.

[17] N. L. Vause and D. W. Grantham, "Effects of Earplugs and Protective Headgear on Auditory Localization Ability in the Horizontal Plane," *Hum. Factors*, vol. 41, no. 2, pp. 282–294 (1999 Jun.). https://doi.org/10.1518/001872099779591213.

[18] V. Zimpfer and D. Sarafian, "Impact of Hearing Protection Devices on Sound Localization Performance," *Front. Neurosci.*, vol. 8, paper 135 (2014 Jun.). https://doi.org/10.3389/fnins.2014.00135.

[19] B. D. Simpson, R. S. Bolia, R. L. McKinley, and D. S. Brungart, "The Impact of Hearing Protection on Sound Localization and Orienting Behavior," *Hum. Factors*, vol. 47, no. 1, pp. 188–198 (2005 Mar.). https://doi.org/10.1518/0018720053653866.

[20] E. Zwicker and U. T. Zwicker, "Dependence of Binaural Loudness Summation on Interaural Level Differences, Spectral Distribution, and Temporal Distribution," *J. Acoust. Soc. Am.*, vol. 89, no. 2, pp. 756–764 (1991 Feb.). https://doi.org/10.1121/1.1894635.

[21] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240 (1997 Apr.).

[22] V. Pulkki, M. Karjalainen, and J. Huopaniemi, "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model," *J. Audio Eng. Soc.*, vol. 47, no. 4, pp. 203–217 (1999 Apr.).

[23] M. Karjalainen, "A Binaural Auditory Model for Sound Quality Measurements and Spatial Hearing Studies," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 985–988 (Atlanta, GA) (1996 May). https://doi.org/10.1109/icassp.1996.543288.

[24] V. Pulkki and M. Karjalainen, "Localization of Amplitude-Panned Virtual Sources. I: Stereophonic Panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 739–752 (2001 Sep.).

[25] K. Ono, V. Pulkki, and M. Karjalainen, "Binaural Modeling of Multiple Sound Source Perception: Methodology and Coloration Experiments," presented at the *111th Convention of the Audio Engineering Society* (2001 Nov.), paper 5446.

[26] K. Ono, V. Pulkki, and M. Karjalainen, "Binaural Modeling of Multiple Sound Source Perception: Coloration of Wideband Sound," presented at the *112th Convention of the Audio Engineering Society* (2002 Apr.), paper 5550.

[27] T. McKenzie, C. Armstrong, L. Ward, D. T. Murphy, and G. Kearney, "Predicting the Colouration Between Binaural Signals," *Appl. Sci.*, vol. 12, no. 5, paper 2441 (2022 Feb.). https://doi.org/10.3390/app12052441.

[28] T. May, S. van de Par, and A. Kohlrausch, "A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 1, pp. 1–13 (2011 Jan.). https://doi.org/10.1109/TASL.2010.2042128.

[29] R. D. Patterson, "The Sound of a Sinusoid: Spectral Models," *J. Acoust. Soc. Am.*, vol. 96, no. 3, pp. 1409–1418 (1994 Aug.). https://doi.org/10.1121/1.410285.

[30] R. Baumgartner, P. Majdak, and B. Laback, "Modeling Sound-Source Localization in Sagittal Planes for Human Listeners," *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 791–802 (2014 Aug.). https://doi.org/10.1121/1.4887447.

[31] R. Baumgartner and P. Majdak, "Modeling Localization of Amplitude-Panned Virtual Sources in Sagittal Planes," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 562–569 (2015 Jul.). https://dx.doi.org/10.17743/jaes.2015.0063.

[32] R. Baumgartner, P. Majdak, and B. Laback, "Modeling the Effects of Sensorineural Hearing Loss on Sound Localization in the Median Plane," *Trends*

*Hear.*, vol. 20, pp. 1–11 (2016 Sep.). https://doi.org/10.1177/2331216516662003.

[33] T. McKenzie, S. J. Schlecht, and V. Pulkki, "Auralisation of the Transition Between Coupled Rooms," in *Proceedings of the International Conference on Immersive and 3D Audio: From Architecture to Automotive*, pp. 1–9 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/I3DA48870.2021.9610955.

[34] N. Meyer-Kahlen and S. J. Schlecht, "Assessing Room Acoustic Self-Localization Using a Virtual Blindfold," in *Proceedings of the Fortschritte der Akustik (DAGA)* , pp. 312–315 (Vienna, Austria) (2021 Aug.).

[35] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), paper 5093.

[36] ITU-R, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," *Recommendation ITU-R BS.1534-3* (2015 Oct.).

[37] M. Schoeffler, S. Bartoschek, F.-R. Stöter, et al., "webMUSHRA—A Comprehensive Framework for Web-Based Listening Tests," *J. Open Res. Softw.*, vol. 6, no. 1, paper 8 (2018 Feb.). http://doi.org/10.5334/jors.187.

[38] T. McKenzie, D. Murphy, and G. Kearney, "Assessing the Authenticity of the KEMAR Mouth Simulator as a Repeatable Speech Source," presented at the *143rd Convention of the Audio Engineering Society* (2017 Oct.), paper 9820.

[39] J. L. Hintze and R. D. Nelson, "Violin Plots: A Box Plot-Density Trace Synergism," *Am. Stat.*, vol. 52, no. 2, pp. 181–184 (1998 May).

[40] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian J. Stat.*, vol. 6, no. 2, pp. 65–70 (1979).

[41] B. C. J. Moore and B. R. Glasberg, "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," *J. Acoust. Soc. Am.*, vol. 74, no. 3, pp. 750–753 (1983 Sep.). https://doi.org/10.1121/1.389861.

[42] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition)* (MIT Press, Cambridge, MA, 1996).

[43] M. Morimoto and H. Aokata, "Localization Cues of Sound Sources in the Upper Hemisphere," *J. Acoust. Soc. Jpn. (E)*, vol. 5, no. 3, pp. 165–173 (1984). https://doi.org/10.1250/ast.5.165.

[44] S. Perrett and W. Noble, "The Effect of Head Rotations on Vertical Plane Sound Localization," *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2325–2332 (1997 Oct.). https://doi.org/10.1121/1.419642.

[45] M. J. Goupell, P. Majdak, and B. Laback, "Median-Plane Sound Localization as a Function of the Number of Spectral Channels Using a Channel Vocoder," *J. Acoust. Soc. Am.*, vol. 127, no. 2, pp. 990–1001 (2010 Feb.). https://doi.org/10.1121/1.3283014.

[46] P. Lladó, P. Hyvärinen, and V. Pulkki, "Auditory Model-Based Estimation of the Effect of Head-Worn Devices on Frontal Horizontal Localisation," *Acta Acust.*, vol. 6, paper 1 (2022 Jan.).

[47] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123 (1993 Jul.). https://doi.org/10.1121/1.407089.

[48] P. Majdak, R. Baumgartner, and B. Laback, "Acoustic and Non-Acoustic Factors in Modeling Listener-Specific Performance of Sagittal-Plane Sound Localization," *Front. Psychol.*, vol. 5, paper 319 (2014 Apr.). https://doi.org/10.3389/fpsyg.2014.00319.

[49] P. X. Zhang and W. M. Hartmann, "On the Ability of Human Listeners to Distinguish Between Front and Back," *Hear. Res.*, vol. 260, no. 1–2, pp. 30–46 (2010 Feb.). https://doi.org/10.1016/j.heares.2009.11.001.

[50] C. Mendonça, "A Review on Auditory Space Adaptations to Altered Head-Related Cues," *Front. Neurosci.*, vol. 8, paper 219 (2014 Jul.). https://doi.org/10.3389/fnins.2014.00219.

## THE AUTHORS


Pedro Lladó


Thomas McKenzie


Nils Meyer-Kahlen


Sebastian J. Schlecht

Pedro Lladó is a Ph.D. student in the Department of Signal Processing and Acoustics at Aalto University, where he studies spatial hearing and psychoacoustics. He received his B.Sc. degree in Audiovisual Systems Engineering in the Polytechnic University of Catalonia, Spain, and completed his M.Sc. degree in Sound and Music Technology in the Pompeu Fabra University, Spain. His main interests are sound perception and auditory modeling.

•

Thomas McKenzie is a post-doctoral researcher in the Department of Signal Processing and Acoustics at Aalto University, where he studies room acoustics and six-degrees-of-freedom spatial audio. He completed a B.Sc. in Music, Multimedia and Electronics at the University of Leeds, UK, in 2013, before completing his M.Sc. in Post-production with Sound Design and Ph.D. in Music Technology at the University of York, UK, in 2015 and 2020, respectively. His research interests include spatial audio and psychoacoustics.

•

Nils Meyer-Kahlen is a doctoral candidate for the Department of Signal Processing and Acoustics at Aalto University in Finland. Before joining the lab in 2019, he studied Electrical Engineering and Audio Engineering at the Technical University and the University of Music and Performing Arts in Graz, Austria. His main interests are the technology and perception of spatial sound. Currently, he studies room acoustic perception in virtual and augmented realities.

•

Sebastian J. Schlecht is a Professor of Practice for Sound in Virtual Reality at the Acoustics Lab, Department of Signal Processing and Acoustics, and at the Media Labs, Department of Art and Media, of Aalto University, Finland. He received a Diploma in Applied Mathematics from the University of Trier, Germany, in 2010 and M.Sc. degree in Digital Music Processing from the School of Electronic Engineering and Computer Science at Queen Mary University of London, UK, in 2011. In 2017, he received a doctoral degree at the International Audio Laboratories Erlangen, Germany, on artificial spatial reverberation and reverberation enhancement systems. From 2012 to 2019, Dr. Schlecht was also an external research and development consultant and lead developer of the 3D Reverb algorithm at the Fraunhofer IIS, Erlangen, Germany.