Audio Engineering Society

Conference Paper

Presented at the International Conference on
Spatial and Immersive Audio
2023 August 23–25, Huddersfield, UK

Do-it-yourself headphones and development platform for augmented reality audio

Alexander Mülleder[1, 3], Michael Romanov[2], Nils Meyer-Kahlen[3], and Franz Zotter[1]

[1] *University of Music and Performing Arts, Graz, Austria*
[2] *Romanov Engineering*
[3] *Acoustics Lab, Dpt. of Information and Communications Engineering, Aalto University, Espoo, Finland*

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

ABSTRACT

We present an open DIY platform for augmented reality audio in the form of ultralight acoustically transparent headphones and a minimalist hardware platform serving as USB audio interface, signal conditioner, and head tracker. We outline the open hardware and electronics designs to permit reproduction, extension, or customization. Furthermore, we show the results of measurements that evaluate passive transparency, variation in the headphone transfer function (HpTF) due to repositioning, as well as tracking latency of the headtracker.
This design has the potential to promote research and development applications in audio augmented reality at low costs and size.

1 Introduction

Audio for Augmented Reality (AR) aims at rendering virtual sound sources that perfectly blend into the real acoustic environment. To achieve this, high quality head-tracked headphones are employed that are as transparent as possible to sounds of the outside world.

Various headphones that may be suitable for studying and experiencing sound for AR have been introduced over the years. For example, the AKG K1000, originally intended as high-quality headphones for music listening, has become a popular device in acoustic research that relies on a mix of virtual and real sound sources, see e.g., [1, 2, 3]. After the manufacturing of the AKG K1000 was discontinued, another commercial pair of headphones with similar aims was introduced, the Mysphere 3[1], which was also explored in experiments already [4]. The drivers of these models hover above the ear at a distance of a few centimeters. As an alternative, a simple and cheap DIY solution based on modifying an AKG K702 [5] was introduced. It was used in [6, 7].

Recently, the MushRoom headphones were presented [8], and because of their very open, DIY design, they also appear suitable for AR audio research. Here, we present a profoundly extended objective analysis of

---

[1] `https://mysphere.at/`

(a) MushRoom **(b)** K1000 (open) **(c)** Mysphere 3 (open) **(d)** Modified K702 **(e)** HD 650

Fig. 1: Tested models on the KEMAR

MushRoom and compare it to the aforementioned models, thereby assessing its suitability for AR. The analysis includes surrounding transparency measurements with height and Headphone Transfer Function (HpTF) measurements including replacement variability. To gain insights into the perceptual consequences of possible acoustic intransparency, we employ the auditory models presented in [9].

Moreover, we present assembly instructions to enhance the MushRoom with an interface: MrHeadTrackerDSP is based on a Teensy 4.0 microcontroller, and serves as an audio interface with integrated head tracking and headphone equalization. Altogether, the MushRoom headphones and the MrHeadTrackerDSP provide an open-source platform for future research and development dealing with sound in AR.

Section 2 shows the extended evaluation of the MushRoom, comparing it to other models regarding objective and modelled perceptual transparency and variability of HpTF. Section 3 describes the construction of the MrHeadTrackerDSP, including the design of the proposed on-board equalization filter, and measurement of the head tracking latency. Finally, Section 4 concludes the report and discusses ideas for future developments based on the new platform.

2 Evaluation of MushRoom

Evaluation is based on acoustic measurements of transparency and HpTF. We compare the MushRoom to four other models shown in Fig. 1. The first two are the AKG K1000 and the Mysphere 3, which were both measured in their most open configuration and are expected to be very transparent. The other two are the modified K702 [5] and the HD650, which are conventional, high quality open-back headphones. Three out of these five models were also tested in [9], which permits comparison to earlier results.

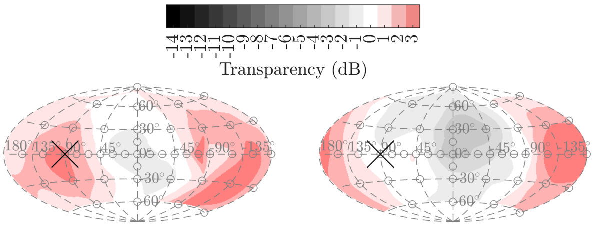
2.1 Measurement Setup

All measurements were conducted in the anechoic chamber “Wilska” at the Aalto Acoustics Lab, Finland. The chamber contains an array of 45 surrounding Genelec 8331A loudspeakers. The loudspeaker positions are marked by circles in Fig. 2. The headphones were measured on a G.R.A.S KEMAR head and torso simulator (HATS) with anthropometric ears.

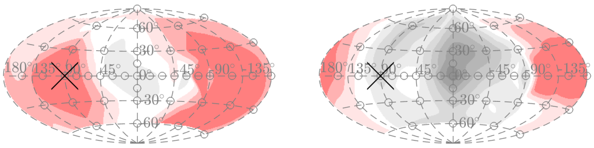
A sequence of 45 overlapping exponential sine sweeps (with a length of 2 s and an overlap of 1.75 s) was played over the self-powered loudspeakers using the RME MadiFace UCX II in combination with the RME ADI-6432 and recorded on G.R.A.S. KEMAR with its 12 AG preamp. The impulse responses obtained via deconvolution were truncated to 256 samples at 48 kHz and saved as a SOFA file. For the evaluation of the HpTFs, the same HATS, preamplifier and audio interface were used. For the transparency measurements, each model was replaced five times, and for HpTF measurements, each model was replaced 20 times.

2.2 Transparency

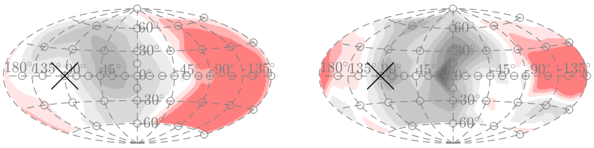
We define transparency as the magnitude ratio of the Head Related Transfer Function (HRTF) when wearing headphones to the HRTF without. Since headphones tend to modify the notches in the HRTFs, the magnitude was smoothed over frequency prior to taking the ratio to avoid unstable behavior of the division.



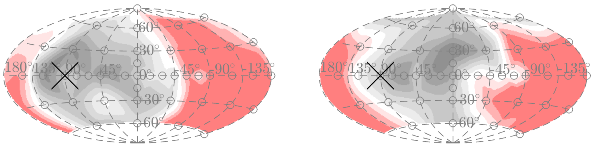
(a) 1kHz MushRoom **(b)** 1kHz K1000 (open)



(c) 2kHz MushRoom **(d)** 2kHz K1000 (open)



(e) 4kHz MushRoom **(f)** 4kHz K1000 (open)



(g) 8kHz MushRoom **(h)** 8kHz K1000 (open)

Fig. 2: Transparency of the K1000 and the MushRoom measured on the left ear, whose approximate location is indicated by the black cross.

Fig. 2 shows transparency of MushRoom and K1000 in detail, in the octave bands 1 kHz, 2 kHz, 4 kHz, 8 kHz for the left ear of the dummy head. In general, the K1000 and the proposed MushRoom exhibit similar behaviour, with an increasing effect of the headphone for higher frequencies. For both models, most impairments are seen on the ipsilateral side. For the K1000 in its open configuration, the regions of strongest attenuation are more frontal than for the MushRoom, where it is slightly more lateral. This is plausible with regards to the location of the acoustically least transparent part of the headphones – the driver. For contralateral directions, both headphones lead to amplifications of the incoming sound, due to reflections off of the contralateral headphone's rigid parts.

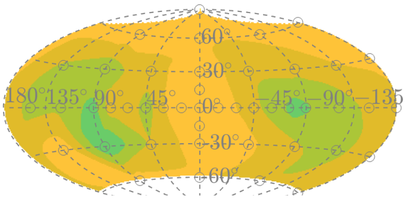
Inspecting the transparency of many headphones as a magnitude ratio in many frequency bands can be tedious, and it is hard to gauge the perceptual relevance of the impairments. Hence, auditory models are used as a helpful tool next, transforming the results to perceptually relevant metrics.

2.3 Models of Perceptual Transparency

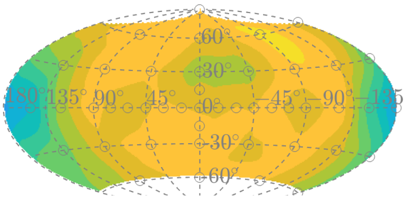
Perceptual consequences of limited transparency were discused in [9]. The authors identified coloration and impairment of localization as the two most important effects. Thus, we adopt the corresponding perceptual models that were tested in [9]. To assess coloration, the composite loudness level (CLL) model was used; Fig. 3 shows the results. K1000, Mysphere and MushRoom show much less coloration than the HD650 and the modified K702. The MushRoom colors sound from the front the least, but does not leave sound from lateral directions completely unaffected. The K1000 has low overall coloration except for sound originating from the back. The Myshere has a slightly higher influence on the sound arriving from the front. The modified K702 headphones show a larger difference in transparency. While they exhibit low coloration values for sound arriving from the front and back, the coloration of lateral sources is closer to the HD650 than to the other models.

Localization errors due to wearing headphones mainly occur in the vertical plane, since they potentially impair natural spectral cues that are important for vertical localization. Lateralization is usually largely preserved. The models presented in [9] allow to compute the percentage of front-back confusions for sources in the horizontal plane and the quadrant error for sources in the median plane. The results are shown in Fig. 4. Measurements were repeated five times; small dots represent a single results and big dots show the median of the five repetitions. Although the K1000 shows the best results here, all tested headphones showed a non-negligible increase in quadrant error and front-back confusion under free-field conditions, compared to open ears. Interestingly, qudrant error and Front-Back confusion do not go hand in hand. The MushRoom has the lowest predicted qudrant error of the tested models, but the expected front-back confusions are lower for K1000 and Mysphere 3.

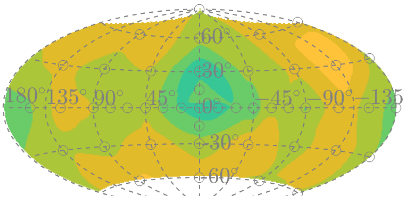
The high variance of the predicted values between repeated measurements for some of the headphones is worth mentioning, as it had not been investigated in [9], where only one measurement was used per headphone model.



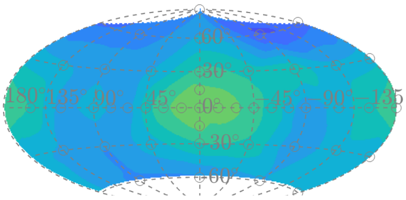
(a) MushRoom ($\overline{\Delta\mathrm{CLL}} = 1.57$ dB)



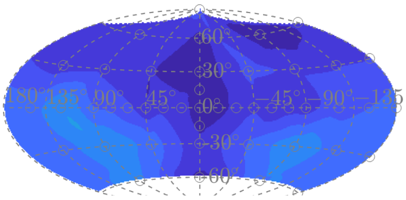
(b) K1000 (open) ($\overline{\Delta\mathrm{CLL}} = 1.81$ dB)



(c) Mysphere 3 (open) ($\overline{\Delta\mathrm{CLL}} = 2.07$ dB)



(d) Modified K720 ($\overline{\Delta\mathrm{CLL}} = 3.85$ dB)



(e) HD650 ($\overline{\Delta\mathrm{CLL}} = 5.95$ dB)

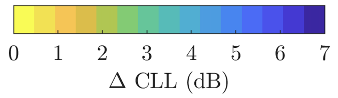


Fig. 3: Predicted coloration according to [9]

.

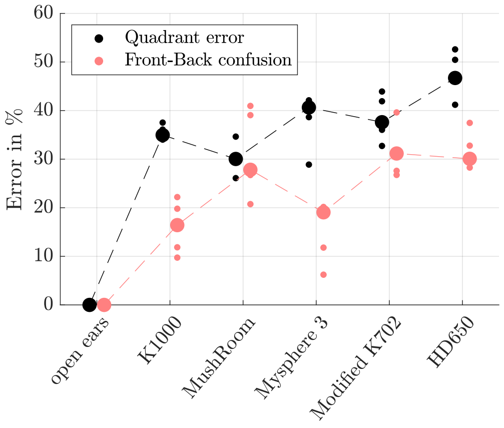


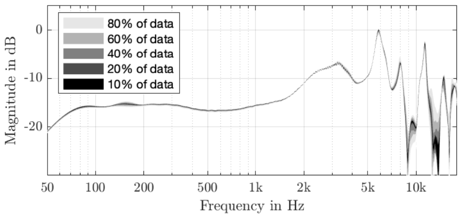
Fig. 4: Predicted percentage of front-back confusions in the horizontal plane and quadrant error in percent, according to [9]

2.4 Headphone Transfer Function and Variability

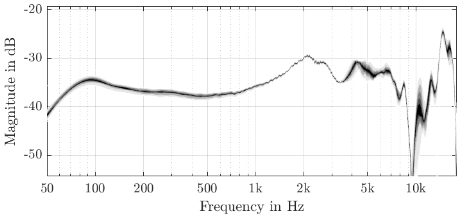
To achieve the best effect of spatialization using HRTFs, it is important to keep the amplitude variation and shifts of the notch frequencies low, which are due to repeatedly putting on and off of the headphones. Also, if headphones are designed to be used in an equalized condition, a distinct HpTF is necessary to obtain the correct filters. If the variation is too high, applying equalization may cause large errors. Therefore, 20 repeated measurements were taken of each headphone model, putting them off and on the HATS after each measurement, as in [10]. For Fig 5, results in HpTF were not smoothed, to show the notches and their variation.

The best results regarding consistency of measured HpTFs in high frequencies were shown by the HD650 This seems reasonable, as the combination of an earcup that creates a defined position on the head, and a small, damped driver work well together in terms of variation. The MushRoom with its position-guiding ear cup has good overall HpTF consistency as well. Nevertheless, at around 9–10 kHz and 13–14.5 kHz notches vary to some extent. The optional felt pad described in [8] could improve this. The modified K702 shows variation above 7 kHz, the K1000 between 4–6 kHz and above 10 kHz. The highest overall variation as well as in variation and quantity of notches was measured on the Mysphere 3.

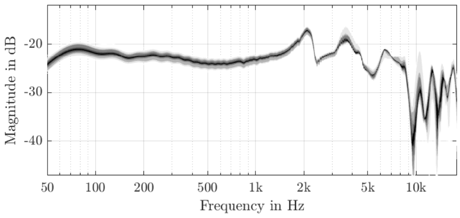
From analyzing the HpTF it becomes clear that equalization should be applied to the MushRoom to perform accurate binaural rendering. The MrHeadTrackerDSP, presented in the next section, can perform such equalization on-device.



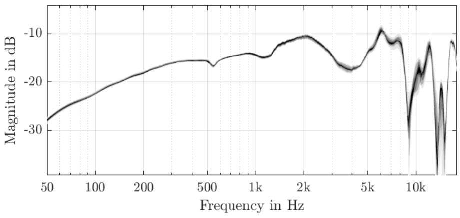
(a) MushRoom (without pads)



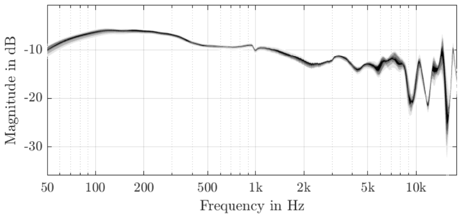
(b) K1000 (open)



(c) Mysphere 3 (open)



(d) Modified K720



(e) HD650

Fig. 5: HpTFs in reference to diffuse response and replacement variability, measured on KEMAR.

3 Design of MrHeadTrackerDSP based on Teensy Microcontroller

As the basis for head tracking, audio IO and filtering, we selected a Teensy 4.0.[2] This USB-based microcontroller development system is compatible with Arduino and has an additional audio adapter board featuring a NXP SGTL5000 Low Power Stereo Codec with headphone amp. Its on-device processing permits headphone equalization. Additionally, we add a BNO055 9-DOF sensor for head tracking.

This new design uses the same IMU sensor as the MrHeadTracker project [11]. Using the Teensy 4.0 with a Teensy Audio Adapter Board instead of the original Arduino Pro Mini and a Arduino USB 2 Serial Micro in the MrHeadTracker project makes tedious soldering and firmware flashing obsolete. By contrast, the Teensy platform allows the device to be easily configured as a class compliant USB audio and MIDI interface. A similar Teensy-based modification of the MrHeadTracker was presented by NOTAM[3], yet without exploiting DSP and audio interfacing capacities.

The required connections are rather simple to make: First we install the Audio Adapter Board below the Teensy 4.0 board by soldering two 14x1 pin headers. This step is straightforward, as the boards are completely pin-compatible as can be seen in Fig. 6b.

Furthermore, we add the BNO055 Adafruit breakout by soldering the SCL and SDA pins of the BNO055 breakout board to Teensys pins 16 (SCL) and 17 (SDA). On both boards, these pins lie next to each other. These two pins correspond to the Teensys second I2C interface while the first one is used for the Audio Adapter Boards I2C connection. Now, only the 3,3V and GND of the BNO055 need to be connected with the corresponding pins on the Teensy 4.0 board using two simple wires. To trigger the calibration one SMD button is added and soldered between pin 1 and ground.

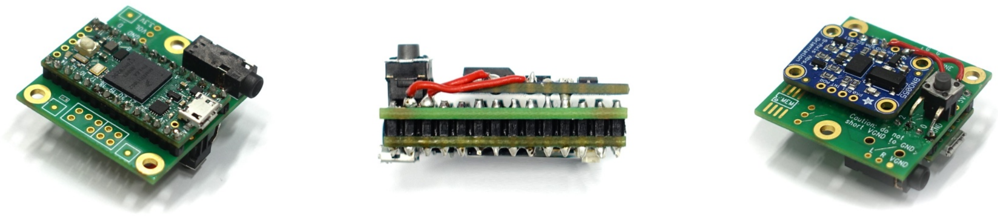
3.1 DSP Signal Flow

The Teensy platform features an audio library containing the most important audio DSP building blocks. It also contains a graphical audio system design tool where a signal path can be configured. This configuration can then be exported as an Arduino header containing the objects and audio routing.

---

[2] `https://www.pjrc.com/teensy/`

[3] `https://github.com/notam02/Teensy-Head-Tracker`



(a) View from above **(b)** View from the side **(c)** View from below

Fig. 6: The Teensy 4.0 board including audio shield and IMU sensor

.

The signal path for the headphone equalisation consists of the stereo usb input object, followed bya 2×12 cascaded biquad-filter objects in parallel, terminating in the stereo I2S output object.

3.2 Code

Some additional simplifications and modifications of the original MrHeadTracker code were made. We got rid of the switch and changed the MIDI library to usbMIDI library. Also, only quaternion data is being sent as they are most meaningfully and unambiguously representing of orientation in space. Furthermore, we added the DSP signal flow and initialisation routines for the audio adapter shield and a peak/notch filter coefficients calculation function. The source code and more detailed building instructions can be found in the MrHeadTracker repository[4].

3.3 Compilation

It is important to note that before compiling and uploading the code to the Teensy board, we need to configure the USB Type in the Arduino IDE Tools section to be "Serial + MIDI + Audio". This unlocks the board, so that it appears as class compliant stereo audio interface as well as an USB-Midi interface.

3.4 Equalization Filter Design

As discussed above, the MushRoom hardware requires digital headphone equalization. For equalization, the mean value of the magnitude response of the 20 HpTF measurements above is used. As a target curve, we use the diffuse field response, so that the signal reproduced at the listeners ears resembles listening in an isotropic field, where sound arrives from all directions with equal energy. The black curve in Fig. 7 shows the magnitude ratio of the mean HpTF X_{HpTF} and the diffuse response X_{Diffuse}, obtained by averaging the magnitude of the HRTFs of a KEMAR HATS. The magnitude responses were smoothed with a 1/3-th-octave wide Gaussian kernel prior to taking the ratio.

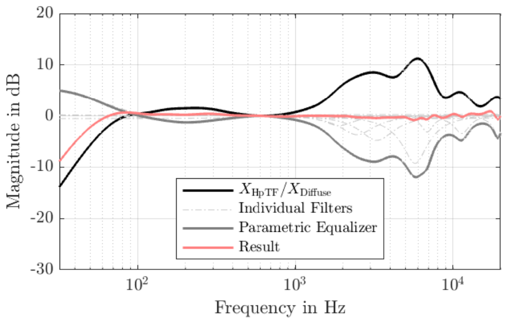


Fig. 7: Equalization filter design for a diffuse field target, using 12 bi-quad sections.

This equalizer can be implemented directly on the Teensy board. As resources are limited, delays should kept short, and the compensation curve is simple, an IIR filter design is proposed. To this end, the impulse responses were imported into the RoomEQ wizard [5] that provides routines for averaging, smoothing, and automated retrieval of the coefficients for a cascade of biquad filters providing equalization to a flat target. The gray curve in Fig. 7 shows the frequency response of this chain of 10 filters. The red curve confirms that the filter cascade inverts the target sufficiently well. Frequencies below 50 Hz were not equalized.

---

[4] `https://git.iem.at/DIY/MrHeadTracker`

[5] `https://www.roomeqwizard.com/`

3.5 Latency

To verify the new head tracking implementation, and to see if it is a viable alternative to the previous MrHeadTracker version, we also performed latency measurements. Latency was measured using the impulsive method described in [12]. Note that for IMU trackers that perform internal sensor fusion, the impulsive latency is not representative of the latency that results from natural head movements. However, it provides the best possible benchmark to date. The mean latency was **34.6 ms**, which is similar to the **30.0 ms** obtained for the MrHeadTracker reported in [12].

4 Conclusion and Outlook

In this contribution, we have shown further measurements of the very transparent MushRoom headphones, and presented as a useful open periphery the new MrHeadTrackerDSP that serves as audio interface, equalizer, and provides basic DSP capacities.

The measurements have shown that the MushRoom headphones have similar transparency properties as existing models, at a fraction of the price.

4.1 Use cases

Together, the MushRoom and the MrHeadTrackerDSP offer a platform for further developments. For example, one may wish to add microphones and connect them to the Teensy board and use the device for prototyping room identification algorithms, required to match the rendering to the real room in AR practise.

Apart from AR, the other main area of application for the enhanced model is production and consumption of spatial music. Together with freely available software tools, it can for example be employed for head-tracked Ambisonics playback.

4.2 Future work

In this paper, we presented comparison of five headphone models. Yet, even more models are available, such as the carefully designed extra-aural headphones BK211 [13]. In addition, some of the models such as the K1000 and the Mysphere can be worn in different configurations. In the future, a full comparison of these options will be pursued. Comparison may include further parameters, such as free-air equivalent coupling. Also, variability should be tested on human users rather than by conducting measurements using a dummy head.

Lastly, the perceptual relevance of the assessed characteristics for research and applications of AR audio should be tested, for example using real/virtual tests under a plausibility [3] or transfer-plausibility [6] paradigm.

References

[1] Neidhardt, A. and Zerlik, A. M., “The Availability of a Hidden Real Reference Affects the Plausibility of Position-Dynamic Auditory AR,” *Frontiers in Virtual Reality*, 2, 2021, doi:10.3389/frvir.2021.678875.

[2] Amengual Garí, S. V., Calamia, P. T., and Robinson, P. W., “Optimizations of the spatial decomposition method for binaural reproduction,” *J. Audio Eng. Soc.*, 68(12), pp. 959–976, 2020.

[3] Lindau, A. and Weinzierl, S., “Assessing the Plausibility of Virtual Acoustic Environments,” *Acta Acust*, 98(5), pp. 804–810, 2012.

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

[5] Meyer-Kahlen, N., Rudrich, D., Wirler, S. A., Winther, S., and Frank, M., “DIY Modifications for Transparent Headphones,” in *148th AES Convention*, 2020.

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

[7] Klanjscek, N., David, L., and Frank, M., “Evaluation of an E-Learning Tool for Augmented Acoustics in Music Education,” *Music & Science*, 4, p. 205920432110375, 2021, doi:10.1177/20592043211037511.

[8] Mülleder, A. and Zotter, F., "Ultralight circumaural open headphones," *154th AES Convention*, 2023.

[9] Llado, P., Mckenzie, T., Meyer-Kahlen, N., and Schlecht, S. J., "Predicting Perceptual Transparency of Head-Worn Devices," *J. Audio Eng. Soc.*, 70(7), pp. 585–600, 2022.

[10] Kulkarni, A. and Colburn, H. S., "Variability in the characterization of the headphone transfer-function," *J. Acoust. Soc. Am.*, 107(2), pp. 1071–1074, 2000, ISSN 0001-4966, doi:10.1121/1.428571.

[11] Romanov, M., Berghold, P., Rudrich, D., Zaunschirm, M., and Frank, M., "Implementation and Evaluation of a Low-cost Head-tracker for Binaural Synthesis," in *142th AES Convention*, 2017.

[12] Meyer-Kahlen, N., Kastemaa, M., Schlecht, S. J., and Lokki, T., "Measuring Motion-to-Sound Latency in Virtual Acoustic Rendering Systems," *J. Audio Eng. Soc.*, 71(6), pp. 390–398, 2023, doi:10.17743/jaes.2022.0089.

[13] Schultz, F., Lindau, A., Makarski, M., and Weinzierl, S., "An extraaural headphone for optimized binaural reproduction," *26th Tonmeistertagung*, pp. 702–714, 2010.