Audio Engineering Society

Conference Paper 9

Presented at the AES 5th International Conference on Audio for Virtual and Augmented Reality
2024 August 19–21, Redmond, WA, USA

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Investigating the Plausibility of Spatially Interpolated Binaural Room Impulse Responses

David Bau[1,2], Hendrik Himmelein[1,2], Jannik Lambert[1], Tim Lübeck[1,2], and Christoph Pörschmann[1]

[1]*Institute of Computer and Communication Technology, TH Köln, Germany*
[2]*Audio Communication Group, TU Berlin, Germany*

Correspondence should be addressed to David Bau (david.bau@th-koeln.de)

ABSTRACT

Rendering virtual acoustic environments with measured binaural room impulse responses (BRIRs) can provide a high level of realism. However, the measurement of BRIRs for multiple head orientations is typically tedious. Appropriate interpolation techniques can lower the demand for spatial resolution and thus reduce the measurement effort. A recent study compared several non-parametric interpolation strategies for very sparse multi-orientational BRIRs. The results suggest a high overall quality for BRIRs interpolated from sparse sets with angular resolutions of 30° and 40°. The present study investigates the realism of such sparsely measured BRIR sets for mixed reality (XR) applications compared to densely measured BRIRs in a plausibility experiment. The participants had to decide if a presented sound source was a real loudspeaker or a headphone simulation, only using their listening expectations as a inner reference. The results indicate that participants could not reliably identify the simulations for either dense or interpolated BRIRs. For dense BRIRs, participants' responses were essentially random, suggesting they were guessing. For interpolated BRIRs, results indicate that participants were able to detect small sensory differences.

1 Introduction

Dynamic binaural auralization is a key component of immersive spatial XR experiences. To achieve convincing auralization via headphones, it is necessary to provide the correct binaural cues and to perceptually match the room acoustics [1]. Several methods exist to derive a spatial representation of the room acoustics from limited data, e.g. from omnidirectional impulse responses (IRs) [2, 3], microphone array IRs [4], or room acoustic simulations [5]. Nevertheless, the long-established approach of measuring Binaural Room Impulse Responses (BRIR) with a dummy head can still be considered the reference method.

Measuring BRIRs at the required angular resolution of 1° [6] is tedious and highly impractical for individual BRIR measurements. In this context, the use of interpolation reduces the measurement effort and allows the coverage of a larger spatial area and thus more flexibility for auralization. A recent study by Bau et al. [7] compared various non-parametric BRIR interpolation methods and found that even with very sparse BRIR sets, having angular resolutions of 30° or 40°, it is possible to achieve a quality comparable to that of a non-interpolated BRIR set. For an accurate interpolation, the alignment of the early reflection peaks is of

key importance, as those peaks are essential for spatial perception [1, 8]. A simple linear interpolation without proper alignment will result in comb-filtering effects and temporal smearing of the reflection peaks [9]. The results of the comparison in [7], conducted as a Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test, provide a useful evaluation of the interpolation methods relative to a non-interpolated BRIR dataset. However, the question of how convincing an auralization with interpolated BRIRs would be in an XR scenario has not yet been addressed.

An established approach to assess the suitability of an auralization system for XR is the plausibility paradigm introduced by Lindau and Weinzierl [10]. It can be described as the credibility of a virtual environment compared to a users expectations based on their listening experience, also referred to as *inner reference* [11]. Lindau and Weinzierl [10] evaluated the plausiblity of a BRIR auralization system by presenting sound stimuli either from real loudspeakers or their virtual headphone-based simulations. Based on signal detection theory (SDT), a framework originally propsed for psychophysics [12], the assessment involved a simple yes/no paradigm to assess the participants' ability to detect the presence of a signal, i.e., audible cues that can reveal a simulation. Therefore, participants were asked if they heard the virtual representation ("yes") or the real loudspeaker signals ("no"). The results showed nearly 50% correct answers suggesting that participants were randomly guessing and thus were unable to distinguish between both presentations. As a result, such a system was considered as *plausible*.

Pike et al. [13] replicated a plausibility experiment similar to Lindau and Weinzierl [10] and added elevated sound sources. The results indicated that participants could detect the simulation in some cases, slightly above the expected guessing rate. The authors suspected that the elevated sound sources might have led to localization errors. Furthermore, the participants had more time to familiarize themselves with the real and virtual stimuli before judging, which may have biased their inner reference. Neidhardt et al. [14] investigated the plausibility of measured BRIRs in a scenario where participants could walk past a real loudspeaker and a corresponding virtual loudspeaker based on BRIR synthesis. Two scenarios were examined: the first involving only virtual sources, and the second incorporating a mixture of virtual sources and real loudspeaker stimuli. In the virtual-only scenario, a high level of plausibility was attained. Conversely, in the scenario that included the real loudspeaker, participants were able to identify the simulation quite reliably. Several other experiments have been conducted for binaural auralizations that used the plausibility paradigm, including parametric renderings [3, 15], acoustic simulations [5], microphone array recordings [16] or anechoic renderings using individual HRTFs [17] and non-individual HRTFs [18].
In general, in most of the tested auralizations, participants could not reliably identify the simulations. But in almost all cases, no pure guessing rate was achieved, indicating small audible differences to the inner reference. The decrease in plausibility was attributed to different factors like perceivable tracking latency, headphone influence, or a biased internal reference. It is important to note that the internal reference in particular can have an unpredictable impact, as it can be biased by a variety of contextual variables such as the duration and intensity of training, the presence visual cues or the selection of stimuli. This must be considered when designing a plausibility experiment.

While the results of the aforementioned studies suggest that plausibility can be achieved with sufficiently fast head-tracking and measured BRIRs of non-elevated sound sources, the effect of interpolating sparsely measured BRIRs on plausibility is yet to be evaluated.

Thus, in this study, we present the results of a plausibility assessment of auralizations based on measured BRIRs, similar to Lindau and Weinzierl [10], but with a focus on interpolated BRIRs. We investigated three conditions: The first condition employed a dense BRIR set, to form a baseline. The other two conditions used BRIR sets interpolated from sparse measurements with angular resolutions of 30° or 45°, respectively. This study extends the comparative evaluation of BRIR interpolation from [7] and provides an estimate of the resolution required for measured BRIRs to achieve plausibility.

2 Methods

2.1 BRIR Interpolation

For the BRIR interpolation in this study, we chose the Spherical Array Interpolation by Time Alignment (SARITA) method, which provided the best interpolation results in [7]. The method was originally proposed for interpolating SMA signals with a limited number of

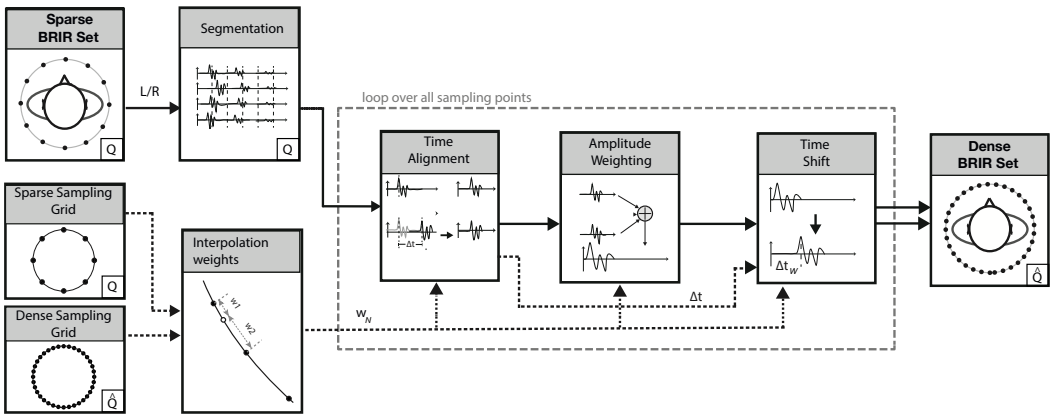


Fig. 1: Block diagram of the SARITA method adapted to BRIR interpolation. The BRIR set with the sparse sampling grid Ω, with Q sampling points, are upsampled to $\widehat{Q}$ sampling points on a dense sampling grid $\widehat{\Omega}$. Processing is applied to left and right ear signals separately.

microphones [19] and has been adapted to interpolate BRIRs measured at a sparse set of view orientations [7].

In the following, we give a brief overview of the method adapted to BRIR interpolation in the horizontal plane. For a comprehensive summary of the method, the reader is referred to [7]. The method employs an interpolation process in short, overlapping time segments. To prevent artifacts resulting from the interpolation of sound incidence peaks at different arrival times, a time-alignment is applied within each segment. A schematic overview of the SARITA method is given in Figure 1. For each segment (32 samples at a sampling rate of 48 kHz, overlap of 50%), the following processing is applied to obtain an interpolated BRIR at a dense view orientation $\widehat{\Omega}_j$ from the two BRIRs at adjacent sparse view orientations Ω_i and Ω_{i+1}. First, the time shift Δt between the dominant peaks in each of the two BRIR segments at Ω_i and Ω_{i+1} is computed by cross-correlation. Both BRIRs are time-aligned with Δt. The time-aligned BRIRs are then summed by means of linear interpolation with amplitude weights $w_i = \frac{\widehat{\Omega}_j - \Omega_i}{\Omega_{i+1} - \Omega_i}$ and $w_{i+1} = 1 - w_i$. Finally, the resulting BRIR is inversely time-shifted with the weighted time shift Δt_w, where w is the interpolation weight corresponding to the more distant BRIR.

It should be noted that the SARITA method does not separate the impulse response into components such as the direct part, early reflections, and late reverb. The described algorithm is most effective when applied to the early part of the impulse response, where distinct peaks allow for meaningful alignment. However, it was shown in [19] that the method does not introduce any systematic coloration to diffuse parts, making it suitable for application to the full-length impulse response.

2.2 Setup

We conducted the experiment in a quiet office room with a floor area of 40 m^2 and a background noise floor of 27 dB(A). The loudspeakers (Genelec 8020A) were placed with a height of 1.29 m on a 1.62 m radius around the listener at various angles, as shown in Figure 3. One loudspeaker was located at the back of the listener (azimuth 225°) so that the listener had to turn the entire body to face it. The placement of the loudspeakers corresponds to typical sound source positions in an office environment, as shown in Figure 2.

We used the HTC Vive Virtual Reality system to determine the participants' head movement. For this, we developed a tracking receiver application[1] in C++ that collects the tracking data from the Vive Tracker through the OpenXR standard[2] and forwards the desired positional data, i.e. listener azimuth view, via Open Sound Control (OSC) [20] messages. It should be noted that while the tracking system tracked motion data in six

[1] https://github.com/AudioGroupCologne/OpenXRTrackingData/
[2] https://www.khronos.org/openxr/



Fig. 2: Setup for the BRIR measurement of the loudspeaker setup in the office room, using the Varisphear measurement system and the Neumann KH100 artificial head.

degrees-of-freedom (6DoF), only the horizontal head rotation, or yaw, was used in this experiment.

We used PyBinSim [21], a Python application for real-time dynamic convolution with binaural impulse responses for the binaural rendering of the virtual sound sources. A control application was implemented to manage the evaluation process. It receives the OSC tracking data from the tracking receiver and controls pyBinSim by sending orientation data (filter selection) and playback control (start/stop playback, select sound file) via OSC messages. Furthermore, it handles audio playback of the real sound sources via the loudspeakers. The participants entered their responses with a Behringer X-Touch mini MIDI controller.

For the binaural auralization, we used AKG K1000 extra-aural headphones. The extra-aural headphones were chosen to minimize the shadowing of the headphones in the sound propagation from the real loudspeaker to the ears [22, 23]. The headphones were equalized using compensation filters [24], which were measured with the same headphones and artificial head as employed for the BRIR measurement (see Section 2.3). An RME Babyface Pro FS, extended with an RME 8 I/O 8 channel AD/DA converter via ADAT, running at 256 samples buffer size at a sampling rate of 48 kHz, was used to send the audio data to the headphones or loudspeakers. The playback level was set to a sound pressure level of 67 dB(A) at the left ear canal entrance for pink noise coming from the frontal loudspeaker. The pink noise had the same loudness as the stimuli used in the experiment. The headphone level was adjusted so that the real loudspeaker and its virtual counterpart had the same loudness at the ear canal entrance.

2.3 Materials

The BRIRs were measured in the same room and for the same loudspeaker arrangement used in the experiment. Figure 2 shows the measurement setup in the experiment room. We used a Neumann KU100 artificial head, mounted on a Varisphear [25] turntable system to obtain BRIRs for 360 equally spaced horizontal orientations for all five loudspeakers at an angular resolution of 1°. During the measurement, the AKG K1000 headphones were placed on the artificial head to account for their effect on the sound from the real loudspeakers. Exponential sine sweeps with a length of 2^{17} samples at a sampling rate of 48 kHz ($\sim$2.5 s) were used for all measurements.

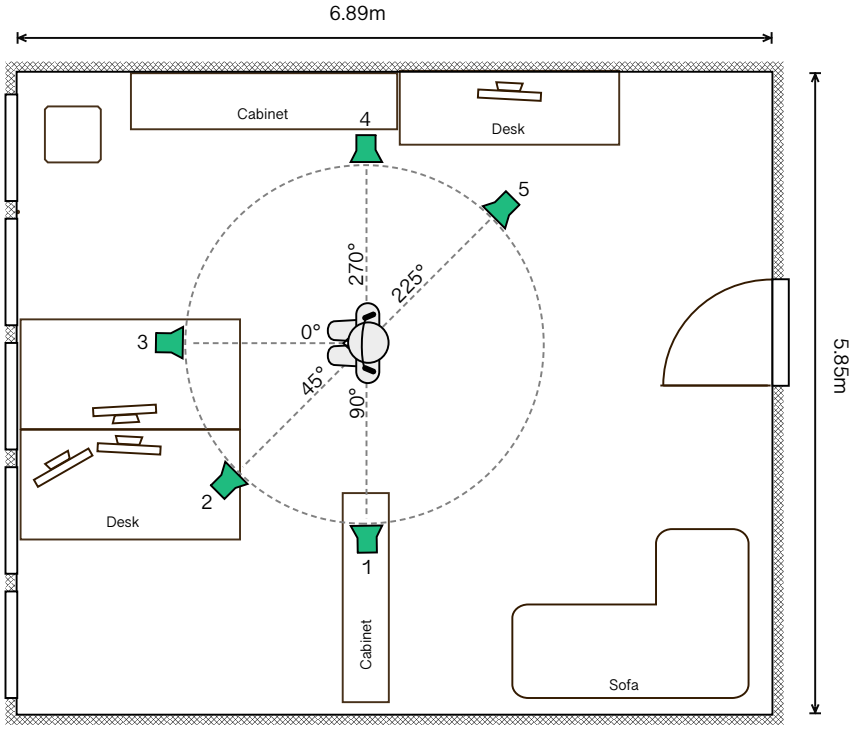


Fig. 3: The office room floor plan and loudspeaker setup that was employed for the plausibility experiment.

For the interpolated datasets, we subsampled the dense BRIR datasets to sparse datasets at two different angular resolutions, i.e., 30° and 45°, corresponding to 12 and 8 horizontal orientations each. Each sparse BRIR set was then upsampled to a dense BRIR set with a resolution of 1°, using the SARITA method described in detail in Section 2.1.

For the assessment, we used 25 stimuli from the sound file database applied by Lübeck and Pörschmann [18]. The files contained small anechoic excerpts of drums, piano, guitar, saxophone, cello, female, and male speech. The duration was limited to 6 s, with a short fade-out to avoid artifacts. The loudness of all sound files was matched to -18 LUFS according to EBU recommendation 128 [26].

2.4 Participants

In total, 48 participants (average age: 32.8 years, 11 female, and 37 male) took part in the experiment, with self-reported normal hearing. 11 of them had substantial experience in listening experiments, while the remaining participants were naive listeners.

2.5 Procedure

After putting the headphones on and calibrating the head-tracker's frontal direction, the participants were presented with a guitar sample from the frontal loudspeaker, to familiarize themselves with the sonic characteristics of the loudspeaker and the room.

The assessment followed the approach based on Lindau and Weinzierl [10]. In each of the 100 randomized trials, the subjects listened to one of the 25 sound samples presented either via one of the loudspeakers or virtually via headphones. To avoid direct comparison, no sound sample was played twice in a row. After each trial, participants judged whether they had heard a real stimulus via the loudspeaker or a virtual stimulus via headphones. Subjects responded by pressing the "yes" or "no" button on the MIDI Controller.

During the experiment, participants were sitting on a swivel chair in the center position of the loudspeaker setup (compare Figure 3). They were allowed to rotate their head and the stool but were instructed to maintain a straight sitting position to prevent excessive translational movements. Notably, the participants were not explicitly instructed to avoid vertical head movements, although the BRIRs were measured in the horizontal plane only. The supervisor was present throughout the experiment and monitored the participants' movements.

We evaluated three conditions with different BRIR datasets. One group of 17 participants assessed a dense BRIR dataset, while two other groups, with 17 and 14 participants respectively, evaluated interpolated datasets synthesized from sparser datasets with resolutions of 30° and 45°. No participant performed the test for more than one condition.

2.6 Data Analysis

We evaluate the results using the SDT framework. The employed yes/no question to assess the participant's ability to detect a virtual sound source has four possible outcomes: *hit* corresponds to a correctly detected virtual representation, *correct rejection* to a correctly detected real loudspeaker, *false alarm* to a falsely detected virtual source, and *miss*, to a virtual source credited as a loudspeaker representation. From the false alarm rate p_{fa} and hit rate p_{hit}, a participant's response behavior can be condensed into two measures: the sensitivity d' and the response bias β. The sensitivity is a measure of how well participants can discriminate between the two stimuli while the response bias indicates whether the tendency is toward "yes" (virtual) or "no" (real). Using the inverse cumulative normal distribution function $Z(p)$, the sensitivity can be computed as:

$$d' = Z(p_{hit}) - Z(p_{fa}) \tag{1}$$

A d' value of 0 indicates no discrimination between virtual and loudspeaker presentations, implying perfect plausibility. Conversely, $d' > 0$ indicates possible discrimination between real and virtual presentations [27]. Sensitivity can be expressed as p_{2FAC} values for a more intuitive interpretation according to

$$p_{2AFC} = \Phi\left(\frac{d'}{\sqrt{2}}\right) \tag{2}$$

with $\Phi(z)$ the cumulative standard normal distribution [28]. The bias β is determined through

$$\beta = e^{-\frac{Z(p_{hit})^2}{2} + \frac{Z(p_{fa})^2}{2}} \tag{3}$$

as per [27], a method that is resilient against errors. A bias value of $\beta > 1$ suggests a tendency among participants to perceive the source as real, whereas a bias value of $\beta < 1$ indicates a decision tendency towards virtual.

In this study, we calculated an individual d' for each participant to quantify their ability to detect a simulation. Additionally, we tested the averaged sensitivity per condition d'_{avg} against the plausibility limit ($d'_{min} = 0.177$ or $p_{2AFC} = 0.55$) proposed by Lindau and Weinzierl [10] using a t-test to determine possible significant differences. It should be noted that this plausibility limit is arbitrarily set by Lindau and Weinzierl [10] but has been used here for better comparability. d' and β values were calculated using the psycho package (0.6.1) for R in RStudio. Normality was tested using the Shapiro-Wilk test since it provides the most robust results for smaller sample sizes [29].

3 Results

Overall, the results show clear differences between the angular resolutions. For the 1° condition, the d'_{avg} is below the given plausibility limit of 0.177 with $d'_{avg} = 0.14$. However, with decreasing angular resolution, d'_{avg} is increasing with $d'_{avg} = 0.40$ for 30° and $d'_{avg} = 0.44$ for the 45° condition. Figure 4 depicts the differences between all three conditions. Individual d' are shown as dots, and colored boxes around the mean (white line) denote the overall confidence interval. The bias β for each group is shown below including confidence intervals as error bars. β depicts a clear tendency towards real loudspeaker sources for all conditions.

The results were first analyzed for normality using the Shapiro-Wilk test. The test showed no violation of the normal distribution for all conditions. A subsequent t-test against the plausibility limit d'_{min} showed no significant differences for the 1° condition [$t(16) = -0.39, p = 0.71$, d = 0.09]. However, significant differences were found for the 30° condition [$t(16) = 2.37, p = 0.031$] and the 45° condition [$t(13) = 2.36, p = 0.035$]. A subsequent pairwise t-test between the three conditions with

benjamin-hochberg adjustment revealed no significant differences ($p.adj_{1°-30°} = 0.08$, $p.adj_{1°-45°} = 0.08$, $p.adj_{30°-45°} = 0.78$).

Since some participants stated that the ratings might fluctuate throughout the experiment, we conducted a further analysis to examine possible changes. We divided the 100 trials into four epochs, assigning 25 trials to each epoch as suggested in Weber et al. [30]. Figure 5 shows the course of each condition, where almost no fluctuation is observable. Results also showed no significant differences for the 1° ($F(3,64) = 0.34, p = 0.79$) nor 30° ($F(3,64) = 0.26, p = 0.86$) or 45° ($F(3,52) = 0.44, p = 0.73$) condition.

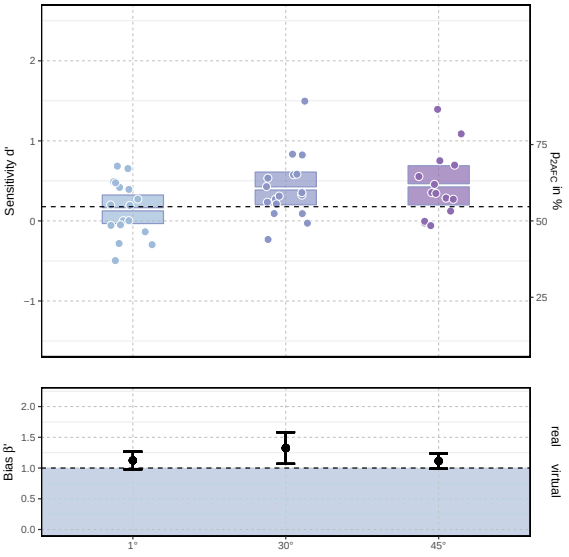


Fig. 4: Differences in sensitivity values (d'_{avg}) for each condition including confidence intervals. Individual d' are displayed as dots (left side). Response bias is displayed as β per condition including confidence intervals as errorbars (right side)

4 Discussion

Overall, the results indicate that participants could not reliably detect the simulation in any of the three rendering conditions. The reference condition without interpolation achieved an average sensitivity d'_{avg} of 0.14, similar to the plausibility level reached in comparable studies by Lindau and Weinzierl [10] ($d'_{avg} = 0.051$) or Pike et al. [13] ($d'_{avg} = 0.195$). The d'_{avg} of 0.14 also

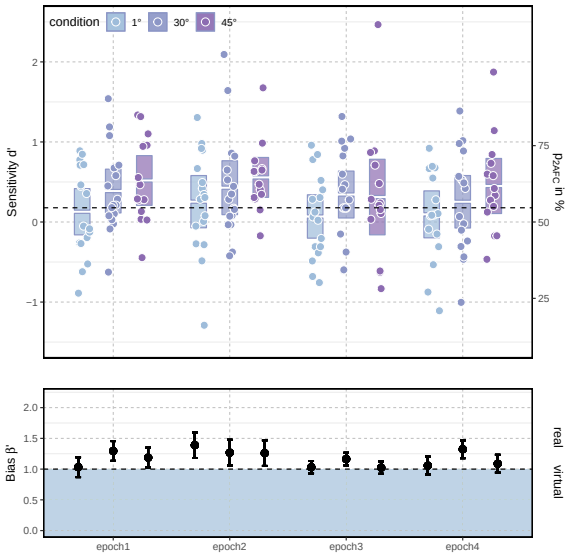


Fig. 5: Overview of the changes in d'_{avg} per epochs in every angular resolution condition. Individual d' are displayed as dots. Response bias is displayed as β per condition including confidence intervals as errorbars

lies well below the threshold $d'_{min} = 0.177$ for a plausible presentation, as specified in Lindau and Weinzierl [10].

However, the sensitivity for the two conditions with interpolated BRIRs was slightly higher ($d'_{avg} = 0.4$ for 30°, $d'_{avg} = 0.44$ for 45°), indicating a noticeable sensory difference between real and virtual sound sources. Surprisingly, the sensitivity for the 45° condition is only slightly higher than the one for 30°. Comparing the individual distributions of d' values in Figure 4, a similar central population trend can be observed for the conditions 30° and 45°. However, there are more outliers in the 45° condition. This could imply that both conditions are similarly convincing for most of the subjects, but a few participants were able to detect more noticeable auditory impairments in the 45° condition. This is consistent with the results in Bau et al. [7], where participants could detect impairments for the 30° and 45° conditions of the SARITA interpolation in a MUSHRA comparison but still gave high ratings to both conditions. However, the suspected indistinguishability between the 30° condition and the reference, as assumed by Bau et al. [7], could not be verified for all participants in the present study. This discrepancy

could be due to the different test paradigms and scales used in the two studies. Subjects in a MUSHRA experiment might rate a stimulus as "good" even if they perceive a minor difference from the reference, whereas subjects in a plausibility experiment rate a stimulus as "no" even as soon they perceive a very slight difference from the internal reference. Another reason for the difference might be the BRIR data employed in the different studies. The MUSHRA experiment in Bau et al. [7] used BRIRs measured without the K1000 headphone. However, in the present study, the BRIRs were measured with the K1000 on the artificial head. The acoustic effect of the headphones in front of the ears might increase the spatial complexity of the BRIRs. Increased spatial complexity is likely to affect interpolation and could lead to audible interpolation artifacts.

The bias β is higher than the threshold 1, indicating that participants tended to rate the simulation as real, which is a common observation in other plausibility experiments [10, 16, 18]. The difference in bias observed in the 30° condition can be attributed to two participants who consistently answered "no" (real loudspeaker) in almost all of the trials.

An interesting observation from Figure 4 is the wide distribution of individual d'_{avg} values, especially for interpolated BRIRs. This implies that the level of sensitivity to interpolation artifacts is highly individual between participants. From observations during the experiments and informal reports by participants, we assume that participants followed individual strategies to detect the simulation. For example, looking for cues that could immediately break immersion, such as unstable source positions or rendering artifacts, appeared to be ineffective. On the other hand, participants who tried to determine small changes in the spatial properties, like diffuseness and perceived source distance, often achieved higher d'_{avg} values.

Furthermore, some participants appeared to be distracted or confused by the initially convincing presentation. No participant, even the outliers, reported finding cues that led them to confident decisions. Hence, we suspect that interpolated BRIRs can generally provide a plausible acoustic scene despite showing slight audible impairments. This raises the question of whether impairments would be noticed in a real-world situation where a user does not focus on searching for audible impairments. However, these informal reports are difficult to verify. To address this, future plausibility studies could incorporate or be combined with additional assessments, such as a Spatial Audio Quality Inventory (SAQI) [31] comparison as in [16], or a questionnaire to link the plausibility results to perceptual attributes. These measures would provide a more comprehensive understanding of the participant's experiences and the validity of their reports.

An analysis of the sensitivity throughout the assessment showed no significant change in d'_{avg} values, indicating no learning effect during the evaluation. This is interesting, especially for the interpolated conditions, where impairments were more or less audible and could have provided cues that participants could identify over time. A small tendency, if any, can be seen in the contrary direction. A decrease of the d'_{avg} value, indicates that the participants became either more uncertain or fatigued. However, the overall constant results over time might also indicate that the number of trials could be reduced in future studies, as each epoch of 25 trials has a similar d'_{avg} as the full 100 trials.

5 Conclusion

In this study, we conducted a plausibility experiment to assess the realism of sparsely measured BRIRs. The results and informal reports indicate that participants could not reliably distinguish between a real loudspeaker and a binaural simulation in a yes/no choice test. Among the three tested conditions, only the non-interpolated reference condition had a sensitivity threshold close to the guessing rate. In contrast, the two interpolated conditions with 30° and 45° angular resolution exhibited a small sensory difference from the real loudspeaker.

Acknowledgements

We would like to express our gratitude to the participants who took part in the experiment. We would also like to thank Tobias Weber for his support in planning the plausibility experiment.

References

[1] Neidhardt, A., Schneiderwind, C., and Klein, F., "Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms - Literature Review and Theoretical Framework," *Trends in Hearing*, 26, 2022, doi:10.1177/23312165221092919.

[2] Pörschmann, C., Stade, P., and Arend, J. M., "Binauralization of Omnidirectional Room Impulse Responses - Algorithm and Technical Evaluation," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 345–352, Edinburgh, UK, 2017.

[3] Arend, J. M., Garí, S. V. A., Schissler, C., Klein, F., and Robinson, P. W., "Six-Degrees-of-Freedom Parametric Spatial Audio Based on One Monaural Room Impulse Response," *J. Audio Eng. Soc.*, 69(7/8), pp. 557–575, 2021, doi: 10.17743/jaes.2021.0009.

[4] McCormack, L., Pulkki, V., Politis, A., Scheuregger, O., and Marschall, M., "Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution," *J. Audio Eng. Soc.*, 68(5), pp. 338–354, 2020, doi:10.17743/jaes.2020.0026.

[5] Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., and Weinzierl, S., "A round robin on room acoustical simulation and auralization," *J. Acoust. Soc. Am.*, 145(4), pp. 2746–2760, 2019, doi:10.1121/1.5096178.

[6] Lindau, A., Maempel, H., and Weinzierl, S., "Minimum BRIR grid resolution for dynamic binaural synthesis," *J. Acoust. Soc. Am.*, 123(5), pp. 3498–3498, 2008, doi:10.1121/1.2934364.

[7] Bau, D., Himmelein, H., and Pörschmann, C., "Comparison of Non-Parametric Interpolation Techniques for Sparsely Measured Binaural Room Impulse Responses [Manuscript accepted for publication]," 2024.

[8] Engel, I. and Picinali, L., "Reverberation and its Binaural Reproduction: The Trade-off between Computational Efficiency and Perceived Quality," in B. F. G. Katz and P. Majdak, editors, *Advances in Fundamental and Applied Research on Spatial Audio*, IntechOpen, Rijeka, 2022, doi:10.5772/intechopen.101940.

[9] Zaunschirm, M., Frank, M., and Zotter, F., "Binaural Rendering with Measured Room Responses: First-Order Ambisonic Microphone vs. Dummy Head," *Appl. Sci.*, 10(5), p. 1631, 2020, ISSN 2076-3417, doi:10.3390/app10051631.

[10] Lindau, A. and Weinzierl, S., "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acustica united with Acustica*, 98(5), pp. 804–810, 2012, doi:10.3813/AAA.918562.

[11] Kuhn-Rahloff, C., *Realitätstreue, Natürlichkeit, Plausibilität*, Springer Berlin, Heidelberg, 1 edition, 2012.

[12] Green, D. M. and Swets, J. A., *Signal Detection Theory and Psychophysics*, Robert E. Krieger, 1974.

[13] Pike, C., Melchior, F., and Tew, T., "Assessing the Plausibility of Non-Individualised Dynamic Binaural Synthesis in a Small Room," in *Audio Engeneering Society 55th International Conference*, Helsinki, Finland, 2014.

[14] Neidhardt, A., Tommy, A. I., and Pereppadan, A. D., "Plausibility of an interactive approaching motion towards a virtual sound source based on simplified BRIR sets," in *Audio Engineering Society Convention 144*, Milan, Italy, 2018.

[15] Meyer-Kahlen, N., Garı, S. A., McKenzie, T., Schlecht, S. J., and Lokki, T., "Transfer-Plausibility of Binaural Rendering with Different Real-World References," in *Proceedings of 48th DAGA*, Stuttgart, Germany, 2022.

[16] Ackermann, D., Fiedler, F., Brinkmann, F., Schneider, M., and Weinzierl, S., "On the Acoustic Qualities of Dynamic Pseudo-Binaural Recordings," *J. Audio Eng. Soc.*, 68(6), pp. 418–427, 2020, doi:10.17743/jaes.2020.0036.

[17] Oberem, J., Masiero, B., and Fels, J., "Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods," *Applied Acoustics*, 114, pp. 71–78, 2016, doi:10.1016/j.apacoust.2016.07.009.

[18] Lübeck, T. and Pörschmann, C., "Evaluating the plausibility of non-individual head-related transfer functions in anechoic conditions," in *10th Convention of the European Acoustics Association: Forum Acusticum*, Torino, Italy, 2023.

[19] Lübeck, T., Arend, J. M., and Pörschmann, C., "Spatial Upsampling of Sparse Spherical Microphone Array Signals," *IEEE/ACM Trans. Audio*

Speech Lang. Process., 31, pp. 1163–1174, 2023, doi:10.1109/TASLP.2023.3250841.

[20] Wright, M. and Freed, A., "Open SoundControl: A New Protocol for Communicating with Sound Synthesizers," Thessaloniki, Greece, 1997.

[21] Neidhardt, A., Klein, F., Knoop, N., and Köllmer, T., "Flexible Python Tool for Dynamic Binaural Synthesis Applications," in *Audio Engineering Society Convention 142*, 2017.

[22] Schneiderwind, C., Neidhardt, A., and Meyer, D., "Comparing the effect of different open headphone models on the perception of a real sound source," in *150th Audio Engineering Society Convention*, Online, 2021.

[23] Pörschmann, C., Arend, J. M., and Gillioz, R., "How wearing headgear affects measured head-related transfer functions," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*, pp. 49–54, 2019, doi:10.25836/sasp.2019.27.

[24] Bernschütz, B., "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *Proceedings of the 39th DAGA*, pp. 592–595, DEGA, Merano, Italy, 2013.

[25] Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S., "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio," in *Proceedings of 36th DAGA*, pp. 717–718, Berlin, Germany, 2010.

[26] "EBU R 128 - Loudness Normalisation and Permitted Maximum Level of Audio Signals V5," 2023.

[27] Stanislaw, H. and Todorov, N., "Calculation of Signal Detection Theory Measures," *Behavior Research Methods, Instruments, & Computers*, 31(1), pp. 137–149, 1999-03, doi:10.3758/BF03207704.

[28] Wickens, T. D., *Elementary Signal Detection Theory*, Oxford University Press, 2001, doi:10.1093/acprof:oso/9780195092509.001.0001.

[29] Mohd Razali, N. and Yap, B. W., "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," *J. Stat. Model. Analytics*, 2, 2011.

[30] Weber, T., Lübeck, T., and Pörschmann, C., "Evaluating the Influence of different Generic Head Related Transfer-Functions on Plausibility of Binaural Rendering," in *Proceedings of 50th DAGA*, Hannover, Germany, 2024.

[31] Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkmann, F., and Weinzierl, S., "A Spatial Audio Quality Inventory (SAQI)," *Acta Acustica united with Acustica*, 100(5), pp. 984–994, 2014, doi:10.3813/AAA.918778.