Binaural reproduction of dummy head and spherical microphone array data—A perceptual study on the minimum required spatial resolution

Tim Lübeck, Johannes M. Arend and Christoph Pörschmann

Citation: The Journal of the Acoustical Society of America **151**, 467 (2022); doi: 10.1121/10.0009277
View online: https://doi.org/10.1121/10.0009277
View Table of Contents: https://asa.scitation.org/toc/jas/151/1
Published by the Acoustical Society of America

JASA ARTICLE

Binaural reproduction of dummy head and spherical microphone array data—A perceptual study on the minimum required spatial resolution

Tim Lübeck,[a),b)] Johannes M. Arend,[a),c)] and Christoph Pörschmann[d)]
Technische Hochschule Köln—University of Applied Sciences, Institute of Communications Engineering, Cologne, Germany

ABSTRACT:
Dynamic binaural synthesis requires binaural room impulse responses (BRIRs) for each head orientation of the listener. Such BRIRs can either be measured with a dummy head or calculated from the spherical microphone array (SMA) data. Because the dense dummy head measurements require enormous effort, alternatively sparse measurements can be performed and then interpolated in the spherical harmonics domain. The real-world SMAs, on the other hand, have a limited number of microphones, resulting in spatial undersampling artifacts. For both of the methods, the spatial order N of the underlying sampling grid influences the reproduction quality. This paper presents two listening experiments to determine the minimum spatial order for the direct sound, early reflections, and reverberation of the dummy head or SMA measurements required to generate the horizontally head-tracked binaural synthesis perceptually indistinguishable from a high-resolution reference. The results indicate that for direct sound, $N = 9$–13 is required for the dummy head BRIRs, but significantly higher orders of $N = 17$–20 are required for the SMA BRIRs. Furthermore, significantly lower orders are required for the late parts with $N = 4$–5 for the early reflections and reverberation of the dummy head BRIRs but $N = 12$–13 for the early reflections and $N = 6$–9 for the reverberation of the SMA BRIRs.
 https://doi.org/10.1121/10.0009277
(Received 21 June 2021; revised 4 December 2021; accepted 17 December 2021; published online 27 January 2022)
[Editor: Efren Fernandez-Grande] Pages: 467–483

I. INTRODUCTION

The binaural auralization of virtual acoustic environments can be achieved by convolution with binaural room impulse responses (BRIRs). Such BRIRs can either be obtained with impulse response measurements using a dummy head (Stade *et al.*, 2012), calculated from spherical microphone array (SMA) captures (Bernschütz, 2016), or generated by parametric synthesis (Mccormack *et al.*, 2020; Merimaa and Pulkki, 2004; Pulkki, 2007; Tervo *et al.*, 2013) or simulation (Brinkmann *et al.*, 2019; Savioja and Svensson, 2015; Vorländer, 2008). The auralization with the BRIRs directly measured with a dummy head can still be regarded as the ground truth (Brinkmann *et al.*, 2014; Lindau, 2014). Many studies, either on SMA auralization, room simulation, or parametric synthesis, compare to a reference measured with a dummy head (Ahrens, 2019; Ahrens and Andersson, 2019; Bernschütz, 2016; Garí *et al.*, 2019). Ideally, the BRIRs calculated from the SMA captures are equivalent to these dummy head BRIRs, which is why we focus on these two methods: the BRIRs based on the dummy head measurements and the BRIRs synthesized based on the SMA measurements together with a set of head-related transfer functions (HRTFs) of the same dummy head.

The dynamic binaural synthesis, where the sound field is adapted to the listeners' orientation, requires BRIRs for the arbitrary directions. Lindau *et al.* (2008) showed that a grid resolution of 2° in the horizontal and vertical directions ensures artifact-free auralization. Because the dense full-spherical dummy head BRIR sets are costly in terms of measurement effort and memory consumption, often interpolation of the sparse BRIR sets to the desired directions is applied, which introduces artifacts that are possibly degrading the auralization quality.

The impulse response measurements with the SMAs and a set of anechoic HRTFs are an alternative to the full-spherical dummy head measurements. Once the sound field is captured with the SMA, BRIRs for the arbitrary head orientations can be synthesized. These impulse responses can be measured simultaneously with the real-world SMAs or sequentially with the single-microphone measurements using automated systems such as the VariSphear (Bernschütz *et al.*, 2010). The binaural synthesis from the SMA captures also has the advantages that individual HRTFs can easily be integrated and real-time applications can be implemented. For the real-world SMAs, the major limitation is the number of microphone capsules on the array surface, which leads to undersampling errors and impairments of the binaural signals (Lübeck *et al.*, 2020a). Currently, commercially available

[a)]Also at: Technical University of Berlin, Audio Communication Group, Berlin, Germany.
[b)]Electronic mail: tim.luebeck@th-koeln.de, ORCID: 0000-0003-2870-095X.
[c)]ORCID: 0000-0002-5403-4076.
[d)]ORCID: 0000-0003-0794-0444.

SMAs have between 4 and 64 microphones. The sequential SMA measurements on a dense grid, on the other hand, are very time-consuming, which is similar to the dummy head measurements.

Thus, the spatial interpolation of sparsely measured BRIRs, as well as the calculation of BRIRs from the SMAs with a limited number of microphones, introduce audible artifacts in the binaural signals. Hence, the number of spatial sampling points, whose density and arrangement can be specified by a spatial sampling grid of a certain (spatial) order N_{grid}, has a significant influence on the binaural synthesis. This spatial order N_{grid} is strongly related to the spatial resolution, which is why both terms are used interchangeably in this paper. The influence of the spatial order on the binaural synthesis has been investigated in several studies. The listening experiments by Pike (2019, Chap. A.8) showed that the auralization of the HRTFs interpolated in the spherical harmonics (SH) domain up to an order of 35 were indistinguishable to the auralizations of the HRTFs measured at that position. Similar thresholds were found by Arend *et al.* (2021). The studies by Ahrens and Andersson (2019) and Bernschütz (2016) showed that the perceptual differences of the dummy head auralizations and binaural renderings of the SMA data significantly decrease above the SH orders of 7–8. However, so far, no study systematically compared the perceptual influence of the spatial order of the measurement grid of the dummy head and SMA captures on the binaural synthesis. With this work, we intend to further contribute to the understanding of the different influences of the spatial order of the dummy head and SMA auralizations. To comparably scale the spatial resolution of both auralizations, we applied the interpolation of the dummy head measurements in the SH domain and also performed the binaural rendering of the SMA data based on the SH representation of the sound field. Thus, the signal processing of both methods is affected by the same artifacts as a result of the order-limited SH processing but differ in the spatial aliasing effects. Thus, as a major hypothesis, we assume that the SMA renderings require higher spatial orders than the dummy head SH interpolation, which is elaborated in more detail in Sec. II.

The BRIRs usually describe the direct sound incidence, early reflections, and diffuse reverberation, which all contribute to the spatial auditory perception of the room acoustics in different ways (Kuttruff, 1973, Chap. 4.2). Humans mainly use the direct sound for the sound source localization. It is followed by a number of early reflections, which evoke other perceptual effects, e.g., the apparent source width, perceived distance, timbre, and spaciousness (Barron, 1971; Olive and Toole, 1988). The diffuse sound field is defined by the uniform sound pressure and incident intensity distribution (Jeong, 2016). Thus, in an ideally diffuse sound field, reverberation has no perceivable directional component but contributes to the perception of other room acoustical features, i.e., the perceived room size or listener envelopment. Hence, the accurate spatial perception becomes less important for the three successive BRIR parts. Engel *et al.* (2021) already showed that when presenting the direct sound with high spatial resolution, the spatial order of the reverberation part can be reduced significantly. However, the study by Engel *et al.* (2021) is based on the order-limited SMA renderings, and they only examined the direct sound and reverberation separately. In this study, we investigate how the limitation of the spatial resolution of each single BRIR component affects the overall perception of the binaural auralizations.

In two listening experiments, we determined the minimum number of sampling points required to achieve the auralizations indistinguishable from a reference auralization based on the high spatial resolution measurements. The grid order N_{grid} is a suitable parameter to scale the number of sampling points and determine the minimum thresholds. We evaluated each part of the BRIRs separately to investigate if N_{grid} has a varying influence on the different BRIR parts. In two adaptive forced choice ABX listening experiments, the participants had to compare the horizontally head-tracked dynamic binaural synthesis based on the measurements on the sparse grids with orders $N_{\mathrm{grid}} = 1$ to $N_{\mathrm{grid}} = 28$ to a high-spatial resolution reference based on the measurements on a 29th-order grid. With experiment 1, we examined the spatial order of SMA measurements, and with experiment 2, which has partly been published in Lübeck *et al.* (2020b), we examined the spatial order of the dummy head measurements. With these experiments, we tested our hypotheses that (1) the dummy head auralizations require a lower spatial order, i.e., less measured sampling points than the SMA auralizations to be perceptually indistinguishable to the high-resolution reference auralizations and (2) the early reflections and reverberation require a significantly lower spatial order, i.e., less sampling points than the direct sound. Although the study and test design are motivated by our two main hypotheses, we were further interested in how various aspects of the binaural reproduction affect the required spatial resolution. We, hence, performed a broad exploratory listening experiment involving different rooms, audio contents and source positions.

II. THEORY

This section briefly summarizes the two methods for accquiring the binaural signals examined in the listening experiments. First, we will outline the concept of binaural synthesis of the SMA data, which is followed by the SH interpolation of the dummy head measurements. Finally, the signal processing of both methods and the undersampling errors resulting from the limited spatial resolution are compared.

A. Binaural synthesis from SMA captures

The signal processing for calculating binaural signals from the SMA captures has been intensively discussed in the literature, and for more details, the reader is referred to, for expample, Bernschütz (2016) or Rafaely (2015).

The sound field S, which has been sampled on the spherical surface Ω of a microphone array with a radius r, is transformed to the SH domain, applying the spatial Fourier transform (SFT) (Williams, 1999)

https://doi.org/10.1121/10.0009277

$$S_{nm}(\omega) = \int_{\Omega} S(\phi, \theta, \omega)\, Y_n^m(\theta, \phi)^*\, \mathrm{d}A_{\Omega}, \tag{1}$$

where ϕ is the horizontal angle ranging from 0 to 2π, θ is the vertical angle ranging from 0 to π, ω is the angular frequency, and dA_{Ω} is an infinitesimal surface element of Ω. Y_n^m are the surface SH of a certain degree n and mode m, and $(\cdot)^*$ denotes the complex conjugate. With a set of order-dependent radial filters d_n, the radial portion, which is introduced by the SMA body, is removed from the sound field. In this way, the sound field S can be decomposed into a continuum of plane waves D, impinging from all directions, which is known as the plane wave decomposition,

$$D(\phi, \theta, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n\, S_{nm}(\omega)\, Y_n^m(\phi, \theta). \tag{2}$$

A HRTF $H(\phi, \theta, \omega)$ is the spatiotemporal transfer function of a plane wave to the listeners' ears. Weighting every sound field plane wave D with the corresponding HRTF from that direction and integrating over the entire surface yields the binaural signals $B(\omega)$, which a listener would be exposed to at the point of the SMA,

$$B(\omega) = \frac{1}{4\pi} \int_{\Omega} H(\phi, \theta, \omega)\, D(\phi, \theta, \omega)\, \mathrm{d}A_{\Omega}. \tag{3}$$

Because the mathematical representation is the same for the left and right ears, for simplification, we omitted the related subscripts throughout this paper. The real-world microphone arrays sample the sound field at discrete positions with a limited number of microphones Q. Consequently, the integrations in Eqs. (1) and (3) become a finite summation, and the plane wave decomposition can solely be calculated up to a certain SH order (Rafaely, 2015). The perceptual consequences of the order-limited plane wave decomposition on the binaural synthesis are mainly degradations of the localization and spaciousness as well as the spectral distortions. These artifacts are discussed, for example, in Ben-Hur *et al.* (2018) or Lübeck *et al.* (2020b) in more detail.

The discretization of Eq. (3) also implies that the sound field can only be decomposed into a limited number of plane waves for the discrete directions. Convolving the limited number of plane waves with the respective head-related impulse responses (HRIRs) is known as the *virtual loudspeaker approach* (Jot *et al.*, 1999). As shown by Bernschütz *et al.* (2014), Ben-Hur *et al.* (2018), or Zaunschirm *et al.* (2018), it is beneficial to perform the convolution with the HRIRs for the plane wave directions on a grid of matched order. The virtual loudspeaker approach can be regarded as the baseline method for binaural decoding, which is why we applied this method in this study.

B. SH interpolation

Nowadays, it is very popular to interpolate the sparsely measured HRTF sets to dense sets in the SH domain (Arend *et al.*, 2021; Aussal *et al.*, 2013; Ben-Hur *et al.*, 2019a; Pörschmann *et al.*, 2020). As this method can be transferred to the BRIRs, we applied the SH interpolation to the sparsely measured dummy head BRIRs in this study. The binaural room transfer functions (BRTFs) are transformed to the spatially continuous SH domain using the SFT [Eq. (1)]. With the inverse spatial Fourier transform (ISFT),

$$B(\phi, \theta, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_{nm}(\omega) Y_n^m(\theta, \phi), \tag{4}$$

the BRTFs for the arbitrary directions can be calculated. Again, a limited number of sampling points negatively affects the SH representation and introduces audible artifacts. According to Rafaely (2015), these undersampling artifacts increase in magnitude above the spatial aliasing frequency $f = Nc/2\pi r$, where c is the speed of sound.

C. Comparison of undersampling errors

For the SMA captures and BRIR SH interpolation, the grid order limits the SH presentation and mainly degrades the high-frequency components. When interpolating the BRIRs measured with a dummy head, each individual measurement has the maximum spatial resolution and the accurate timbre and, therefore, accurately encodes the entire room information. Here, the artifacts arise from a single back-and-forth SFT, which introduces the SH interpolation errors (Ben-Hur *et al.*, 2018).

On the other hand, the SMA renderings are based on an undersampled sound field, which suffers from spatial aliasing. Moreover, the order-limited SFT of the sound field further impairs the SH representation, leading to plane waves with impaired spectra and blurry spaciousness. This plane wave sound field of limited spatial resolution is then convolved with a set of HRIRs of matched order.

Thus, for both of the methods, the spatial order of the sampling grid degrades the SH presentation of the sound field, resulting in the same artifacts on the rendering side. However, for the SMA renderings, additional spatial aliasing artifacts arise when capturing the sound field. This becomes mathematically clear when considering a sound field consisting of a single plane wave. Substituting a single (ideal) plane wave into Eq. (3) and rearranging it yields a perfectly sampled HRTF from the direction of the plane wave (the derivation can be found in Appendix A). To illustrate this, Fig. 1 depicts the different influences of the spatial undersampling.[1] It shows the binaural signals calculated from a single plane wave impinging from the frontal direction. For the following, we employed the Neumann KU100 HRTF set provided by Bernschütz (2013). As a reference, a HRTF directly measured for the frontal direction is shown as a dark gray curve. As an HRTF describes the transformation of one plane wave to the human ears, this represents the ideal case of an artifact-free rendered plane wave from the corresponding direction. The binaural signal depicted as the red curve represents the real-world SMA case. For this, we simulated the plane wave, which was spatially sampled

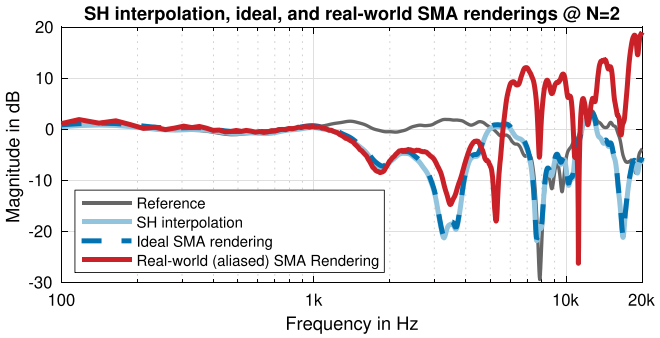

FIG. 1. (Color online) The binaural signals, resulting from the SMA renderings of a simulated plane wave impinging on an ideal virtual SMA (bright blue line), impinging on a real-world SMA with six microphones (red line), and resulting from the second-order SH interpolation of the Neumann KU100 HRTF set (dashed blue line).

at positions of a second-order Lebedev grid. Based on this sampled plane wave, we calculated the binaural signals as described in Sec. II at an order of $N = 2$. The dashed blue curve illustrates the binaural signal resulting from an ideal SMA. For this, again, we simulated an ideal plane wave, which was not sampled by a SMA, and applied the binaural rendering (again, at an order of $N = 2$). These binaural signals are not affected by the spatial aliasing, which occurs when spatially sampling the sound field with the SMA. They are only impaired by the interpolation errors introduced by the order-limited SH processing. Last, the binaural signal depicted as the bright blue line shows the SH interpolation case. For this, we resampled the HRTF set to a second-order Lebedev grid. We then transformed the resampled HRTF set to the SH domain at an order of $N = 2$ and inverse transformed it for the frontal direction. It can be seen that the binaural signal resulting from the ideal SMA (dashed blue curve) is identical to the signal from the SH interpolation (bright blue curve). The real-world SMA binaural signal is notably more impaired. This example shows that on the rendering side, the SH interpolation and SMA rendering are impaired by the order-limited SH processing to the same extent. The binaural signals from the real-world SMA renderings, however, are further impaired by the spatial aliasing on the capturing side. The aliasing and truncation errors are mathematically derived in Ben-Hur *et al.* (2019b). It is worthwhile to mention that for the ideal and real-world SMA renderings in this example, we used ideal radial filters. Thus, the plot only shows the nonideal behaviour of the real-world SMA renderings in terms of the undersampling errors and neglects the constraints of the nonideal radial filters such as the soft-limited radial filters (Bernschütz *et al.*, 2011b).

III. EXPERIMENT 1: AURALIZATION OF SMA DATA

In the first listening experiment, we determined the minimum grid order of the sparse SMA sampling grids, which results in binaural auralizations that are indistinguishable from the auralizations of the SMA data measured on a 29th-order grid. We determined this minimum order as the point of subjective equality (PSE) in an adaptive ABX listening experiment.

A. Method

1. Participants

A total of 36 participants, 29 males and 7 females with a mean age of 24.6 years old [standard deviation (SD), 5.4 years], took part in the listening experiment. Most of them were media technology students, and all of them had self-reported normal hearing.

2. Setup

We applied the dynamic binaural synthesis using the SoundScape Renderer (Geier *et al.*, 2008, 2019). It convolves a set of BRIRs with arbitrary anechoic input signals according to the listener's head orientation, which was tracked with a Polhemus Fastrak (U - 05446-Vermont, US) at a sampling rate of 120 Hz. The experiments were performed in the anechoic chamber of TH Köln with a background noise level of less than 20 dB(A). We used an RME Fireface UFX (D-85778 Haimhausen, Germany) as a digital-analog converter at 48 kHz and a buffer size of 256 samples and Sennheiser HD600 headphones (DE - 30900 Wedemark, Germany) for playback with a playback level of about 66 dB(A). We equalized the binaural chain of the Neumann KU100 dummy head (DE-10117 Berlin, Germany) and Sennheiser HD600 headphones using a 2048 tap minimum phase compensation filter designed according to a regularization method proposed in Erbes *et al.* (2017). The test was implemented and performed with the MATLAB software Scale (Giner, 2013).

3. Stimuli

a. Employed data. For the listening experiments in this study, we used the SMA impulse responses captured in four different rooms at the WDR broadcast studios (Stade *et al.*, 2012). The impulse responses were sampled on a 1202 node Lebedev grid, which allows the SH representation up to the 29th order. At an order of $N = 29$, the spatial aliasing and SH order truncation artifacts can be neglected up to approximately 18 kHz (with a radius of 0.0875 m and a speed of sound of 343 ms^{-1}; Rafaely, 2015, p. 80). Thus, the 29th-order SMA captures are well-suited as the high-spatial resolution ground truth in this study. The database consists of the measurements in the four different rooms with different reverberation times as presented in Table I. For synthesizing the binaural signals, again, we used the Neumann KU100 HRTFs (Bernschütz, 2013).

Because we intended to investigate the spatial resolution separately for the three parts of the BRIR, i.e., the direct sound, early reflections, and reverberation, we defined the transition times of these parts as follows. The direct sound corresponds to the duration of a HRTF measured in anechoic conditions (Blauert, 1996; Møller *et al.*, 1995; Zahorik, 2002) and is approximately 2.5–3.5 ms. For some cases, it is difficult to separate the first floor reflection from

https://doi.org/10.1121/10.0009277

TABLE I. The RT_{60} and the transmission times at which the early reflections and the reverberation part start for all of the rooms examined. RT_{60}, The reverberation time 60 (500 Hz and 1 kHz).

Room	RT_{60}	Early reflections starting time	Reverberation starting (mixing) time
Control room 1 (CR1)	<0.25 s	3.5 ms after onset	71.02 ms
Control room 7 (CR7)	<0.25 s	3.5 ms after onset	39.03 ms
Small broadcast studio (SBS)	0.9 s	3.5 ms after onset	43.34 ms
Large broadcast studio (LBS)	108 s	3.5 ms after onset	46.22 ms

the direct sound, which is why we decided to define the duration of the direct sound for all rooms as 3.5 ms. The direct sound is followed by a number of early reflections, and at the so-called mixing time, the number of reflections has increased such that the sound pressure is equally distributed over the entire room, and the sound field can be considered as diffuse. Different methods to estimate the mixing time have been proposed in the literature. A comprehensive comparison and evaluation of the various methods has been presented in Lindau *et al.* (2010). In this study, the mixing times have been estimated with a procedure introduced by Abel and Huang (2006) and as proposed in Lindau *et al.* (2010), which is averaged across the left and right ear signals for the frontal direction with a window length of 20 ms and a safety margin of 100 samples. The resulting mixing times are presented in Table I.

b. Simulation of sparse measurements. The reference BRIRs were calculated from the 29th-order Lebedev grid SMA impulse responses according to Eq. (3). To simulate the SMA measurements on the sparse grids, we spatially resampled the 29th-order grid to order $N = 1$–28 Gauss grids by interpolation in the SH domain. The SMA impulse responses are transformed to the SH domain up to the maximum order of 29. Subsequently, the ISFT [Eq. (4)] is applied to yield the 28 SMA impulse response sets defined for the sampling directions according to $N = 1$–28 Gauss grids. This procedure results in sparse SMA impulse responses, which suffer from spatial aliasing and SH order truncation, as would be the case with measurements with the real-world SMAs. In contrast, truncating the SH order series of the SH representation of the 29th-order grid would solely lead to the SH order truncation artifacts. Gauss grids are rather inefficient, i.e., they need a relatively large number of sampling points to resolve certain SH orders. However, in contrast to the more efficient grids, such as the Lebedev or Fliege grid, Gauss grids are defined for every grid order. Therefore, we decided to use Gauss grids to scale the spatial resolution in steps of one order. The influence of different grids has been discussed, for example, in Rafaely (2015, Chap. 3), Zotter (2009, Chap. 4.2), or Bernschütz (2016, Chap. 3.2.2).

From each of the resampled sparse SMA impulse responses, we calculated the BRIRs as described in Sec. II up to the corresponding SH order. According to the virtual loudspeaker approach, for each BRIR rendering, the plane wave components were calculated for the Gauss grids of corresponding order and weighted with the HRTFs for these directions. The plane wave components were calculated with radial filters with a 20 dB soft limit (Bernschütz *et al.*, 2011b). As an alternative to the rotation of the entire sound field in the SH domain, the BRIRS for the different listener head orientations can be synthesized by accounting for the head orientation when selecting the HRTFs for the plane wave incident directions [in Eq. (3)]. This procedure has been discussed by Bernschütz (2016, p. 66) in more detail. Besides the reference BRIRs, we calculated 28 BRIR sets based on 28 SMA impulse response sets of order 1–28. Each of the BRIR sets were calculated for 360 directions in 1° steps along the horizontal plane. However, in the listening experiment, the binaural synthesis was adapted according to the listeners' head orientations only for ±60° along the horizontal plane to save the working memory. The signal processing was performed in MATLAB using the SOFiA toolbox (Bernschütz *et al.*, 2011a). The block diagram in Fig. 2 presents an overview of the signal processing.

c. Splitting the BRIRs. To determine the minimum grid order for each BRIR component separately, we split each BRIR set at the transmission times specified in Table I. This resulted in the direct sound part, consisting of the first 3.5 ms, the early reflection part up to the mixing time, and the final reverberation part. Subsequently, we recomposed the sparse BRIR sets with a reduced spatial resolution in (a) just the reverberation (REV) part, (b) just the early reflections and reverberation parts (ER), and (c) in all three parts, resulting in the BRIRs being completely reduced in their spatial resolution (DS). To ensure the artifact-free recomposition, we applied linear fading over 128 samples between the parts. In the following, these BRIRs are denoted as

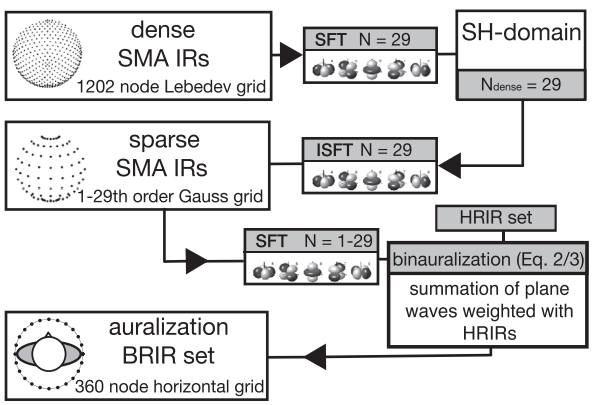

FIG. 2. The block diagram of the signal processing for the generation of the BRIRs based on the sparse SMA impulse responses examined in experiment 1.

Rev-BRIRs (reduced spatial resolution starting at the mixing time), *ER-BRIRs* (reduced spatial resolution starting at the early reflections), and *DS-BRIRS* (reduced spatial resolution starting at the direct sound).

d. Test signals and sound source positions. Because several studies (Ahrens and Andersson, 2019; Arend *et al.*, 2021; Pörschmann *et al.*, 2019; Zaunschirm *et al.*, 2018) showed that lateral sound sources are perceptually more critical than frontal sources, we presented one sound source from the side at $\phi = 270°$. Further, we examined a second position at $\phi = 30°$ to present a frontal source that induces the interaural time and level differences. As anechoic test signals, we used a pink noise burst with a length of 0.75 s (including 10 ms cosine-squared onset/offset ramps) and a male speech sample, consisting of a short German sentence with a length of 1.5 s.

4. Procedure

The ABX three-interval/two-alternative forced choice (3I/2AFC) test design is a simple, robust, and widely used paradigm in psychophysics. In combination with the adaptive one-up one-down staircase procedure (Kingdom and Prins, 2010; Levitt, 1971, Chap. 3 and 5), it is well-suited to determine the so-called PSE (Meese, 1995). The PSE, also denoted as the threshold of recognition, is the 50% point on the psychometric function. It defines the point at which no relevant differences can be detected anymore, as in the present case, the differences between the BRIR auralizations based on the sparse and dense SMA measurements.

According to the ABX test paradigm, three intervals (A, X, and B) were presented to the participants in each trial. Two of the intervals consisted of the same stimuli (i.e., auralizations based on the BRIR with the same spatial resolution). Either the reference stimulus (based on the high-resolution BRIR) or the stimulus of the lower-resolution BRIR was assigned randomly to the middle interval X. Accordingly, either A or B consisted of the same stimuli as X. This assignment ensures the (1) direct comparison of the stimuli with different spatial resolutions and (2) that either the lower or higher resolution was presented two times randomly. After the presentation of the three intervals A, X, and B, the participants were asked to decide if A or B equaled X by pressing the corresponding button on the experiment graphical user interface (GUI).

Following the adaptive one-up-one-down staircase method, if the participants were correct and could indicate the difference between the intervals, the stimulus based on the BRIR with the next higher spatial order was assigned in the next trial. If they gave a wrong answer, i.e., they could not indicate any differences, the BRIRs based on the next lower grid were picked for the next trial. Each run started with the low resolution stimuli of order $N_{\text{grid}} = 1$ and was terminated after 12 reversals. One reversal is defined as a correct decision followed by a wrong decision or vice versa. Each A, B, X sequence was automatically played back one time and, during the playback, the participants were free to move their heads if it helped them to distinguish between the sparse and dense BRIR auralizations.

According to a $4 \times 3 \times 2 \times 2$ mixed factorial design with the between-subject factor room (CR1, CR7, SBS, LBS) and the within-subject-factors BRIR component (DS, ER, REV), source position ($\phi = 30°$, $\phi = 270°$), and test signal (noise, speech), the participants were divided into four groups with nine participants each. The participants from each group performed 12 runs in total (3 BRIR components, 2 source positions, and 2 test signals). To ensure that the subjects fully understood the task, each participant was introduced to the experimental design at the beginning of the experiment. Afterward, each participant had to perform two training runs to get familiar with the test procedure. The training runs consisted of the speech test signal at position $\phi = 30°$ and the noise test signal at position $\phi = 270°$ for the corresponding room. The training runs were completed after four reversals or five wrong decisions at the lowest spatial resolution. The duration of the experiment varied depending on the group and participants. On average, the experiment took about 50 min, including a short break.

B. Data analysis

The PSEs were determined as the averaged grid order N_{grid} over the last nine reversals. We, thus, omitted the first three reversals. Because the rendering requires an integer number for the grid order N, we employed discrete values for the statistical analysis. According to Yap and Sim (2011) or Bee Wah and Mohd Razali (2011), the Shapiro-Wilk test is powerful for testing the assumption of the normality distribution, which is why we decided to use it in this study. Moreover, Chen *et al.* (2017) presented a comprehensive comparison of the different adjustments for multiple testing. We decided to use the rather conservative Bonferroni method for all of the adjustments. A Shapiro-Wilk test with Bonferroni correction showed no violations of the assumption of the normality distribution of the data. Therefore, we analyzed the PSEs with a four-way mixed analysis of variance (ANOVA) with the between-subject factor room (CR1, CR7, SBS, and LBS) and within-subject factors BRIR component (DS, ER, REV), test signal (noise, speech), and source position (30°, 270°). A Mauchly test for the sphericity revealed that for the factor BRIR, the component sphericity was not met, which is why we applied the Greenhouse-Geisser correction where applicable. Girden (1992) proposed to use the Greenhouse-Geisser correction for $\varepsilon \leq 0.75$. The ANOVA for experiment 2 revealed $\varepsilon \leq 0.75$. Therefore, we decided to apply the Greenhouse-Geisser correction for all of the applied tests. For a more detailed analysis, we further applied the various *post hoc* Bonferroni corrected independent-samples t-tests.

C. Results

A graphical overview of the results is presented as boxplots in Fig. 3. First, it can be seen that the PSEs become

https://doi.org/10.1121/10.0009277

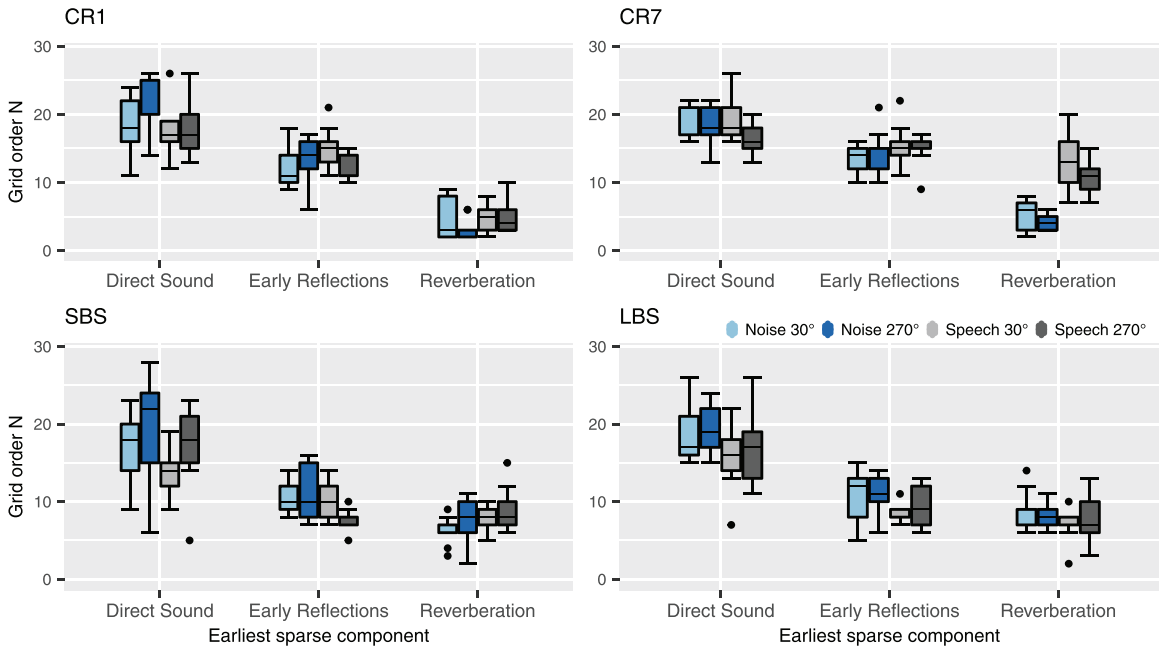

FIG. 3. (Color online) The interindividual variation in the determined PSEs (grid orders N) for the tested rooms CR1, CR7, SBS, and LBS with respect to the earliest sparse component (x axis) are shown for experiment 1. The sound source position and test signal are depicted separately as indicated by the colors. Each box shows the interquartile range (IQR), median value (black line), outliers (black points), and black whiskers, displaying the 1.5 × IQR below the 25th or above the 75th percentile. Note that in some cases, the median is exactly on the upper or lower IQR. The median of the CR1 direct sound noise at 270° is 20, the median of theCR1 early reflections speech at 270° is 17, the median of the CR7 early reflections noise at 270° is 12, and the median of the LBS reverberation noise at 30° is 9.

smaller for the successive BRIR components. The reverberation part of the CR7 room is an exception as it shows the relatively large PSEs for the speech test signal. Furthermore, it can be observed that for the direct sound part, the medians of the PSEs for the more critical test signal noise are always higher. For the more reverberant rooms SBS and LBS, the medians of the PSEs for the lateral sound source position at 270° are higher than those for the frontal 30° condition. The boxplots further indicate the PSE outliers. The highest PSE of 28 was detected for the direct sound part of the SBS at the lateral position and noise test signal. It is worth mentioning that this PSE of 28 is below the upper whisker and was, thus, not indicated as an outlier.

The results of the four-way mixed ANOVA are shown Table II of Appendix A. The significant main effect of the BRIR component together with the observation from Fig. 3 indicates a strong dependency of the BRIR component on the required grid order. The ANOVA further revealed a significant main effect of the room as well as the interaction effects of room × BRIR component, room × source position, and room × signal. These significant differences of the room might be due to the exception of CR7 in the reverberation part. The interaction of BRIR component × signal shows that the test signal has varying influence on the PSE for the different BRIR components. Although the position is not a significant main effect, it has a varying influence with respect to the room as indicated by the interaction effect of position × room. Last, we found the significant interaction effect of room × BRIR component × signal. A *post hoc* power-analysis with G*power (Erdfelder *et al.*, 2009) based on the calculated Cohens' f values revealed an achieved power ≥0.9 for all of the significant effects.

To further investigate the significant effect of the room, we applied a series of *post hoc* independent-samples t-tests between the pooled data of each room. Only the t-tests between the data for the room CR7 and SBS and CR7 and LBS were significant, which supports the assumption that the effect of the room is due to the exception in CR7 (the results of all of the t-tests are displayed in Appendix B 1).

For further inspection of the significant effect of the BRIR component, Fig. 4 displays the mean values for each room and the BRIR component separately pooled over the signal and position. It can be seen that the means between all of the BRIR components vary for each room. Only for the more reverberant rooms SBS and LBS, the visual inspection does not indicate a clear difference between the early

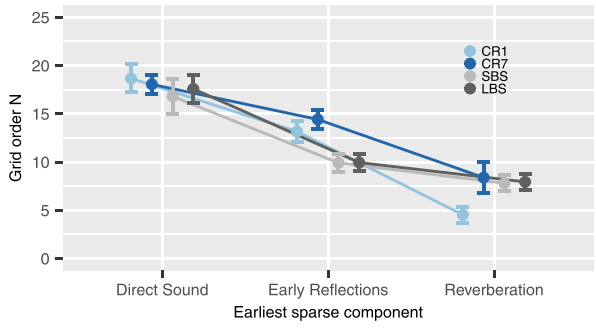

FIG. 4. (Color online) The marginal mean plot with respect to the room and BRIR pooled over the position and signal, including 95% within-subject confidence intervals, is shown for experiment 1.

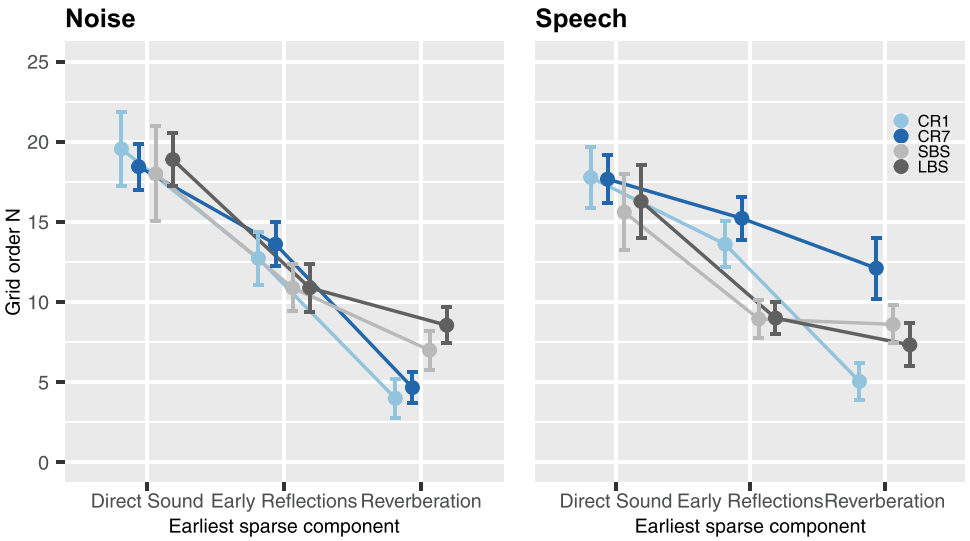

FIG. 5. (Color online) The average PSEs pooled over the signal with respect to the earliest sparse BRIR component (x axis) and room (color) for the frontal and lateral positions separately and the 95% within-subject confidence intervals are shown for experiment 1.

reflections and reverberation part. However, a t-test showed that there is a significant difference between them for both rooms. This suggests that there is a significant difference between each BRIR component for all of the rooms, and the required spatial orders descend over the successive BRIR components. The crossing of the interaction line of the room CR1 with the LBS and SBS illustrates the interaction between the BRIR component and the room. For the later BRIR components, CR1 seem to require significantly less spatial orders than CR7, SBS, and LBS. Probably, this is simply due to the higher estimated mixing time in CR1 (see Table I) and, thus, a shorter reverberation time part. A *post hoc* nested ANOVA involving the PSEs of the direct sound part with the between-subject factor room and the within-subject factors source position and test signal only showed a significant effect of the test signal $[F(1,32) = 8.36, p < 0.007, \eta_p^2 = 0.21, \varepsilon = 1.0]$, which suggests that the dependency of the factor room might be attributed to the later BRIR parts.

For further investigation of the interactions of the room × BRIR component, room × signal, and room × position, we performed *post hoc* nested ANOVAs with the within-subject factors BRIR component, source position, and test signal for each room separately (see Tables III–VI in Appendix B 1). Each ANOVA revealed a significant main effect BRIR component. Only for CR7, we found a significant effect of the position. This indicates that CR7 causes the interaction of the room × position. For all of the other rooms, the position has no significant influence on the required spatial order. This can also be seen in Fig. 5, which displays the mean values with respect to the room and BRIR pooled over the signal for both positions separately. For CR7 and LBS, we further found a significant effect of the signal. This indicates that CR7 and LBS cause the significant interaction effect of the room × signal. This is strongly supported by Fig. 6, which displays the mean values with respect to the room and BRIR component pooled over the positions for both signals separately. The reverberation part

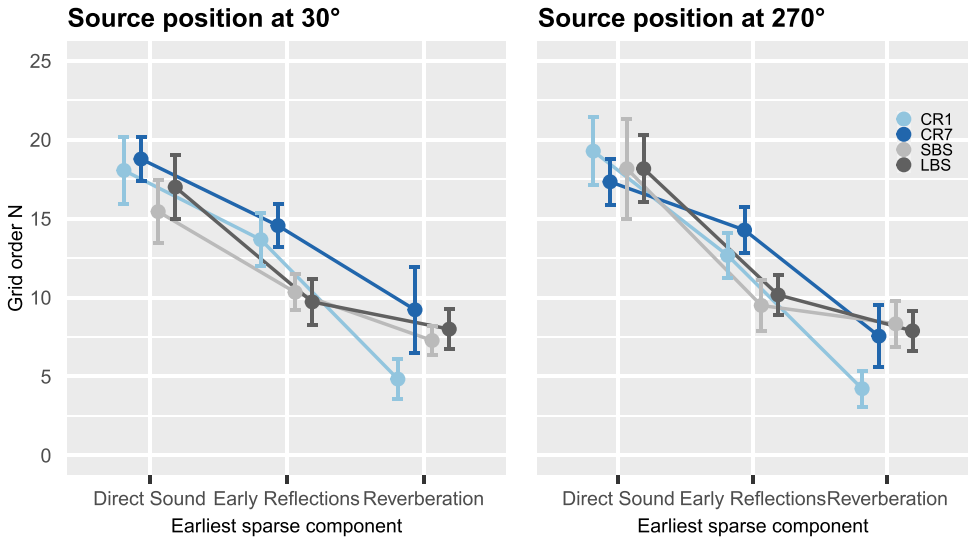

FIG. 6. (Color online) The average PSEs pooled over the position with respect to the earliest sparse BRIR component (x axis) and room (color) for the frontal and lateral positions separately and the 95% within-subject confidence intervals are shown for experiment 1.

https://doi.org/10.1121/10.0009277

of CR7 requires significantly higher orders for the speech than for the noise signal. For CR7, SBS, and CR1, we found the significant interaction BRIR comp × signal. They cause the three-way interaction room × BRIR component × signal. The p*ost hoc* paired *t*-test between both signals for each room and BRIR component separately revealed that for only the reverb part of CR7, the signal leads to a significant difference in the PSEs.

For a better overview and comparison with the results of experiment 2, Table VIII (Appendix B 1) displays all of the mean values across the subjects with respect to the BRIR component, test signal, and source position. Although the factor room had a significant effect, we decided to pool the data of all of the rooms. Thus, the interpretation of the mean values, at least for the early reflections and the reverberation, should be performed with reservation.

Based on the results of experiment 1, we, thus, conclude as follows. The minimum required spatial resolution varies and mostly decreases for the three successive BRIR components, which supports our second main hypothesis. For the direct sound, the average PSEs range from 17 to 20 for the early reflections from 12 to 13 and for the reverberation from 7 to 9. The PSEs of each BRIR component vary significantly. We could observe a significant influence of the auralized room, whereby there are indications that this dependency is evoked by the later BRIR parts of the early reflections and reverberation. At this point, it should be noted that different participants with different experiences, for example, in spatial audio, might result in different test results. Thus, especially for the between-subject test designs, a significant effect of the between-subject factor (such as, in this case, the room) could be due to the participant group.

IV. EXPERIMENT 2: AURALIZATION OF DUMMY HEAD DATA

In the second listening experiment, we investigated the dummy head auralizations. We determined the minimum number of sampling points, which after interpolation in the SH domain results in auralizations that are indistinguishable to the auralizations of the reference measurements on a dense 29th-order grid. The results of this listening experiment have partly been presented in Lübeck *et al.* (2020b). The setup, test design, and procedure were exactly the same as for those in experiment 1.

A. Method

We were interested in a later comparison of the results of both experiments. To have a balanced number of observations, we extended the data by four participants compared to the results presented in Lübeck *et al.* (2020b). A total of 36 participants, 25 male and 11 female, took part in the listening experiment (mean age = 28.1 years old, SD = 7.4 yr). Most of them were media technology students and all had self-reported normal hearing.

The stimuli for this experiment are based on the same impulse responses as in experiment 1. Because the employed database does not contain a full-spherical dummy head BRIR measurement set on a 29th-order grid, the high-spatial resolution references were also calculated from the 29th-order SMA impulse responses for 1202 head orientations of a 29th-order Lebedev grid, according to Eq. (3). The so-computed BRIRs are nearly perceptually equivalent to the BRIRs directly measured with a dummy head as has been extensively discussed and evaluated, e.g., in Ahrens and Andersson (2019) and Bernschütz (2016). Therefore, we considered these BRIR sets as the high-resolution ground truth.

To simulate the sparse dummy head measurements, we transformed these 29th-order Lebedev BRIR sets to the SH domain at the maximum order of 29. Subsequently, applying the ISFT, we resampled the dense set to 28 BRIR sets defined for the sampling directions according to $N = 1$–28 Gauss grids. Finally, for the continuous dynamic binaural synthesis, all of the sparse BRIR sets were transformed to the SH domain again, this time with the corresponding order of the sparse BRIR set, and then resampled to a 360 sampling point grid with 1° steps. We split the BRIRs in the direct sound part, early reflections part, and reverberation part, and recombined them exactly as was done for experiment 1. Again, the binaural synthesis was adapted according to the listeners' head orientations only for ±60° along the horizontal plane. The entire signal processing workflow is illustrated as a block diagram in Fig. 7. In all other aspects, the setup, materials, procedure, and analysis were identical to those of experiment 1.

B. Results

A Bonferroni corrected Shapiro-Wilk test rejected the hypothesis of normal distribution in 1 of 48 conditions. However, the parametric tests are robust to slight violations of normality assumptions (Bortz and Schuster, 2010; Pearson, 1931). Therefore, for the statistical analysis, again, we applied a four-way mixed ANOVA with the between-subject factor room and the within-subject factors BRIR

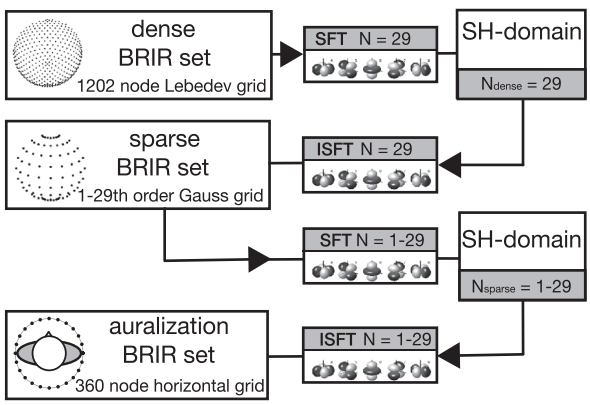

FIG. 7. The block diagram of the signal processing for the generation of the sparse BRIR sets in experiment 2.

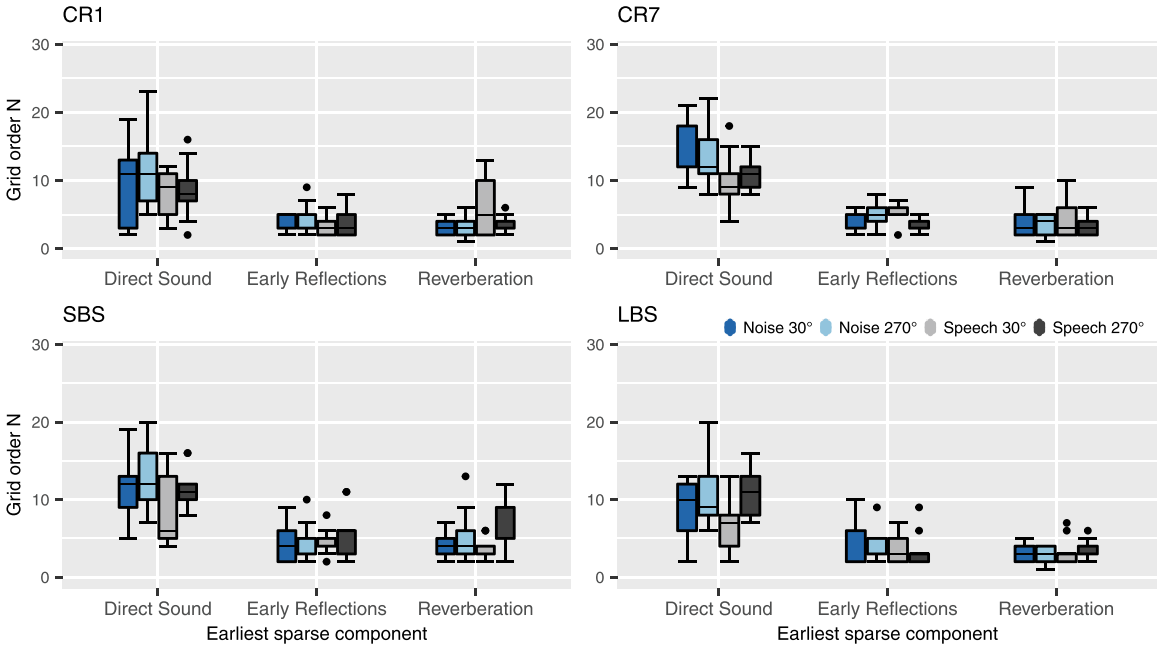

FIG. 8. (Color online) For experiment 2, the interindividual variation in the determined PSEs (SH orders N) for the tested rooms CR1, CR7, SBS, and LBS with respect to the earliest sparse component (x axis). The sound source position and test signal are depicted separately as indicated by the colors. Each box specifies the interquartile range (IQR), median value (black line), outliers (black points), and black whiskers, displaying the 1.5 × IQR below the 25th or above the 75th percentile. Again, in some cases, the median is exactly on the upper or lower IQR. The median of the CR7 direct sound noise at 30° is 12, the median of the CR7 early reflections noise at 270° is 3, the median of the SBS early reflections noise at 270° is 5, the median of the SBS early reflections speech at 270° is 3, the median of the SBS reverberation speech at 270° is 5, and the median of the LBS early reflections noise at 30° is 2, and the median of the LBS early reflections noise at 270° is 3.

component, test signal, and sound source position. Just as with experiment 1, the Mauchly test rejected the assumption of sphericity for the factor BRIR component, and we applied the Greenhouse-Geisser correction. Figure 8 presents an overview of the results of experiment 2. For each room, the PSEs significantly decrease for the BRIRs with a limited resolution of the early reflections and reverberation. Between the early reflections and reverberation, we cannot observe a difference by visual inspection. For all of the rooms, noise was the more critical test signal for the direct sound. This dependency of the test signal seems to become smaller for the early reflections and reverberation part. It can be seen that for all of the rooms, none of the participants could detect differences of the grids with orders higher than 23. The absolute maximum value of all of the rooms was 23 for CR1 with the noise test signal at 270°.

The results of the four-way mixed ANOVA are displayed in Appendix B 2, Table IX. Because we could neither observe any significant main effect of the room nor any interaction effect involving the factor room, we pooled the data over the room for Fig. 9, which supports the results of the ANOVA.

The ANOVA revealed a significant main effect of the BRIR component, which is strongly supported by the observations from Figs. 8 and 9. The significant effect of the source position also matches the observation from Figs. 8 and 9 and shows that the lateral source positions mostly required higher grid orders than frontal grid orders. Furthermore, we found significant interaction effects of the BRIR component × source position and BRIR component × signal. This indicates that the signal and source position have varying influences on the PSE with respect to the BRIR component, which was already observed in Figs. 8 and 9.

To disentangle the interaction effects, we applied a series of Bonferroni corrected independent-samples t-tests between the data of positions 1 and 2 for each BRIR component separately. For the direct sound only, we found a significant difference between the frontal and lateral source positions. We performed the same t-tests between the noise and speech test signal and also found that only for the direct sound, the PSEs of the noise and speech signal differ significantly. This supports the assumption that for the direct sound, the source position and signal have an influence on the PSE but this is not the case for the later BRIR components.

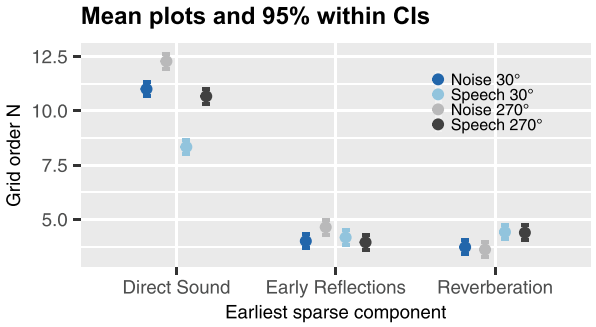

FIG. 9. (Color online) For experiment 2, the average PSEs pooled over all of the rooms with respect to the earliest sparse BRIR component (x axis) are shown. The 95% within-subject confidence intervals were calculated according to Jarmasz and Hollands (2009) and Loftus (1994). The test signal and sound source positions are displayed separately as indicated by the colors.

https://doi.org/10.1121/10.0009277

Moreover, we performed the pairwise *t*-tests between all of the BRIR components and found that the PSEs of the early reflections and reverberation parts are not significantly different. The PSEs of the direct sound significantly differ from both.

Similar to experiment 1, Table X in Appendix B 2 shows the mean values across the subjects with respect to the BRIR component, test signal, and source position.

The results of experiment 2 lead to the following assumptions. As similar to experiment 1, the required grid order decreases significantly for the successive BRIR parts but with notable smaller PSEs: 9–13 for the direct sound and 4–5 for the early reflections and the reverberation. Together with the results of experiment 1, this proves the second main hypothesis. We could not detect a significant difference for the PSEs in the early reflections and reverberation. For the direct sound part, the test signal and source position have a statistical influence; for the early reflection and reverberation part, this influence vanishes. We did not observe any statistical significance of the room. Because in experiment 2, the ANOVA did not reveal any significant three-way interaction, we do not show the estimated marginal mean plots as were shown for experiment 1.

The results plots in Figs. 3 and 8 as well as the averaged PSEs in Tables VIII and X suggest that the PSEs for experiment 1 are significantly higher than those for experiment 2. To prove that these differences are statistically significant, we further applied a series of independent-samples *t*-tests (two-sided) with the Bonferroni correction. The *t*-test comparing the pooled PSEs from experiment 1 and experiment 2, respectively, revealed a significant difference in the estimated PSEs between both of the experiments and, thus, between the minimum required SH order for the SMA and dummy head BRIRs [$t(862) = 17.175$, $p < 0.001$, $d = 1.17$]. The separate pairwise comparisons of the PSEs estimated in experiments 1 and 2 for the three different BRIR components (DS, ER, and REV) showed significant differences for all of the BRIR components [DS, $t(286) = 13.79$, $p < 0.001$, $d = 1.63$; ER, $t(286) = 22.661$, $p < 0.001$, $d = 2.67$; REV, $t(286) = 8.78$, $p < 0.001$, $d = 1.03$]. This suggests that the first main hypothesis is also valid.

V. GENERAL DISCUSSION

A. Comparison of both experiments

Both of the experiments show that the required grid order varies and mostly decreases for the later BRIR parts. As expected, the dummy head SH interpolation requires significantly lower grid orders than the SMA renderings. Moreover, it is noticeable that in experiment 1, there is a significant difference between the PSEs of the early reflections and reverberation, whereas for experiment 2 this is not the case. Further, the results of experiment 2 show that for the direct sound part, the test signal and source position have a significant effect, whereas this influence vanishes for the later BRIR part. We conclude that the SH interpolation of the dummy head data mainly affects the direct sound part and seems to produce fewer perceptual artifacts in the later BRIR components. On the contrary, limiting the order of the SMA renderings seems to have more impact on the synthesis of the later BRIR parts. Certainly, it is due to the different signal processing applied for both methods. One explanation could be that the SMA renderings synthesize the BRIRs with a superposition of (order-limited) plane waves. The limitation of the plane wave decomposition results not just in impaired plane waves but also in fewer directions of the impinging plane waves, which might introduce the comb-filtering artifacts. We, thus, assume that in contrast to the dummy head SH interpolation, the SMA renderings require relatively high SH orders to synthesize the diffuseness of the sound field and its timbre. This could also be an explanation for the room dependency of the SMA renderings, which was not observed for the dummy head SH interpolation.

B. Comparison to previous studies

Ahrens and Andersson (2019) presented a listening experiment, comparing the dummy head auralizations to the order-limited SMA auralizations. They found that mostly above orders of eight, the perceptual differences decrease. Our experiments show that even up to an order of 28, the perceptual differences persist. However, in Ahrens and Andersson (2019), the participants rated the difference in terms of the spaciousness and timbre on a quality scale, whereas we examined the overall indistinguishability in experiment 1.

Recently, Engel *et al.* (2021) and Engel *et al.* (2019) presented a comprehensive investigation on the different Ambinsonic-based binaural renderers, in which the direct sound and the reverberation were rendered separately. They found that when the direct sound is auralized with the high spatial resolution, the listeners could hardly distinguish between the first-, second-, third-, and fourth-order binaural reverberations. However, the participants compared to a reference rendering at a SH order of four. Our experiment 1 shows that when comparing to a high-order reference, in our case $N_{\text{grid}} = 29$, orders 6–9 are necessary for the indistinguishability of the reverberation. Unlike Engel *et al.* (2021) and Engel *et al.* (2019), who truncated the SH order of the Ambisonics representation, we resampled the SMA data to the sparse grids. However, the general findings of both of the studies are similar: The reverberation part in the binaural auralizations can be rendered with lower SH orders than the earlier parts. In addition, Engel *et al.* (2021) and Engel *et al.* (2019) did not distinguish between the required spatial orders of the dummy head and SMA data. Our study shows that these orders are significantly different.

In the past studies, there are inconsistent and even contrary observations regarding the room dependency of the SMA renderings. Bernschütz (2016, p. 224), as well as Ahrens and Andersson (2019), did not detect any statistical differences across the rooms. On the contrary, Ahrens *et al.* (2017) reported that reverberant rooms require higher SH

orders than less reverberant rooms for the indistinguishability compared to the dummy head auralizations. Engel *et al.* (2021) also found that certain room characteristics affect the required spatial order. In line with this finding, the ANOVA for our experiment 1 involving all of the rooms showed a significant effect of the room. However, the ANOVA that only considers the direct sound condition did not reveal the room as a significant effect. Furthermore, there are indications that the significant effect of the room is only evoked by the later BRIR components of the room CR7. For the reverberation part of CR7, we found relatively high PSEs. Interestingly, they were detected for the speech signal in the front, although it is the less critical signal at the less critical position. Visual inspection of the impulse responses around this part did not yield any anomalies such as strong reflections. It is worth mentioning that for experiment 2 in the reverberation part of the CR1, which has a comparable RT_{60}, we could observe a similar tendency, again, for the speech signal in the front. However, this observation was not indicated as significant by the ANOVA. We could not find a clear explanation for this interaction of the dry rooms and the speech signal in the reverberation part.

We assume that, in general, there is a weak influence of the room on the binaural SMA renderings, which is certainly more significant in the later BRIR parts.

Experiment 2 indicates that the SH interpolation of the dummy head data is not dependent on the room. In this context, it is interesting to compare experiment 2 to the studies of Pike (2019, Chap. A.8] or Arend *et al.* (2021). Pike (2019) compared the auralizations of the anechoic HRTFs, interpolated in the SH domain to the HRTFs directly measured at that direction. They showed that above an order of 35, no differences were audible anymore. Arend *et al.* (2021) conducted a similar experiment as the present experiment 2 but just for the anechoic HRTFs and reported the PSEs between 13 and 25 for the frontal and lateral noise and speech sound sources. In contrast, our study revealed the PSEs between 9 and 13 for the direct sound. It can, thus, be assumed that in the presence of early reflections and reverberation, the artifacts in the direct sound are perceptually less relevant. Therefore, the SH interpolation of the dummy head impulse responses, which also encode the reflections and reverberation, requires significantly smaller SH orders than the SH interpolation of the dummy head impulse responses, which encode only the direct sound in the anechoic conditions, i.e., the HRIRs.

To determine the PSE thresholds for both of the methods, we used a baseline approach, i.e., the virtual loudspeaker method to synthesize the BRIRs from the SMA data, and the classical SH transform for the interpolation of the BRIRs. However, in the last years, several approaches have been developed to perceptually improve the binaural renderings of the SMA captures, for example, as discussed in Zaunschirm *et al.* (2018) or Lübeck *et al.* (2020a). Also, the interpolation of the BRIRs in the SH domain could be improved by the spectral equalization, matrix regularization, or time-alignment approaches. Therefore, it should be noted that different decoding or interpolation methods may lead to different thresholds. However, the baseline methods allow us to keep the signal processing for both of the methods similar (see Sec. II) and determine the generally valid but rather conservative thresholds.

VI. CONCLUSION

In this paper, we presented two listening experiments with the aim of finding the minimum required spatial orders of the SMA and dummy head BRIR measurements, which result in an auralization that is indistinguishable from a high-resolution reference. We applied the dynamic binaural synthesis, which was adapted only with respect to the horizontal head orientation of the listener. The found thresholds may shift for the full-spherical auralizations. For the horizontally head-tracked auralizations, we could show that the BRIR components encoding the early reflections or the reverberation for the dummy head data and SMA data require fewer sampling points than the direct sound component. Furthermore, the dummy head impulse responses require lower orders than the SMA impulse responses to achieve the perceptually similar binaural auralization. Last, the room has no influence on the interpolation of the dummy head BRIRs in the SH domain, whereas for the SMA renderings, it has an influence. The thresholds can be used to further simplify the data acquisition of the binaural rendering. Furthermore, the computational effort can be reduced enormously when rendering the direct sound, early reflection, and reverberation separately. In this study, we determined the thresholds in terms of the indistinguishability. It can be assumed that the quality-based listening experiments would lead to significantly lower spatial orders.

ACKNOWLEDGMENTS

The authors would like to thank all of the participants for their support as well as Melissa Ramírez and Kai Altwicker for assistance in conducting the listening experiments. We are grateful to the two anonymous reviewers for their constructive comments on previous versions of this manuscript. The work was funded by ERDF (European Regional Development Fund) under the funding reference code EFRE-0801444.

APPENDIX A: UNDERSAMPLING ERRORS IN DUMMY HEAD AND SMA RENDERINGS: MATHEMATICAL DERIVATIONS

In the following, it is mathematically shown that if undersampling artifacts due to the discrete sampling of the sound field are neglected, the spatial interpolation of the dummy head data in the SH domain is equivalent to the binaural rendering of the SMA data. The interpolation of a HRTF set $H(\phi_q, \theta_q)$ to a HRTF $H(\phi_d, \theta_d)$ can be performed by (order-limited) SH transform at an order N [Eq. (A1)] and inverse SH transform [Eq. (A2)],

https://doi.org/10.1121/10.0009277

$$H_{nm}(\omega) = \int_{\Omega} H(\phi_q, \theta_q, \omega) Y_n^m(\theta, \phi)^* \mathrm{d}A_{\Omega}, \tag{A1}$$

$$H(\phi_d, \theta_d, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} H_{nm}(\omega) Y_n^m(\theta_d, \phi_d). \tag{A2}$$

Substituting the plane wave density function calculated with Eq. (2) into the binaural reproduction described with Eq. (3) yields

$$B(\omega) = \frac{1}{4\pi} \int_{\Omega} H(\phi, \theta, \omega) \times \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n j_n\left(\frac{\omega}{c} r_0\right)} S_{nm}(\omega) Y_n^m(\phi, \theta) \mathrm{d}A_{\Omega}. \tag{A3}$$

According to Williams (1999, p. 259), the SH coefficients of a unity plane wave impinging from (ϕ_d, θ_d) are

$$\tilde{S}_{nm}(\omega) = 4\pi \mathrm{i}^n j_n\left(\frac{\omega}{c} r_0\right) Y_n^m(\phi_d, \theta_d)^*. \tag{A4}$$

Inserting $\tilde{S}_{nm}$ for the sound field S_{nm} in Eq. (A3) yields

$$B(\omega) = \int_{\Omega} H(\phi, \theta, \omega) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\phi_d, \theta_d)^* Y_n^m(\phi, \theta) \mathrm{d}A_{\Omega}. \tag{A5}$$

The HRTFs can be expressed as the SH sum $H_{nm}(\omega)$,

$$B(\omega) = \int_{\Omega} \sum_{n=0}^{N} \sum_{m=-n}^{n} H_{nm}(\omega) Y_n^m(\theta, \phi) \times \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\phi_d, \theta_d)^* Y_n^m(\phi, \theta) \mathrm{d}A_{\Omega}, \tag{A6}$$

assuming that there are no undersampling effects resulting from the discrete sampling of the sound field. Hence, the orthogonality property of the SH function holds such that

$$B(\omega) = \int_{\Omega} \sum_{n=0}^{N} \sum_{m=-n}^{n} H_{nm}(\omega) Y_n^m(\theta, \phi) \times \delta(\phi - \phi_d) \delta(\cos(\theta - \cos\,\theta_d)) \mathrm{d}A_{\Omega}. \tag{A7}$$

Resolving the integral leads to

$$H(\phi_d, \theta_d, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} H_{nm}(\omega) Y_n^m(\theta_d, \phi_d), \tag{A8}$$

which is exactly Eq. (A2). Hence, the binaural signals in both of the experiments are affected by exactly the same artifacts due to the order-limited SH processing. The signals in experiment 1 are additionally impaired by the undersampling artifacts because of the sampling with the SMA.

TABLE II. The results of the mixed $4 \times 3 \times 2 \times 2$ ANOVA with the between-subject factor room (R) and the within-subject factors BRIR component (B), position (P), and signal (S) for experiment 1.

Effect	Degrees of freedom (df)	F	p[a]	ε[b]	η_G^2[c]
R	3,32	3.38	0.030*	1.0	0.064
B	2,64	266.01	<0.001*	0.75	0.066
P	1,32	0.05	0.817	1.0	0.00
S	1,32	0.0	1.0	1.0	0.00
$R \times B$	6,64	7.5	<0.001*	0.75	0.14
$R \times P$	3,32	3.66	0.022*	1.0	0.016
$R \times S$	3,32	8.51	<0.001*	1.0	0.073
$B \times P$	2,64	2.73	0.078	0.93	0.01
$B \times S$	2,64	19.12	<0.001*	0.94	0.07
$P \times S$	1,32	2.17	0.150	1.0	0.003
$R \times B \times P$	6,64	1.35	0.254	0.93	0.014
$R \times B \times S$	6,64	3.13	0.011*	0.94	0.035
$R \times \mathrm{P} \times S$	3,32	2.23	0.104	1.0	0.01
$B \times P \times S$	2,64	1.37	0.261	0.89	0.005
$R \times B \times P \times S$	6,64	0.8	0.560	0.89	0.009

[a]p, The Greenhouse-Geisser corrected p-values with the statistical significance at the 5% level as indicated by the asterisks.
[b]ε, the Greenhouse-Geisser epsilons (note that only the factors with more than one level can be corrected for the sphericity).
[c]η_G^2, the generalized eta squared according to Olejnik and Algina (2003).

TABLE III. The results of the nested $3 \times 2 \times 2$ ANOVA with the within-subject factors BRIR component (B), position (P), and signal (S) for the data of room CR1 are shown for experiment 1.

Effect	df	F	p[a]	ε[b]	η_G^2[c]
B	2,16	90.905	<0.001*	0.744	0.772
P	1,8	0.175	0.687	1	0.0
S	1,8	0.006	0.939	1	0.0
$B \times P$	2,16	2.120	0.159	0.893	0.023
$B \times S$	2,16	4.485	0.039	0.814	0.04
$P \times S$	1,8	4.392	0.069	1	0.014
$B \times P \times C$	2,16	1.196	0.321	0.75	0.031

[a]p, The Greenhouse-Geisser corrected p-values and the statistical significance at the 5% level as indicated by the asterisks.
[b]ε, the Greenhouse-Geisser epsilons (note that only the factors with more than one level can be corrected for the sphericity).
[c]η_G^2, the generalized eta squared according to Olejnik and Algina (2003).

TABLE IV. The results of the nested $3 \times 2 \times 2$ ANOVA with the within-subject factors BRIR component (B), position (P), and signal (S) for the data of room CR7 are shown for experiment 1.

Effect	df	F	p[a]	ε[b]	η_G^2[c]
B	2,16	59.115	0.000	0.906	0.683
P	1,8	8.902	0.018	1	0.041
S	1,8	30.921	0.001	1	0.205
$B \times P$	2,16	0.632	0.541	0.975	0.012
$B \times S$	2,16	20.20	0.001	0.626	0.288
$P \times S$	1,8	2.778	0.134	1	0.021
$B \times P \times S$	2,16	0.121	0.850	0.825	0.001

[a]p, The Greenhouse-Geisser corrected p-values with statistical significance at the 5% level as indicated by asterisks.
[b]ε, the Greenhouse-Geisser epsilons (note that only the factors with more than one level can be corrected for the sphericity.
[c]η_G^2, the generalized eta squared according to Olejnik and Algina (2003).

TABLE V. The results of the nested $3 \times 2 \times 2$ ANOVA with the within-subject factors BRIR component (B), position (P), and signal (S) for the data of room SBS are shown for experiment 1.

Effect	df	F	p[a]	ε[b]	η_G^2[c]
B	2,16	39.642	0.000	0.573	0.549
P	1,8	2.200	0.176	1	0.019
S	1,8	1.408	0.269	1	0.017
$B \times P$	2,16	2.447	0.137	0.750	0.042
$B \times S$	2,16	10.083	0.003	0.832	0.062
$P \times S$	1,8	0.805	0.396	1	0.004
$B \times P \times S$	2,16	2.186	0.151	0.907	0.018

[a]p, The Greenhouse-Geisser corrected p-values with statistical significance at the 5% level as indicated by asterisks.
[b]ε, the Greenhouse-Geisser epsilons (note that only the factors with more than one level can be corrected for the sphericity).
[c]η_G^2, the generalized eta squared according to Olejnik and Algina (2003).

TABLE VI. The results of the nested $3 \times 2 \times 2$ ANOVA with the within-subject factors BRIR component (B), position (P), and signal (S) for the data of room LBS are shown for experiment 1.

Effect	df	F	p[a]	ε[b]	η_G^2[c]
B	2,16	143.959	0.000	0.661	0.66
P	1,8	1.087	0.328	1	0.007
S	1,8	6.503	0.034	1	093
$B \times P$	2,16	0.939	0.402	0.865	0.008
$B \times S$	2,16	0.287	0.691	0.741	0.01
$P \times S$	1,8	1.800	0.217	1	0.012
$B \times P \times S$	2,16	0.020	0.966	0.834	000

[a]p, The Greenhouse-Geisser corrected p-values with statistical significance at the 5% level as indicated by the asterisks.
[b]ε, the Greenhouse-Geisser epsilons (note that only the factors with more than one level can be corrected for the sphericity).
[c]η_G^2, the generalized eta squared according to Olejnik and Algina (2003).

TABLE VII. The results of the nested $4 \times 2 \times 2$ ANOVA with the between-subject factor room and within-subject factors position (P) and signal (S) for the data of the direct sound are shown for experiment 1.

Effect	df	F	p[a]	η_G^2[b]
R	3,32	0.569	0.639	0.0280
P	1,32	3.046	0.091	0.013
S	1,32	8.364	0.007	0.034
$R \times P$	3,32	2.716	0.061	0.053
$R \times S$	3,32	0.394	0.758	0.008
$P \times S$	1,32	0.285	0.597	0.001
$R \times P \times S$	3,32	0.855	0.474	0.01

[a]p, The Greenhouse-Geisser corrected p-values with statistical significance at the 5% level as indicated by the asterisks.
[b]η_G^2, the generalized eta squared according to Olejnik and Algina (2003).

APPENDIX B: ADDITIONAL INFORMATION OF THE APPLIED STATISTICS

This appendix displays additional information of the applied statistics presented in Secs. III and IV.[2]

1. Experiment 1

Olejnik and Algina (2003) proposed the generalized eta squared as a measure for the effect size in repeated measures ANOVAs. This was supported by Bakeman (2005). Therefore, we report the generalized eta squared in ANOVA Tables II and III.

a. t-test results

1. Pairwise t-tests between rooms with the pooled data of signal, position, and BRIR component

(1) CR1 vs CR7: $(t(107) = 3.068, p < 0.055, d = 0.246)$
(2) CR1 vs SBS: $(t(107) = 1.121, p < 1.000, d = 0.100)$
(3) CR1 vs LBS: $(t(107) = 0.599, p < 1.000, d = 0.049)$
(4) CR7 vs SBS: $(t(107) = 3.991, p < 0.002^*, d = 0.391)$
(5) CR7 vs LBS: $(t(107) = 3.458, p < 0.016^*, d = 0.338)$
(6) SBS vs LBS: $(t(107) = 0.701, p < 1.000, d = 0.059)$

2. t-tests between early reflections and reverberation part for SBS and LBS

(1) SBS $(t(35) = 3.466, p < 0.028^*, d = 0.79)$
(2) LBS $(t(35) = 3.651, p < 0.017^*, d = 0.762)$

3. Pairwise t-tests between signal 1 and signal 2 for each room and BRIR component separately

(1) CR1 DS − signal 1 vs signal 2: $(t(17) = 1.714, p < 1.000, d = 0.416)$
(2) CR1 ER − signal 1 vs signal 2: $(t(17) = 0.788, p < 1.000, d = 0.282)$
(3) CR1 REV − signal 1 vs signal 2: $(t(17) = -1.118, p < 1.000, d = 0.436)$
(4) CR7 DS − signal 1 vs signal 2: $(t(17) = 1.092, p < 1.000, d = 0.264)$
(5) CR7 ER − signal 1 vs signal 2: $(t(107) = 0.599, p < 1.000, d = 0.593)$
(6) CR7 Rev − signal 1 vs signal 2: $(t(17) = 10.547, p < 0.001, d = 2.451)$
(7) SBS DS − signal 1 vs signal 2: $(t(17) = 2.496, p < 0.463, d = 0.445)$
(8) SBS ER − signal 1 vs signal 2: $(t(17) = 2.268, p < 0.733, d = 0.722)$
(9) SBS Rev − signal 1 vs signal 2: $(t(17) = 1.916, p < 1.000, d = 0.671)$

TABLE VIII. The determined PSEs averaged across subjects and rooms with respect to BRIR component, source position, and test signal are shown for experiment 1; additionally, the 95% between-subject confidence intervals are presented. The room had a significant effect, which is why the interpretation of the mean values should be performed with reservation.

		DS	ER	REV
		PSE ± CI	PSE ± CI	PSE ± CI
Noise	30°	19 ± 1.3	12 ± 1.0	7 ± 1.0
	270°	20 ± 1.6	13 ± 1.16	6 ± 0.96
Speech	30°	17 ± 1.3	13 ± 1.35	9 ± 1.4
	270°	18 ± 1.44	12 ± 1.15	9 ± 1.2

https://doi.org/10.1121/10.0009277

TABLE IX. The results of the mixed $4 \times 3 \times 2 \times 2$ ANOVA with the between-subject factor room (R) and the within-subject factors BRIR component (B), position (P), and signal (S) are shown for experiment 2.

Effect	df	F	p[a]	ε[b]	η_G^2[c]
R	3,32	2.41	0.086	1.0	0.039
B	2, 64	122.29	<0.001*	0.65	0.52
P	1, 32	6.80	0.014*	1.0	0.012
S	1, 32	3.80	0.06	1.0	0.009
$R \times B$	6, 64	0.91	0.466	0.65	0.024
$R \times P$	3, 32	2.24	0.102	1.0	0.012
$R \times S$	3, 32	0.17	0.913	1.0	0.001
$B \times P$	2, 64	5.75	0.006*	0.95	0.02
$B \times S$	2, 64	13.98	<0.001*	0.80	0.04
$P \times S$	1, 32	0.04	0.854	1.0	0.00
$R \times B \times P$	6, 64	1.43	0.221	0.95	0.015
$R \times B \times S$	6, 64	0.96	0.447	0.8	0.009
$R \times P \times S$	3, 32	1.84	0.159	1.0	0.01
$B \times P \times S$	2, 64	1.07	0.34	0.85	0.004
$R \times B \times P \times S$	6, 64	0.54	0.75	0.85	0.007

[a]p, The Greenhouse-Geisser corrected p-values with statistical significance at the 5% level as indicated by the asterisks.
[b]ε, the Greenhouse-Geisser epsilons (note that only the factors with more than one level can be corrected for the sphericity).
[c]η_p^2, the generalized eta squared according to Olejnik and Algina (2003).

(10) LBS DS − signal 1 vs signal 2: $(t(17) = 1.648, p < 1.000, d = 0.650)$
(11) LBS ER − signal 1 vs signal 2: $(t(17) = 2.665, p < 0.327, d = 0.736)$
(12) LBS REV − signal 1 vs signal 2: $(t(17) = 1.479, p < 1.000, d = 0.489)$

2. Experiment 2

a. t-test results

1. t-tests between signal 1 vs signal 2 for each BRIR component separately

(1) DS: $(t(71) = 3.764, p < 0.003^*, d = 0.481)$
(2) ER: $(t(71) = 0.705, p < 1.000, d = 0.125)$
(3) REV: $(t(71) = 2.261, p < 0.242, d = 0.311)$

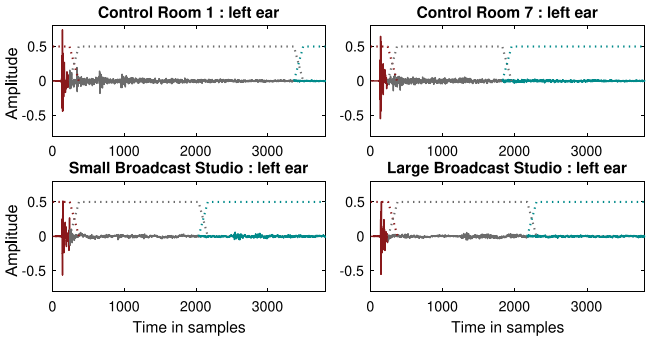

FIG. 10. (Color online) The different BRIR components and the corresponding fading windows at the transition points, as an example, for the left ear signal and each room examined are depicted.

TABLE X. The determined PSEs averaged across subjects and rooms for the conditions BRIR component, source positon, and test signal separately are shown for experiment 2; additionally, the 95% between-subject confidence intervals are presented.

		DS	ER	REV
		PSE ± CI	PSE ± CI	PSE ± CI
Noise	30°	11 ± 1.67	4 ± 0.73	4 ± 0.63
	270°	13 ± 1.66	5 ± 0.70	4 ± 0.79
Speech	30°	9 ± 1.4	5 ± 0.6	5 ± 1.0
	270°	11 ± 1.14	4 ± 0.82	5 ± 0.75

2. t-tests between position 1 vs position 2 for each BRIR component separately

(1) DS: $(t(71) = 3.240, p < 0.016^*, d = 0.403)$
(2) ER: $(t(71) = 0.587, p < 1.000, d = 0.098)$
(3) REV: $(t(71) = 0.177, p < 1.000, d = 0.029)$

3. Pairwise t-tests between all BRIR components with the pooled data of room, signal, and position

(1) DS vs ER $(t(143) = 15.986, p < 0.001^*, d = 1.796)$
(2) DS vs REV $(t(143) = 15.488, p < 0.001^*, d = 1.796)$
(3) ER vs Rev $(t(143) = 0.609, p < 1.000, d = 0.068)$

APPENDIX C: BRIR COMPONENTS AND FADING WINDOWS

Supplementary to Table I, Fig. 10 shows the left ear BRIRs for the source to the left (such that the left ear is ipsilateral) for each room and the 29th-order reference BRIR. Additionally, the linear fades are marked as dashed lines. The linear fade was performed over 128 samples so that the last 64 samples of the corresponding BRIR component were faded in and out, respectively. The employed Neumann KU100 HRIRs have a length of 128 samples. The direct sound component was defined as the first 3.5 ms (168 samples) after the onset. The mixing times were calculated for the frontal direction, which were averaged over the left and right ears according to Abel and Huang (2006) with *AKmixingTimeAbel* from the MATLAB toolbox AKtools (Brinkmann and Weinzierl, 2017).[3]

[1]See https://doi.org/10.5281/zenodo.5862771 for a detailed MATLAB code to generate Fig. 1 (Last viewed 1/22/2021).
[2]See https://doi.org/10.5281/zenodo.5862771 for a full data set of statistical results in.mat and.R format, as well as R scripts for the presented statistical analysis (Last viewed 1/22/2021).
[3]See https://doi.org/10.5281/zenodo.5862771 for a detailed MATLAB code to generate Fig. 10 (Last viewed 1/22/2021).

Abel, J. S., and Huang, P. (**2006**). "A simple, robust measure of reverberation echo density," in *Proceedings of 121th AES Convention*.
Ahrens, J. (**2019**). "Perceptual evaluation of binaural auralization of data obtained from the spatial decomposition method," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, pp. 2–6.
Ahrens, J., and Andersson, C. (**2019**). "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," J. Acoust. Soc. Am. **145**, 2783–2794.

Ahrens, J., Hohnerlein, C., and Andersson, C. (**2017**). "Authentic auralization of acoustic spaces based on spherical microphone array recordings," in *ASA/EAA Meeting*, Boston, Vol. 40, pp. 303–310.

Arend, J. M., Brinkmann, F., and Pörschmann, C. (**2021**). "Assessing spherical harmonics interpolation of time-aligned head-related transfer functions," J. Audio Eng. Soc. **69**(1/2), 104–117.

Aussal, M., Alouges, F., and Katz, B. (**2013**). "A study of spherical harmonics interpolation for HRTF exchange," in *Proceedings of Meetings on Acoustics*, Montreal, Canada, Vol. 19.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods **37**(3), 379–384.

Barron, M. (**1971**). "The subjective effects of first reflections in concert halls-The need for lateral reflections," J. Sound Vib. **15**(4), 475–494.

Bee Wah, Y., and Mohd Razali, N. (**2011**). "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," J. Stat. Model. Anal. **2**, 21–33.

Ben-Hur, Z., Alon, D. L., Mehra, R., and Rafaely, B. (**2019a**). "Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment," in *Proceedings of the IEEE/ACM Transactions on Audio Speech and Language Processing*, Vol. 27, pp. 2249–2262.

Ben-Hur, Z., Alon, D. L., Rafaely, B., and Mehra, R. (**2019b**). "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions," EURASIP J. Audio, Speech, Music Process. **2019**(1), 1–14.

Ben-Hur, Z., Sheaffer, J., and Rafaely, B. (**2018**). "Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction," Appl. Acoust. **134**, 138–144.

Bernschütz, B. (**2013**). "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 39th DAGA*, Meran, pp. 592–595.

Bernschütz, B. (**2016**). "Microphone arrays and sound field decomposition for dynamic binaural recording," Ph.D. thesis, Technische Universität Berlin, available at https://doi.org/10.14279/depositonce-5082 (Last viewed 1/22/2021).

Bernschütz, B., Giner, A. V., Pörschmann, C., and Arend, J. M. (**2014**). "Binaural reproduction of plane waves with reduced modal order," Acta Acust. Acust. **100**(5), 972–983.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (**2010**). "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio" (Design and construction of a variable spherical microphone array for research applications in room acoustics and virtual audio.), in *Proceedings of 36th DAGA*, pp. 717–718.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (**2011a**). "SOFiA Sound Field Analysis Toolbox," in *Proceedings of the International Conference on Spatial Audio (ICSA)*, pp. 8–16.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (**2011b**). "Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren" ("■"), in *Proceedings of the 37th DAGA*, 2, Düsseldorf, pp. 661–662.

Blauert, J. (**1996**). *Spatial Hearing* (Hirzel, Stuttgart), p. 459.

Bortz, J., and Schuster, C. (**2010**). *Statistik Für Human- Und Sozialwissenschaftler* Statistics For Human And Social Scientists, 7th ed. (Springer, Gießen), pp. 117–136.

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

Brinkmann, F., Lindau, A., Vrhovnik, M., and Weinzierl, S. (**2014**). "Assessing the authenticity of individual dynamic binaural synthesis," in *EAA Joint Symposium on Auralization and Ambisonics*, April, Vol. 71, pp. 3–5, available at https://depositonce.tu-berlin.de/handle/11303/168 (Last viewed 1/22/2021).

Brinkmann, F., and Weinzierl, S. (**2017**). "Aktools—An open software toolbox for signal acquisition, processing, and inspection in acoustics," in *Proceedings of the 142nd AES Convention*, AES, Berlin, Germany, pp. 1–6.

Chen, S. Y., Feng, Z., and Yi, X. (**2017**). "A general introduction to adjustment for multiple comparisons," J. Thorac. Dis. **9**(6), 1725–1729.

Engel, I., Henry, C., Garí, S. V. A., Robinson, P. W., and Picinali, L. (**2021**). "Perceptual implications of different Ambisonics-based methods for binaural reverberation," J. Acoust. Soc. Am. **149**, 895.

Engel, I., Henry, C., Garí, S. V. A., Robinson, P. W., Poirier-Quinot, D., and Picinali, L. (**2019**). "Perceptual comparison of Ambisonics-based reverberation methods in binaural listening," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*, Paris pp. 121–126.

Erbes, V., Wierstorf, H., Geier, M., and Spors, S. (**2017**). "Free database of low-frequency corrected head-related transfer functions and headphone compensation filter," in *Proceedings of the 142nd AES Convention*, pp. 1–5.

Erdfelder, E., Faul, F., Buchner, A., and Lang, A. G. (**2009**). "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses," Behav. Res. Methods **41**(4), 1149–1160.

Garí, S. V. A., Brimijoin, W. O., Hassager, H. G., and Robinson, P. W. (**2019**). "Flexible binaural resynthesis of roomi responses for augmented reality research," in *Spatial AudioSignal Processing Symposium*, pp. 161–166, available at https://hal.archives-ouvertes.fr/hal-02275193/document (Last viewed 1/22/2021).

Geier, M., Ahrens, J., and Spors, S. (**2008**). "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *Proceedings of the 124th AES Convention*, Amsterdam, NE, pp. 179–184.

Geier, M., Ahrens, J., and Spors, S. (**2019**). "The SoundScape Renderer," available at http://spatialaudio.net/ssr/ (Last viewed 1/22/2021).

Giner, A. V. (**2013**). "Scale—A software tool for listening experiments," in *Proceedings of the 39th DAGA*, pp. 1–4.

Girden, E. R. (**1992**). *ANOVA: Repeated Measures* (Sage, Newbury Park, CA).

Jarmasz, J., and Hollands, J. G. (**2009**). "Confidence intervals in repeated-measures designs: The number of observations principle," Can. J. Exp. Psychol. **63**(2), 124–138.

Jeong, C. H. (**2016**). "Diffuse sound field: Challenges and misconceptions," in *Proceedings of the 45th International Congress and Exposition on Noise Control Engineering: Towards a Quieter Future*, Vol. 4, 1015–1021.

Jot, J.-M., Larcher, V., and Pernaux, J.-M. (**1999**). "A comparative study of 3-D audio encoding and rendering techniques," in *AES 16th International Conference*, pp. 281–300.

Kingdom, F. A., and Prins, N. (**2010**). *Psychophysics: A Practical Introduction*, 1st ed. (Academic, London, UK).

Kuttruff, H. (**1973**). *Room Acoustics*, 4th ed. (Spon, London, UK).

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2B), 467–477.

Lindau, A. (**2014**). "Binaural resynthesis of acoustical environments—Technology and perceptual evaluation," Ph.D. thesis, pp. 1–279, available at https://depositonce.tu-berlin.de/handle/11303/4382 (Last viewed 1/22/2021).

Lindau, A., Kasanke, L., and Weinzierl, S. (**2010**). "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," in *Proceedings of the 128th AES Convention.*

Lindau, A., Maempel, H., and Weinzierl, S. (**2008**). "Minimum BRIR grid resolution for dynamic binaural synthesis," in *Proceedings of Acoustics 2008*, Paris, Vol. 123, pp. 3498–3498.

Loftus, G. R. (**1994**). "Using confidence intervals in within-subject designs," Psychon. Bull. Rev. **1**(4), 476–490.

Lübeck, T., Helmholz, H., Arend, J. M., Pörschmann, C., and Ahrens, J. (**2020a**). "Perceptual evaluation of mitigation approaches of impairments due to spatial undersampling in binaural rendering of spherical microphone array data: Dry acoustic environments," in *Proceedings of the International Conference on Digital Audio Effects 2020*, Vienna, Vol. 68, pp. 428–440.

Lübeck, T., Pörschmann, C., and Arend, J. M. (**2020b**). "Perception of direct sound, early reflections, and reverberation in auralizations of sparsely measured binaural room impulse responses," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality*, Redmond, WA, pp. 1–10.

McCormack, L., Pulkki, V., and Marschall, M. (**2020**). "Higher-order spatial impulse response rendering: Investigating the perceived effects of spherical order, dedicated diffuse rendering," J. Audio Eng. Soc. **68**(5), 338–354.

Meese, T. (**1995**). "Using the standard staircase to measure the point ofsubjective equality: A guide based on computer simulations," Percept. Psychophys. **25**(1), 16–18.

Merimaa, J., and Pulkki, V. (**2004**). "Spatial impulse response rendering," in *Proceedings of the 7th International Conference on Digital Audio Effects*, Naples, pp. 139–144.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**(5), 300–321.
Olejnik, S., and Algina, J. (**2003**). "Generalized eta and omega squared statistics: Measures of effect size for some common research designs," Psychol. Methods **8**(4), 434–447.
Olive, S. E., and Toole, F. E. (**1988**). "The detection of reflections in typical rooms," in *AES 85th Convention*, Ottawa, Canada, Vol. 2719.
Pearson, E. S. (**1931**). "The analysis of variance in cases of non-normal variation," Biometrika **23**(1/2), pp. 114–133.
Pike, C. W. (**2019**). "Evaluating the perceived quality of binaural technology," Ph.D. thesis, University of York.
Pörschmann, C., Arend, J. M., Bau, D., and Lübeck, T. (**2020**). "Comparison of spherical harmonics and nearest-neighbor based interpolation of head-related transfer functions," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality*, Redmond, WA, pp. 1–10.
Pörschmann, C., Arend, J. M., and Brinkmann, F. (**2019**). "Directional equalization of sparse head-related transfer function sets for spatial upsampling," IEEE/ACM Trans. Audio, Speech, Lang. Process. **27**(6), 1060–1071.
Pulkki, V. (**2007**). "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc. **55**(6), 503–516.
Rafaely, B. (**2015**). *Fundamentals of Spherical Array Processing* (Springer, Berlin).
Savioja, L., and Svensson, U. P. (**2015**). "Overview of geometrical room acoustic modeling techniques," J. Acoust. Soc. Am. **138**(2), 708–730.
Stade, P., Bernschütz, B., and Rühl, M. (**2012**). "A spatial audio impulse response compilation captured at the WDR Broadcast Studios," in *Proceedings of the 27th Tonmeistertagung—VDT International Convention*, pp. 551–567.
Tervo, S., Pätynen, J., Kuusinen, A., and Lokki, T. (**2013**). "Spatial decomposition method for room impulse responses," J. Audio Eng. Soc. **61**(1/2), 17–28.
Vorländer, M. (**2008**). *Auralization*, 1st ed. (Springer, Berlin), pp. 1–335.
Williams, E. G. (**1999**). *Fourier Acoustics* (Academic, London), p. 302.
Yap, B. W., and Sim, C. H. (**2011**). "Comparisons of various types of normality tests," J. Stat. Comput. Simul. **81**(12), 2141–2155.
Zahorik, P. (**2002**). "Direct-to-reverberant energy ratio sensitivity," J. Acoust. Soc. Am. **112**(5), 2110–2117.
Zaunschirm, M., Schörkhuber, C., and Höldrich, R. (**2018**). "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," J. Acoust. Soc. Am. **143**(6), 3616–3627.
Zotter, F. (**2009**). "Analysis and synthesis of sound-radiation with spherical arrays," Ph.D. thesis, University of Music and Performing Arts, Austria.