OPEN ACCESS Freely available online

T. Lübeck, C. Scheer, D. Ackermann, F. Brinkmann,
C. Pörschmann, S. Weinzierl, J. Ahrens, and J. M. Arend,
"Perceptual Evaluation of Different Methods for
Binaural Rendering of Recordings With Various Microphone Arrays,"
J. Audio Eng. Soc., vol. 74, no. 5, pp. 256–269 (2026 May).
https://doi.org/10.17743/jaes.2022.0259.

Perceptual Evaluation of Different Methods for Binaural Rendering of Recordings With Various Microphone Arrays

TIM LÜBECK,[1, 2, *] (tim.luebeck@icloud.com) **CHRISTIAN SCHEER,**[2] (christian.scheer@tu-berlin.de) **DAVID ACKERMANN,**[2, 3] (david.ackermann@hs-duesseldorf.de)

FABIAN BRINKMANN,[2] (fabian.brinkmann@tu-berlin.de) **CHRISTOPH PÖRSCHMANN,**[1] (christoph.poerschmann@th-koeln.de)

STEFAN WEINZIERL,[2] (stefan.weinzierl@tu-berlin.de) **JENS AHRENS,**[4] (jens.ahrens@chalmers.se) **AND JOHANNES M. AREND**[5] (johannes.arend@aalto.fi)

[1]*Institute of Computer and Communication Technology, TH Köln - University of Applied Sciences, Cologne, Germany*
[2]*Audio Communication Group, Technische Universität Berlin, Berlin, Germany*
[3]*Institute of Sound and Vibration Engineering, HSD - University of Applied Sciences, Düsseldorf, Germany*
[4]*Audio Technology Group, Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden*
[5]*Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Espoo, Finland*

Binaural reproduction of microphone array recordings has become an important technology in the research and consumer sectors. Several commercially available spherical microphone arrays have been introduced over the years along with various methods for binaural rendering of array recordings. Most of these methods have been evaluated individually, typically using only one specific microphone array. However, a comprehensive and systematic perceptual evaluation combining different methods and various microphone arrays is lacking. This study presents the results of a listening experiment comparing the motion-tracked binaural method, various Ambisonic binaural decoders, and the parametric binaural rendering method COMPASS using loudspeaker orchestra recordings with six different microphone arrays from two rooms, the Berliner Philharmonie and a laboratory space resembling a small chamber music venue. The experiment assessed the binaural renderings with respect to overall listening experience and four perceptual attributes from the Spatial Audio Quality Inventory in comparison to a reference recorded with a head and torso simulator. The results provide detailed insights into which rendering method and array combination provides a high overall listening experience while preserving the assessed perceptual attributes externalization, coloration, source position, and presence. Moreover, the results indicate the extent to which the assessed perceptual attributes contribute to overall listening experience.

Keywords: ambisonics, magnitude-least squares, parametric audio, stereo

1 INTRODUCTION

The binaural reproduction of spherical microphone array (SMA) recordings has been intensively researched in recent years and has long since found its way into commercial audio production. Although binaural recordings with head and torso simulators are still the ground truth for binaural reproduction, binaural synthesis from spherical array recordings offers a lot of flexibility. This includes dynamic binaural synthesis that adapts to the listener's head orientation in real-time, or the incorporation of individual head-related transfer functions (HRTFs). These features make microphone arrays flexible tools for applications such as live broadcasting of concerts or teleconferencing.

Over the years, different methods for binaural rendering of array recordings have been introduced. A simple approach is the motion-tracked binaural (MTB) rendering [1]. This method is based on microphone arrays consisting of a rigid spherical body equipped with microphones along the equator, hereafter referred to as equatorial microphone arrays (EMAs). So-called pseudobinaural signals are extracted from the recordings of two opposing micro-

*To whom correspondence should be addressed, email: tim.luebeck@icloud.com.

phones, resulting in interaural time and level differences of a spherical head model. Dynamic binaural synthesis is achieved by next-neighbor interpolation of the two microphone signals closest to the actual position of the listener's ears. Although this method does not involve HRTF processing and thus cannot produce the cues caused by the human pinnae, it can provide good externalization and plausible rendering [2].

More complex methods apply linear Ambisonic processing. SMA recordings can be encoded into Ambisonic signals by means of spherical harmonic (SH) decomposition and radial filtering [3, 4]. Ambisonics is a standardized scene-based format that provides a variety of sound field manipulations, such as rotation. The sound field encoded in the Ambisonic domain, also referred to as the SH domain, can then be decoded either to headphones for dynamic binaural synthesis or to different loudspeaker arrangements. The encoding of SMA signals into the SH domain is significantly limited by the number of microphones. On the one hand, the limited number of microphones determines the maximum order N of the SH decomposition and thus the spatial resolution that can be encoded ($N \leq \sqrt{Q} - 1$, with Q being the number of microphones). On the other hand, the associated spatial undersampling introduces spatial aliasing artifacts. Both limitations lead to audible impairments in the high-frequency components [5–8].

In recent years, several approaches have been developed to mitigate these artifacts, such as spectral equalization [9, 10]; the magnitude least squares (MagLS) optimization [11, 4]; side-lobe suppression algorithms, such as max-$\boldsymbol{r}_\mathrm{E}$ [12]; or combinations of these methods [13]. The MagLS decoder has become an established method and can be considered state of the art for binaural decoding of Ambisonic signals. Several open-source applications are available for Ambisonic encoding and decoding, such as the IEM Plug-in Suite [4, 14], SPARTA [15], or ReTiSAR [16, 17].

Recordings from EMAs cannot only be rendered using the MTB approach, but they can also be used to compute SH signals, as recently demonstrated by Ahrens et al. [18, 19]. This method has two advantages over encoding from SMA recordings: 1) significantly fewer microphones are needed to decompose to the same SH order N, and 2) the microphone arrangement along the equator is more suitable for consumer devices than the uniform distribution of sensors on the spherical body. In this sense, EMAs can be considered a precursor to head-worn microphone arrays. This comes at the cost that the SH decomposition of the EMA data assumes a height-invariant sound field. However, because the human auditory system is less sensitive in the median plane [20], this limitation may be acceptable from a perceptual point of view.

The above mentioned methods can be classified as scene-based and linear methods. Another class of array-based rendering methods is parametric spatial audio. The sound field is first encoded into a set of perceptually important parameters, such as the direction of arrival (DOA) of sound sources and the direct-to-diffuseness balance, for each time-frequency index. Then, the sound field is decoded for either playback over loudspeakers or binaurally to headphones, with most of the available methods typically relying on similar sound field model assumptions and rendering techniques.

Methods, such as the spatial impulse response rendering (SIRR) [21], higher-order (HO) SIRR [22], and the spatial decomposition method (SDM) [23] operate on impulse responses and are not suitable for real-time applications. However, other methods, such as Directional Audio Coding (DirAC) [24], HO-DirAC [25], and COMPASS [15], are not limited to impulse responses and are alternatives for real-time rendering. The main goal of parametric approaches is to improve the perceptual quality of the rendering. This is typically accomplished by spatially sharpening the directional components of a few individual sound sources using beamforming and HRTF convolution while decorrelating the remainder of the sound field, which is assumed to contribute to the diffuse reverberation. Conventional linear Ambisonic processing cannot achieve this. However, parametric approaches carry the risk of the input sound scene violating the underlying sound field assumptions, which can sometimes give rise to audible artifacts. Therefore, contrary to traditional Ambisonic processing, their perceptual performance can be scene dependent.

Several studies have perceptually evaluated the aforementioned rendering techniques in different ways. The performance of the MTB rendering based on two different EMAs was evaluated by Ackermann et al. [2]. Bernschütz [26] presented results of a large number of listening experiments evaluating the perceptual quality of binaural rendering of SMA recordings compared with binaural recording with an artificial head.

The main foci were to evaluate the influence of the SH rendering order, the number of microphones, or the sampling grid of the SMA. This work was continued by Ahrens and Andersson [27]. A comparison of different binaural decoding methods has been done by Zaunschirm et al. [28] or Lübeck et al. [29, 30]. Recently, Helmholz et al. [31] presented results of a perceptual evaluation comparing binaural renderings of recordings from SMAs, EMAs, and head-worn microphone arrays. COMPASS was evaluated by Politis et al. [32], and a binaural version of SDM was evaluated by Amengual et al. [33]. McCormack et al. [34] presented a comparison of the different parametric methods SDM, SIRR, and REPAIR. Furthermore, Pawlak et al. [35] conducted listening experiments comparing the two parametric approaches SDM and HO-SIRR for different array configurations. However, only a few studies have systematically compared linear Ambisonics with parametric rendering, such as that of McCormack et al. [36, 37].

In recent years, a wide variety of rendering methods have been developed, and numerous SMAs with different radii and channel counts have become commercially available. Consequently, a wide range of methods and combinations for recording and reproduction are now available for professional audio production. However, a systematic perceptual evaluation that covers different array configurations in combination with linear and parametric rendering under real-world conditions is lacking. Here, real-world conditions denote the use of physical (nonsimulated) arrays and the real-

time rendering of continuous program material (e.g., live concert broadcasts), rather than impulse response–based workflows. Unlike impulse response–based rendering, program material and its processing are subject to practical limitations, particularly a limited signal-to-noise ratio. This can result in perceivable artifacts in recordings with real-world microphone arrays, as demonstrated by Helmholz et al. [38]. In light of this, it is particularly important for practicing audio engineers to understand which microphone array and rendering approach yields the optimal perceptual results in live broadcasting and recording scenarios. To date, a systematic, comparative assessment providing such guidance has been missing.

To address this gap, this work presents a comprehensive comparison of different binaural rendering methods based on recordings with different microphone arrays. The focus lays on compact microphone arrays featuring microphones arranged on a spherical surface, including rigid, open, and equatorial configurations. The study uses the database provided by Ackermann et al. [39], which contains recordings of different microphone arrays of an orchestra assembled with 18 loudspeakers in two different rooms. The database includes recordings with the FABIAN head and torso simulator [40], an 8- and 16-channel EMA built at Technische Universität (TU) Berlin, the first-order Ambisonic microphone Sennheiser AMBEO [41], the third-order SMA ZYLIA ZM-1 [42], the fourth-order SMA mh acoustics' Eigenmike 32 [43, 44], and the 7th-order SMA HØSMA-7N MKII [45] constructed at Technische Hochschule (TH) Köln. The composition of the orchestra with loudspeakers instead of real instruments and musicians allowed a reproducible recording with the different microphone arrays under the same conditions, which is why the database is perfectly suited for a broad and systematic comparison.

Using the orchestra recordings, a two-part listening experiment was conducted. The first part of the experiment employed a test design evaluating the overall listening experience (OLE). OLE is a term derived from the quality of experience developed in the field of telecommunications, where it has been defined as: "The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state" [46].

Following this example, Schöffler et al. [47] used OLE as a construct to describe the degree of enjoyment while listening to music, including all factors that influence enjoyment. Although traditional paradigms, such as the Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA) [48], measure audio quality by comparing the perceptual difference to an explicit reference, OLE is measured by asking participants about the overall enjoyment of listening without any reference. Schöffler et al. [49] consider OLE to be well suited for assessing whether the performance of an audio system is important to potential customers.

In contrast to the reference-free OLE assessment, the perceptual evaluation based on the Spatial Audio Quality Inventory (SAQI) developed by Lindau et al. [50] focuses on specific audio quality metrics. The SAQI catalog contains a variety of attributes to describe the quality of a spatial audio reproduction. To gain deeper insight into the performance of each binaural reproduction method across different perceptual attributes, the second phase of the experiment evaluated audio quality based on the four SAQI attributes externalization, source position, tone color, and presence. Beyond determining the quality achievable with each rendering method, the combined analysis of these metrics also provides valuable insight into how each attribute affects the listener's OLE.

The paper is structured as follows. SEC. 2 presents the methodology in full, including the participants, setup, recordings used, the different binaural rendering methods investigated, the experimental procedure, and the statistical analysis. SEC. 3 presents the results for OLE and the assessed SAQI attributes and predicts how the SAQI attributes affected the OLE ratings. SEC. 4 discusses the findings, limitations, and implications, and SEC. 5 concludes with key takeaways and directions for future work.

2 METHODS

2.1 Participants

Fifty participants with an average age of 28.3 years participated in the listening experiment. Twenty-six of them performed the experiment at TH Köln and 24 at TU Berlin. Seventeen subjects had already participated in more than two listening experiments in the context of binaural technology. All participants had self-reported normal hearing.

2.2 Setup

Dynamic binaural synthesis was applied using the SoundScape Renderer (v.0.3.4) [51, 52] in binaural playback mode, which fades between precomputed binaural signals according to the listener's instantaneous horizontal head orientation. The listener's head orientation was tracked with a Polhemus PATRIOT tracker at a sampling rate of 120 Hz. The binaural signals were calculated for head orientations $\pm$ 45° from the frontal direction in 1° resolution. Hence, the dynamic binaural synthesis only accounted for horizontal head movements.

The experiments were conducted at TH Köln and TU Berlin. At TH Köln, the experiments were conducted in the anechoic chamber with a background noise level of less than 20 dB(A) using the head tracker and an RME Fireface UFX II at 48 kHz and a buffer size of 256 samples as audio interface and headphone digital-to-analog converter. At TU Berlin, the experiments were conducted in a soundproof and acoustically treated audiometric booth [Desone A:BOX System ZS, size G, background noise level of less than 20 dB(A)] using the same head tracker and an RME Fireface UFX at 48 kHz and a buffer size of 256 samples. In both laboratories, the identical Sennheiser HD 600 headphones were used for playback.

2.3 Recordings

To obtain reproducible stimuli, recordings of a loudspeaker orchestra were used [39]. A 2:08 minute excerpt

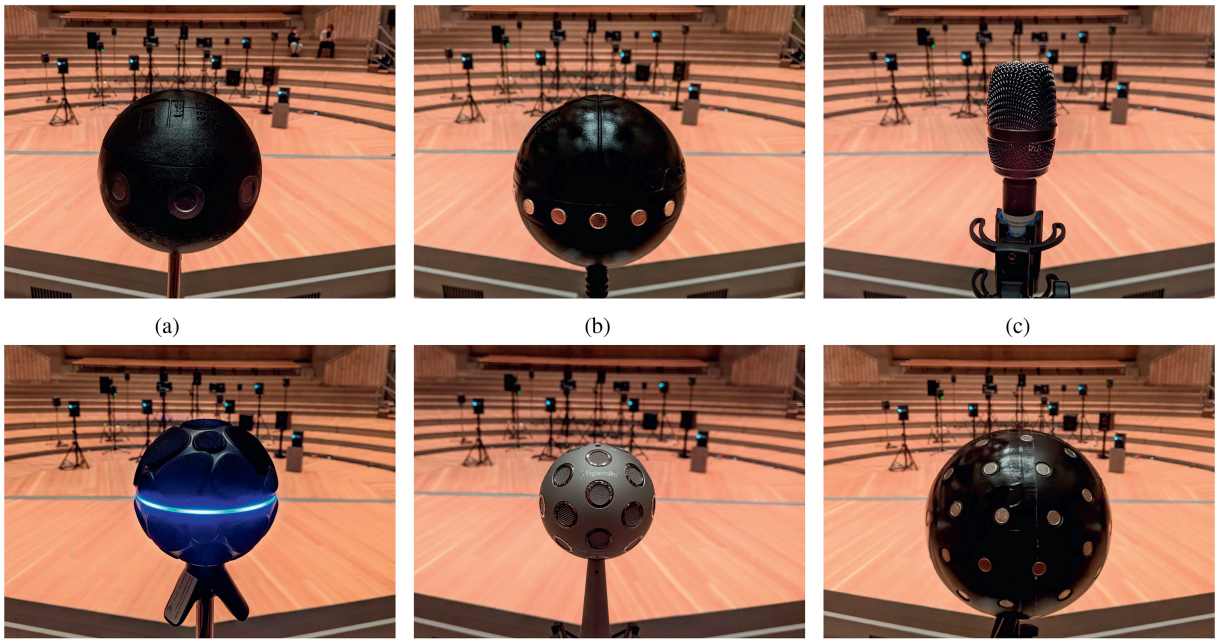


Fig. 1. The six microphone arrays in the main hall of the BPH with the 18-loudspeaker orchestra on stage in the background. The pictures are taken from Ackermann et al. [39]. (a) EMA 8. (b) EMA 16. (c) Sennheiser AMBEO. (d) Zylia ZM-1. (e) mh acoustics em32 Eigenmike. (f) HØSMA-7N MKII.

of "Golliwogg's Cakewalk" from Claude Debussy's suite *Children's Corner* was recorded with the Konzerthausorchester Berlin in the anechoic chamber of TU Berlin. Each musician was recorded separately to avoid cross talk during the recording. To simulate an orchestra on stage, an assemble of 18 loudspeakers was placed according to the "American" seating on the stage in the Berliner Philharmonie (BPH; reverberation time $T_{20,\ \mathrm{mean}} = 1.95$ s) and the Berlin Open Lab (BOL; $T_{20,\ \mathrm{mean}} = 1.15$ s). This setup allowed sequential recording of the loudspeaker orchestra with the FABIAN head and torso simulator and each of the following microphone arrays: eight-channel EMA (EMA8), 16-channel EMA (EMA16), Sennheiser Ambeo, Zylia ZM-1, mh acoustics em32 Eigemike, and HØSMA-7N MKII. A detailed description of the loudspeaker orchestra, the actual recordings, and the data postprocessing is provided by Ackermann et al. [39]. Fig. 1 shows all microphone arrays employed in this study.

2.4 Binaural Rendering Methods

The present study aims to represent the current state-of-the-art and publicly available rendering methods. Therefore, the synthesis of the binaural signals is mainly based on the software provided by the array manufacturers or on publicly available research code. Table 1 summarizes the microphone arrays and rendering methods used in the present study, as described below. In the remainder of this article, *rendering method* refers to a specific combination of microphone array and rendering technique, as indicated in Table 1 by the *denotation*. The project website[1] presents audio examples of all binaural renderings for the frontal head orientation.

Motion tracked binaural sound: The recordings from EMA8 and EMA16 were rendered to pseudobinaural signals according to the MTB method [1]. Because the sound field is sampled at discrete points, signals from two adjacent capsule pairs must be interpolated using weights that are determined by the current position of the head. This allows pseudobinaural signals to be reproduced continuously in space when the head is turned.

Algazi et al. [1] proposed five different interpolation methods. The two-band spectral interpolation restoration algorithm produced the best results when evaluated with different numbers of microphone capsules and different types of audio content (see, for example, [53]). Furthermore, it has been shown that the recorded sound field can be plausibly reproduced via headphones using this particular interpolation approach [2]. In this method, which was also employed in the present study, the low-frequency component is linearly interpolated in the time domain, whereas the high frequencies are interpolated in the frequency domain using short-time Fourier transforms (128 samples, 75% overlap, and a Hanning window). Although the high-frequency magnitude response is obtained by linear interpolation, the phase response is taken from the nearest neighbor to avoid comb filter coloration.

EMA Ambisonic encoding: The recordings from the EMA8 and EMA16 were additionally encoded into SH signal using the method of Ahrens et al. [18]. The computation of the SH signals was performed in MATLAB with the implementation that was presented in [54]. The processing assumes that the impinging sound field is height invariant. A circular harmonic decomposition of the sound field is

[1] https://audiogroupcologne.github.io/binaural_examples_jaes2026/

Table 1. Combinations of microphone arrays and rendering methods assessed in the perceptual evaluation.

Microphone Array	Encoding	Decoding	Denotation
EMA8	···	MTB	MTB8
EMA16	···	MTB	MTB16
EMA 8	EMA $N = 4$	MagLS $N = 4$	EMA N4
EMA16	EMA $N = 7$	MagLS $N = 7$	EMA N7
Sennheiser Ambeo	Ambisonics $N = 1$	MagLS $N = 1$	Ambeo
Zylia ZM-1	Ambisonics $N = 3$	MagLS $N = 3$	Zylia
mh acoustics em32 Eigenmike	Ambisonics $N = 4$	MagLS $N = 4$	em32
HØSMA-7N MKII	Ambisonics $N = 7$	MagLS $N = 7$	HØSMA
EMA8	EMA $N = 4$	COMPASS	EMA N4 C
EMA16	EMA $N = 7$	COMPASS	EMA N7 C
Sennheiser Ambeo	Ambisonics $N = 1$	COMPASS	Ambeo C
Zylia ZM-1	Ambisonics $N = 3$	COMPASS	Zylia C
mh acoustics em32 Eigenmike	Ambisonics $N = 4$	COMPASS	em32 C
HØSMA-7N MKII	Ambisonics $N = 7$	COMPASS	HØSMA C
Sennheiser Ambeo	Ambisonics $N = 1$	MS Stereo	Stereo

Note. All Ambisonic data were processed using ACN channel ordering and N3D normalization.

computed from which the effect of the microphone baffle is removed before being converted to SH signals. Finally, a global equalization is applied to account for potential effects of SH truncation and spatial aliasing on the spectral balance of the binaural signals. When rendered binaurally, EMAs preserve binaural elevation cues for sounds that impinge from nonhorizontal directions [19].

SMA Ambisonic encoding: The SMAs were encoded into SH signals using virtual studio technology (VST) plug-ins in Reaper. For the commercially available SMAs Sennheiser AMBEO, Zylia ZM-1, and mh acoustics em32 Eigenmike, the manufacturers' software was used. The Ambeo microphone was encoded using the VST plug-in `AMBEO A-B format converter` with the Ambisonic correction filter activated, low-cut filter deactivated, microphone rotation set to 0°, position set to upright, and output format set to ambiX. The Zylia microphone was encoded using the VST plug-in `ZYLIA Ambisonics Converter` with order set to 3rd, channel ordering set to Ambisonic channel number (ACN), normalization set to fully normalized (N3D), microphone position set to upright, and rotation and elevation set to 0°. The em32 was encoded using the VST plug-in `EigenUnit-em32-Encoder` with Ambisonic order set to 4th, channel ordering set to ACN, and normalization set to N3D. The HØSMA array was encoded using the VST plug-in `array2sh` from the SPARTA plug-in suite [15] with the diffuse-field equalization enabled (see Eq. (3) in [37]) and regularized radial filters limited to 20 dB using the soft-limiting approach [26].

Ambisonic binaural decoding: Binaural decoding of the SH signals from the EMAs and SMAs was performed in MATLAB using the toolbox provided by Deppisch et al. [13]. It contains an implementation of the MagLS algorithm [11, 4], which the authors used for all renderings. As recommended by the MagLS developers [4], the authors applied a diffuse covariance constraint for order $N = 1$, which in the present case concerned the Ambeo SMA. As the HRTF set for binaural decoding, the FABIAN HRTFs were used [40], which were measured with the same artificial head that was also used for the binaural reference recordings in [39].

Parametric binaural decoding: As a state-of-the-art representative parametric rendering approach, the COMPASS algorithm was chosen [32, 15]. The COMPASS renderings are based on the identical SH signals and the same HRTF set as the binaural Ambisonic decodings and were provided by the developers of COMPASS using the COMPASS plug-in [55].

MS Stereo: To also include a traditional and static stereophonic reproduction in the perceptual evaluation, the authors additionally synthesized stereo signals from the W and Y channels of the Ambeo microphone using the VST plug-in `Voxengo MSED` in decode mode, with mid gain set to 0 dB, side gain to −3.0 dB, and both mid and side pan set to center. This resulted in an MS-to-stereo conversion because the microphone setup ensured that the Y channel of the B-format corresponded to the side signal and the W signal to the mid signal in MS terminology. The applied 3-dB attenuation of the side signal resulted in a stereo recording angle of roughly 60° [56; p. 7, Fig. 9], matching the geometric arrangement of the loudspeaker orchestra and microphone.

The reference recordings with the FABIAN head and torso simulator were diffuse-field equalized. Accordingly, all renderings except the MTB renderings were processed with the same diffuse-field compensation filter.

2.5 Procedure

The listening experiment consistent of two phases that were performed directly one after the other by the participants. This took approximately 60 minutes per participant. In the first phase, listeners rated the OLE by answering the question, "How much did you enjoy listening to the following pieces of music?" Responses were recorded on a

five-point Likert scale with the labels "very much," "quite," "neutral," "not a lot," and "not at all," which has been shown to yield consistent OLE ratings [57, 58]. In the OLE phase, no reference was provided, and the FABIAN artificial head recordings were part of the stimuli that participants rated. The 16 conditions were randomly divided into two rating pages, with each page consisting of eight sliders with eight play buttons, each assigned to one rendering. The stimuli presented in a 15 s loop, and between the renderings, it could be toggled as often as desired by pressing the play button on the graphical user interface (GUI).

In the second phase, participants rated the SAQI attributes "externalization," "source position," "tone color," and "presence" (where "presence" is not an acoustic attribute but refers to the perception of being in the scene; cf. [50]). At the beginning of each experiment, the experimenter introduced the SAQI attributes and clearly explained their meanings to the participants. Furthermore, a handout was provided for reference during the test in case any clarification was needed. Again, the conditions were randomly divided into two rating pages with eight and seven sliders, respectively, and an "A" and "B" button was assigned to each slider on the GUI. Pressing the "A" button would always play the FABIAN reference, and pressing the "B" button would play the stimulus to be rated. Accordingly, the GUI asked the participants to rate "B" as compared with "A" with respect to the corresponding SAQI attribute. Participants were aware that "A" was the reference. Again, the stimuli were presented in a 15 s loop, and it was possible to toggle between the renderings as often as desired. When toggling, the stimuli played continuously without interruption.

2.6 Statistical Analysis

The OLE and SAQI results were separately analyzed using linear mixed-effects models (LMM) with restricted maximum likelihood estimation of variance components and type III analysis of variance via Satterthwaite's degrees of freedom method. The assumption of normal distribution of the residuals was checked by visual inspection of the quantile-quantile plots and Kolmogorov-Smirnov tests. The LMMs with the fixed-effects room, rendering method, and laboratory (see Table 4 in the supplementary materials [59]), and the random intercept subject for each SAQI attribute and for OLE showed no significant effect for the lab. For this reason, the data were pooled across laboratory, so the reported results are based on LMMs that do not include the fixed effect of laboratory.

For further analysis of the OLE results, Holm-corrected [60] post hoc comparison between each rendering method and FABIAN were performed. The SAQI results were analyzed in more detail with Holm-corrected one sample t tests against 0 for each rendering method and room separately.

To examine the influence of the four SAQI attributes on OLE, an additional LMM was fitted with the dependent variable OLE, the fixed effects room, rendering method, externalization, tone color, source position, presence, and the random effect subject. The four SAQI attributes acting as predictors were standardized (z-scores) before analysis. The absolute bivariate Pearson correlations between the SAQI attributes were all below 0.35, falling short of the established threshold of 0.5 to 0.7, above which multicollinearity can severely distort model estimation [61].

Table 2. Fixed-effects results from the LMM on OLE, including F values, degrees of freedom (df), and p values.

Effect	F	df	p
Rendering method	36.19	15	< 0.001
Room	9.43	1	0.002
Rendering method × room	1.75	15	0.036

The models were calculated with Jamovi using the GAMLj module (version 3) [62]. More detailed statistical results, which are not reported below, are provided in the supplementary materials [59].

3 RESULTS

3.1 Overall Listening Experience

The model explains $R^2_{\text{conditional}} = 29.4\%$ of the variance of fixed and random effects, and $R^2_{\text{marginal}} = 25.5\%$ of the variance of only the fixed effects. The analysis revealed significant fixed effects of rendering method, room, and the interaction of both (see Table 2). The estimated marginal means with 95% confidence intervals (CIs) of the interaction rendering method × room are shown in Fig. 2(a). The MTB16 achieved the highest mean OLE ratings followed by MTB8, em32 C, Zylia C, and FABIAN. Notably, the lowest mean OLE was obtained for the HØSMA C. For the Zylia, a notable difference between BOL and BPH can be observed, which could evoke the significant interaction room × rendering method.

A Holm-corrected post hoc comparison between FABIAN and each rendering method revealed that EMA N4 ($p < 0.001$), Ambeo ($p = 0.023$), Zylia ($p = 0.001$), HØSMA ($p = 0.011$), EMA N4 C ($p < 0.001$), EMA N7 C ($p = 0.004$), Ambeo C ($p < 0.001$), HØSMA C ($p < 0.001$), and Stereo ($p < 0.001$) led to significantly lower OLE ratings than FABIAN. The corresponding estimated marginal means of rendering method, including 95% CIs, with the data pooled over both rooms, are shown in Fig. 2(b). Asterisks indicate a significant difference between the respective rendering method and the FABIAN reference. Thus, the results suggest that MTB8, MTB16, EMA N7, em32, and the COMPASS renderings of Zylia and em32 are the only rendering methods that provide an OLE similar (i.e., not significantly different) to that of FABIAN.

3.2 Spatial Audio Quality Inventory

The model for externalization explains $R^2_{\text{conditional}} = 16\%$ of the variance of the fixed and random effects and $R^2_{\text{marginal}} = 9.6\%$ of the variance of only the fixed effects. The model yielded a significant effect only for the fixed effect rendering method (see Table 3). Corresponding estimated marginal means of the interaction between room and rendering method are shown in Fig. 2(c). It can be

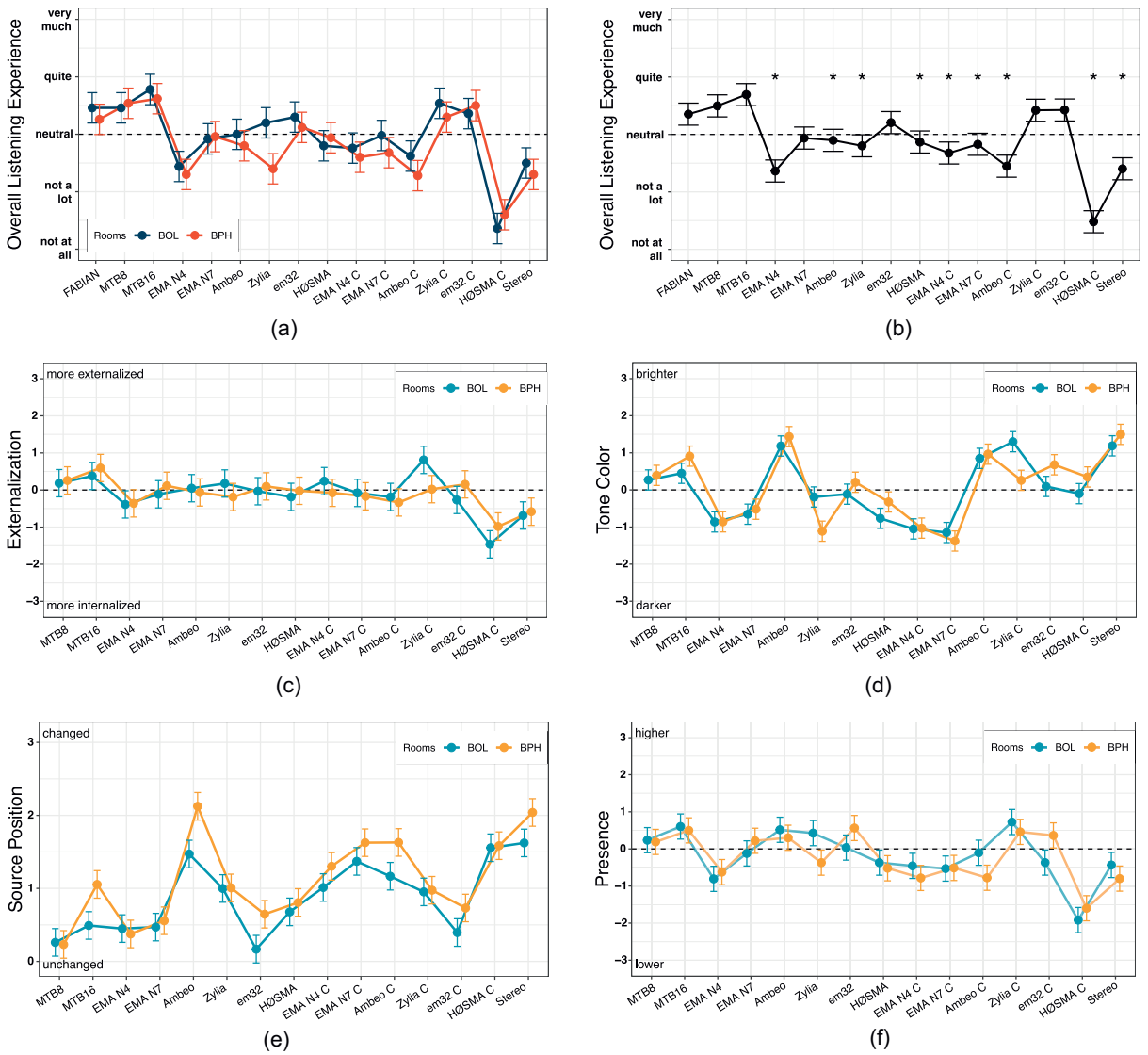


Fig. 2. Estimated marginal means and 95% CIs for the OLE ratings with respect to (a) rendering method and room and (b) pooled across rooms and [(c) through (f)] for the SAQI ratings with respect to rendering method and room. Asterisks in (b) indicate significant differences from FABIAN.

observed that most rendering methods performed similarly and close to zero, suggesting that for most rendering methods, participants could not perceive strong differences in externalization compared to FABIAN.

HØSMA C and Stereo have the lowest mean externalization ratings, indicating that these rendering methods lead to more internalized sound than the FABIAN reference, whereas Zylia C for the room BOL and MTB16 for both rooms led to the highest mean ratings, suggesting slightly improved externalization by these rendering methods compared to FABIAN. The Holm-corrected one-sample t tests statistically supported most of those observations, yielding significant differences from zero toward more internalized for HØSMA C in both rooms and toward more externalized for Zylia C renderings in the room BOL (see Table 27 in the supplementary materials [59]).

For tone color, the model resulted in $R^2_{\text{conditional}} = 44.3\%$ and $R^2_{\text{marginal}} = 43.1\%$ and yielded a significant effect for the fixed effect rendering method and the interaction rendering method × room (see Table 3). The marginal means of the interaction in Fig. 2(d) show clear deviations from zero for various rendering methods, suggesting clear audible differences in terms of coloration when compared to the reference. In general, Ambeo, Zylia C in the room BOL, and Stereo have the highest mean values, indicating that these renderings were substantially brighter in tone color than the FABIAN reference, whereas EMA N7 and EMA N4, with and without COMPASS, and Zylia in the room BPH yielded the lowest mean ratings, suggesting a clearly perceptible darker tone color obtained with these rendering methods than the reference. The t tests revealed a significant difference from zero for all rendering methods except MTB8 in BOL, Zylia in BOL, em32 in both rooms, HØSMA in BPH, Zylia C in BPH, em32 C in BOL, and HØSMA C in both rooms (see Table 28 in the supplementary materials [59]).

Table 3. Fixed-effects results from the LMM on each SAQI attribute, including F values, degrees of freedom (df), and p values.

Effect	F	df	p
Externalization			
Rendering method	10.69	14	< 0.001
Room	0.0005	1	0.981
Rendering method × room	1.60	14	0.072
Tone color			
Rendering method	76.71	14	< 0.001
Room	1.87	1	0.172
Rendering method × room	6.03	14	< 0.001
Source position			
Rendering method	75.73	14	< 0.001
Room	59.77	1	< 0.001
Rendering method × room	3.78	14	< 0.001
Presence			
Rendering method	27.94	14	< 0.001
Room	0.83	1	0.363
Rendering method × room	3.23	14	< 0.001

For source position, the model yielded $R^2_{\text{conditional}} = 50.9\%$ and $R^2_{\text{marginal}} = 38.4\%$ and showed significant effects for the fixed effects rendering method and room and for the rendering method × room interaction (see Table 3). Also for source position, many of the rendering methods led to clear deviations from the reference, as can be seen in Fig. 2(e), which shows the marginal means of the interaction. Here, especially Ambeo and Stereo in the room BPH stand out, indicating strong spatial impairments with these rendering methods compared to the FABIAN reference. Furthermore, HØSMA C, EMA N7 C, and Ambeo C obtained mean values clearly different from zero, suggesting that the image was also slightly impaired when using these rendering methods. The t tests revealed a significant difference from zero for all renderings in all rooms (see Table 29 in the supplementary materials [59]).

The model for the attribute presence yielded $R^2_{\text{conditional}} = 27.1\%$ and $R^2_{\text{marginal}} = 21.3\%$ and showed significant effects for the fixed effect rendering method and the rendering method × room interaction (see Table 3). The corresponding marginal means plot of the interaction in Fig. 2(f) indicates that participants' perceived presence depended considerably on the rendering method and partly on the room. The presence ratings show the lowest means for HØSMA C, followed by EMA N4 and stereo and the highest means for Zylia C and MTB16. The t tests revealed a significant difference from 0 for MTB16 in BOL, EMA N4 in both rooms, em32 in BPH, HØSMA in BPH, EMA N4 C in BPH, Zylia C in BOL, and HØSMA C in both rooms (see Table 30 in the supplementary materials [59]).

3.3 Comparison of Overall Listening Experience and Spatial Audio Quality Inventory Ratings

The model analyzing the influence of the SAQI attributes on OLE accounts for $R^2_{\text{conditional}} = 32.2\%$ of the variance of fixed and random effects and $R^2_{\text{marginal}} = 28.6\%$ of the fixed effects only. It revealed significant fixed effects for rendering method and room and the four SAQI attributes externalization, tone color, source position, and presence, suggesting that all assessed SAQI attributes significantly predict OLE (see Table 4). Among the SAQI attributes, presence is the strongest predictor of OLE, with the highest standardized coefficient ($\beta = 0.162$), indicating that an increase of one standard deviation in presence is associated with a meaningful improvement in OLE. In comparison, the effects of externalization ($\beta = 0.062$), tone color ($\beta = 0.071$), and source position ($\beta = -0.073$) were statistically significant, although they were smaller in magnitude. These results suggest that, even though all SAQI attributes contribute to OLE, presence has a particularly strong influence.

Table 4. Fixed effects results from the LMM predicting OLE, including F values, degrees of freedom (df), and p values. SAQI attributes were included as standardized (z-scores) covariates.

Effect	F	df	p
Rendering method	18.10	14	< 0.001
Room	5.73	1	0.017
Externalization	5.43	1	0.020
Tone color	4.76	1	0.029
Source position	4.95	1	0.026
Presence	31.91	1	< 0.001
Rendering method × room	1.10	14	0.354

Fig. 3 illustrates the predicted OLE scores (estimated marginal means) as a function of each standardized SAQI attribute, based on the fitted LMM. The solid blue lines show the model's predicted OLE scores for each z-scored attribute. The shaded areas indicate the corresponding 95% CIs. The steeper slope for presence compared with the other attributes highlights its stronger effect, as revealed by the model's predictions.

4 DISCUSSION

The presented perceptual evaluation of various microphone array–rendering method combinations provides insightful details about the current state of binaural rendering of microphone array recordings. One clear result of the study is that the HØSMA array in combination with the COMPASS rendering received low ratings for OLE and almost all SAQI attributes. However, this was not observed for the Ambisonic decodings, which are based on the same input as the COMPASS renderings (i.e., the HØSMA Ambisonic signals). Therefore, it can be assumed that the COMPASS decoding of higher orders does not perform optimally with the applied parametrization. One possible explanation is the relatively small operational bandwidth of the HØSMA array, which is caused by the the rather low spatial aliasing frequency due to the array's comparably large diameter of 23.5 cm and the radial filter limitation. These factors could affect the DOA estimation, which is an essential step in the COMPASS encoding.

Furthermore, in the default COMPASS parameterization, the number of sources for which the DOAs are estimated and are rendered in the encoding, scales with the input Ambisonic order as $(N + 1)^2$. Rendering such a large number

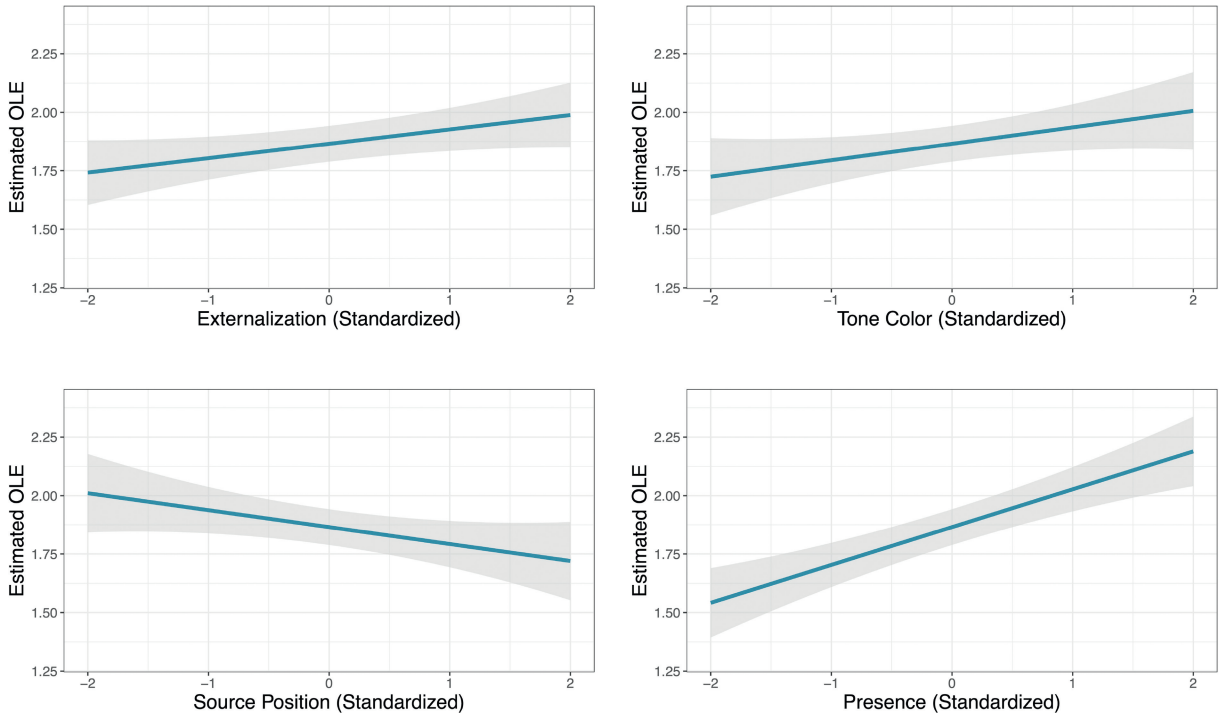


Fig. 3. Predicated OLE scores (estimated marginal means) as a function of standardized (z-scores) SAQI attributes. The blue lines show the predicted scores and the gray shaded areas indicate 95% CIs.

of individual sources can also lead to audible artifacts such as comb filtering. A solution might be to limit the maximum number of detected DOAs and thus the maximum number of rendered sound sources. McCormack et al. [34] observed similar issues with overestimation of the number of sources and the resulting degradation of perceptual quality. They proposed an alternative estimation approach that is, however, notably more computationally demanding and may therefore be unsuitable for real-time applications.

It should be noted that the HØSMA array is a self-built prototype and may therefore contain minor manufacturing inaccuracies. However, analytical radial filters (included in the SPARTA `array2sh` plug-in) were applied, which assume perfect spherical geometry and ideal microphone placement. This may lead to nonideal estimation of the SH signals. Measurement-based radial filters or direct estimation of the SH signals using measured steering vectors (see [63]) might better account for these imperfections. Yet, this factor alone is unlikely to explain the poor performance of HØSMA in combination with COMPASS. Because HØSMA with COMPASS consistently performed poorly in the evaluation, the HØSMA COMPASS renderings will not be discussed in more detail in the following overall comparison of the results.

Another general observation is that there are no significant differences in the results obtained in Berlin and Cologne for any of the tested attributes. This indicates good reproducibility of the results, which is notable given the limited number of studies that have conducted the same listening experiment in multiple locations.

Regarding the OLE ratings, MTB renderings, COMPASS renderings of Zylia and em32, and Ambisonic decoding of the em32 were rated the best, closely followed by EMA N7. Remarkably, the MTB method, with its simple processing without HRTF integration, achieved the highest mean ratings, even though not significantly higher than the reference FABIAN recording. One possible explanation is that the MTBs employed for the recordings in this study are equipped with high-quality Sennheiser KE14 electret condenser capsules. Moreover, no spatial aliasing or any other artifacts introduced by complex signal processing affect the recorded signals. The fact that the HØSMA microphone is equipped with the same capsules as the MTBs yet the Ambisonic renderings of the HØSMA have generally received lower ratings than the MTB renderings suggests that the perceived differences are due rather to artifacts induced by Ambisonic processing.

Similarly, the em32 is a high-quality microphone array with superior microphone capsules compared with the Zylia. The Ambeo is also of high quality, but the limited order $N = 1$ may have negatively affected the OLE ratings. The high OLE ratings of the COMPASS renderings of Zylia and em32 indicate the gain in quality through parametric processing, which reduces the artifacts of linear Ambisonic processing (i.e., spatial aliasing and order truncation).

The trends seen in the OLE ratings are similarly reflected in the SAQI results, with the MTB renderings, COMPASS renderings of Zylia and em32, and Ambisonic rendering of em32 performing best across all SAQI attributes. Close behind was the EMA N7, which was rated only slightly lower for the attribute tone color.

For the attribute externalization, most of the microphone method combinations were rated close to the reference. With the exception of Zylia C, which was rated slightly

better than the reference, it can hence be concluded that all tested methods lead to the same perceived externalization as the reference artificial head recording, even the Stereo reproduction without head tracking.

The strongest differences across rendering method can clearly be observed for tone color and source position. The highest differences in tone color were obtained for Ambeo, Ambeo C, and Stereo. Because all three methods are based on measurements taken with the Ambeo microphone array, the ratings are likely related to the microphone's characteristics combined with the manufacturer's Ambisonic encoder. The tone color of the Ambisonic renderings of EMA N4, EMA N7, and their COMPASS renderings were rated as slightly darker, indicating a coloration difference in the EMA encoding. This can likely be mitigated by a more refined global equalization.

Similar to the tone color, Ambeo, Ambeo C, and Stereo showed the strongest difference in source position. This is probably caused by the low spatial resolution of the first-order Ambisonic encoding and the associated blurry sound image, or mislocalization caused by spatial aliasing, mainly affecting higher time frequency components of certain instruments. Notably, the EMA encodings rendered with COMPASS seem to evoke differences in the source position as well, which could not be observed for the Ambisonic decodings of the EMA. This is probably due to the missing information of elevated sound incidences in the EMA encoding, which might conflict with the DOA estimation of the COMPASS algorithm that is per default performed on a uniformly distributed grid.

For the presence ratings, higher variances were observed between the methods. MTB16, em32, and Zylia C were rated significantly higher; EMA N4, HØSMA, EMA N4 C, EMA N7 C, and Stereo were rated significantly lower compared to the FABIAN recording.

The results further show that the attribute presence had the strongest impact on OLE, suggesting that a strong sense of presence in the scene leads to a greater OLE. However, the assessed attributes externalization, tone color, and source position also contributed to the OLE, although to a lesser extent.

A potential limitation of this study is that signal-dependent parametric approaches may inherently perform differently across various acoustic scenes. Because both examined scenarios are recordings of the same orchestra in rather reverberant rooms, the findings may not be generalizable. A broader range of acoustic scenes could therefore provide even more comprehensive insights. Nevertheless, evaluating a large number of array-rendering method combinations under real-world conditions, as was done in this study, provides valuable insight and practical guidance. Including additional acoustic scenes would have exceeded the scope of this work.

5 CONCLUSION

This study provided a comprehensive perceptual comparison of different microphone array and binaural rendering combinations. The results offer detailed insights into the most suitable combination of microphone array and rendering method for achieving a high OLE while preserving the assessed perceptual SAQI attributes. In particular, MTB and the COMPASS renderings of em32 and Zylia received the highest OLE ratings, whereas MTB also performed well across all SAQI attributes. In general, the MTB method produced convincing results across all assessed parameters. One reason for its good overall performance may be the minimal signal processing combined with high-quality microphones, which leads to artifact-free pseudobinaural signals.

Overall, many of the assessed array-method combinations performed similarly in the tested scenario and were perceptually close to the head and torso simulator reference. Consistent with this finding, even static Stereo rendering received reasonably good ratings. However, specific methods lead to certain impairments. For example, Ambeo renderings exhibit noticeable coloration and source position impairments, likely due to the characteristics of the Ambeo microphone, its low order, and the Ambisonic encoder. Moreover, COMPASS seems to encounter problems with higher-order Ambisonic signals with low operational bandwidth (e.g., HØSMA with $N = 7$), likely due to limitations in the parameterization.

Future research should conduct similar experiments across a wider range of recording scenarios. To develop a more comprehensive understanding of the various rendering methods, these scenarios could include sound field configurations that differ from an orchestra, with either fewer or more dominant sound sources.

6 ACKNOWLEDGMENT

The authors would like to thank Leo McCormack for providing the COMPASS renderings. They also thank all the voluntary participants for their support.

6 REFERENCES

[1] V. R. Algazi, R. O. Duda, and D. M. Thompson, "Motion-Tracked Binaural Sound," *J. Audio Eng. Soc.*, vol. 52, no. 11, pp. 1142–1156 (2004 Nov.). https://aes.org/publications/elibrary-page/?id=13028.

[2] D. Ackermann, F. Fiedler, F. Brinkmann, M. Schneider, and Weinzierl Stefan, "On the Acoustic Qualities of Dynamic Pseudobinaural Recordings," *J. Audio Eng. Soc.*, vol. 68, no. 6, pp. 418–427 (2020 Jun.). https://doi.org/10.17743/jaes.2020.0036.

[3] B. Rafaely, *Fundamentals of Spherical Array Processing* (Springer, Berlin, Germany, 2019), 2nd ed. https://doi.org/10.1007/978-3-319-99561-8.

[4] F. Zotter and M. Frank, *Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Springer Nature, Cham, Switzerland, 2019). https://doi.org/10.1007/978-3-030-17207-7.

[5] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial Perception of Sound Fields Recorded by Spherical Microphone arrays With Varying Spatial Resolution," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2711–2721 (2013 May). https://doi.org/10.1121/1.4795780.

[6] Z. Ben-Hur, J. Sheaffer, and B. Rafaely, "Joint Sampling Theory and Subjective Investigation of Plane-Wave and Spherical Harmonics Formulations for Binaural Reproduction," *Appl. Acoust.*, vol. 134, pp. 138–144 (2018 May). https://doi.org/10.1016/j.apacoust.2018.01.016.

[7] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness Stability of Binaural Sound With Spherical Harmonic Representation of Sparse Head-Related Transfer Functions," *EURASIP J. Audio Speech Music Process.*, vol. 2019, paper 5 (2019 Mar.). https://doi.org/10.1186/s13636-019-0148-x.

[8] T. Lübeck, J. M. Arend, and C. Pörschmann, "Binaural Reproduction of Dummy Head and Spherical Microphone Array Data—A Perceptual Study on the Minimum Required Spatial Resolution," *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 467–483 (2021 Jan.). https://doi.org/10.1121/10.0009277.

[9] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral Equalization in Binaural Signals Represented by Order-Truncated Spherical Harmonics," *J. Acoust. Soc. Am.*, vol. 141, no. 6, pp. 4087–4096 (2017 Jun.). https://doi.org/10.1121/1.4983652.

[10] F. Salmon, G. Berthomieu, J. Palacino, and M. Paquier, "The Influence of Diffuse-Field Equalization on the Perception of Spatial Aliasing Introduced by Spherical Microphone Arrays," *J. Audio Eng. Soc.*, vol. 72, no. 10, pp. 650–663 (2024 Feb.). https://doi.org/10.17743/jaes.2022.0157.

[11] C. Schörkhuber, M. Zaunschirm, and R. Holdrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in Proc. *44th DAGA*, vol. 44, pp. 339–342 (Munich, Germany) (2018 Mar.).

[12] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *J. Audio Eng. Soc.*, vol. 60, no. 10, pp. 807–820 (2012 Oct.).

[13] T. Deppisch, H. Helmholz, and J. Ahrens, "End-to-End Magnitude Least Squares Binaural Rendering of Spherical Microphone Array Signals," in Proc. *Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–7 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/i3da48870.2021.9610864.

[14] Institute of Electronic Music and Acoustics, "IEM Plug-in Suite" https://plugins.iem.at (assessed May 19, 2025).

[15] L. McCormack and A. Politis, "SPARTA & COMPASS: Real-Time Implementations of Linear and Parametric Spatial Audio Reproduction and Processing Methods," in Proc. *AES Conf. Immersive and Interaktive Audio* (2019 Mar.), paper 111. https://aes.org/publications/elibrary-page/?id=20417.

[16] H. Helmholz, C. Andersson, and J. Ahrens, "Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals," in Proc. *45th DAGA*, pp. 2–5 (Rostock, Germany) (2019 Mar.).

[17] H. Helmholz, T. Lübeck, J. Ahrens, S. V. A. Garí, D. L. Alon, et al., "Updates on the Real-Time Spherical Array Renderer (ReTiSAR)," in Proc. *46th DAGA*, pp. 1169–1172 (Hannover, Germany) (2020 Mar.).

[18] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. A. Garí, "Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer," *J. Acoust. Soc. Am.*, vol. 150, pp. 805–815 (2021 Aug.). https://doi.org/10.1121/10.0005754.

[19] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "Spherical Harmonic Decomposition of a Sound Field Using Microphones on a Circumferential Contour Around a Non-Spherical Baffle," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 3110–3119 (2022 Sep.). https://doi.org/10.1109/TASLP.2022.3209940.

[20] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, Massachusetts, 1996).

[21] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering," in Proc. *7th Intl. Conf. Digital Audio Effects (DAFx)*, pp. 139–144 (Naples, Italy) (2004 Oct.).

[22] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall, "Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution," *J. Audio Eng. Soc.*, vol. 68, no. 5, pp. 338–354 (2020 May). https://aes.org/publications/elibrary-page/?id=20852.

[23] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28 (2013 Jan./Feb.). https://aes.org/publications/elibrary-page/?id=16664.

[24] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516 (2007 Jun.). https://aes.org/publications/elibrary-page/?id=14170.

[25] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 5, pp. 852–866 (2015 Aug.). https://doi.org/10.1109/JSTSP.2015.2415762.

[26] B. Bernschütz, *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Ph.D. thesis, Technische Universität Berlin, Berlin, Germany (2016 Feb.). https://doi.org/10.14279/depositonce-5082.

[27] J. Ahrens and C. Andersson, "Perceptual Evaluation of Headphone Auralization of Rooms Captured With Spherical Microphone Arrays With Respect to Spaciousness and Timbre," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2783–2794 (2019 Apr.). https://doi.org/10.1121/1.5096164.

[28] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural Rendering of Ambisonic Signals by Head-Related Impulse Response Time Alignment and a Diffuseness Constraint," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627 (2018 Jun.). https://doi.org/10.1121/1.5040489.

[29] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *J. Audio Eng. Soc.* vol. 68, no. 6, pp. 428–440 (2020 Jun.) https://doi.org/10.17743/jaes.2020.0038.

[30] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments Due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data: Dry Acoustic Environments," in Proc. *23rd Intl. Conf. Digital Audio Effects (DAFx)*, pp. 250–257 (Vienna, Austria) (2020 Sep.).

[31] H. Helmholz, J. Crukley, S. V. Amengual Garí, Z. Ben-Hur, and J. Ahrens, "Perceived Quality of Binaural Rendering From Baffled Microphone Arrays Evaluated Without an Explicit Reference," *J. Audio Eng. Soc.*, vol. 72, no. 10, pp. 691–704 (2024 Feb.). https://doi.org/10.17743/jaes.2022.0164.

[32] A. Politis, S. Tervo, and V. Pulkki, "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in Proc. *IEEE Intl. Conf. Acoustics, Speech and Signal Processing*, pp. 6802–6806 Calgary, Canada) (2018 Apr.). https://doi.org/10.1109/ICASSP.2018.8462608.

[33] S. V. Amengual Garí, J. M. Arend, P. Calamia, and P. W. Robinson, "Optimizations of the Spatial Decomposition Method for Binaural Reproduction," *J. Audio Eng. Soc.*, vol. 68, no. 12, pp. 959–976 (2020 Dec.). https://doi.org/10.17743/jaes.2020.0063.

[34] L. McCormack, N. Meyer-Kahlen, and A. Politis, "Spatial Reconstruction-Based Rendering of Microphone Array Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 71, no. 5, pp. 267–280 (2023 May). https://doi.org/10.17743/jaes.2022.0072.

[35] A. Pawlak, H. Lee, A. Mäkivirta, and T. Lund, "Spatial Analysis and Synthesis Methods: Subjective and Objective Evaluations Using Various Microphone Arrays in the Auralization of a Critical Listening Room," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 32, pp. 3986–4001, (2024 Aug.). https://doi.org/10.1109/TASLP.2024.3449037.

[36] L. McCormack and S. Delikaris-Manias, "Parametric First-Order Ambisonic Decoding for Headphones Utilising the Cross-Pattern Coherence Algorithm," in Proc. *EAA Spatial Audio Signal Processing Symposium*, pp. 173–178 (Paris, France) (2019 Sep.). https://doi.org/10.25836/sasp.2019.26.

[37] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric Ambisonic Encoding of Arbitrary Microphone Arrays," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 2062–2075 (2022 Jun.). https://doi.org/10.1109/TASLP.2022.3182857.

[38] H. Helmholz, J. Ahrens, D. L. Alon, S. V. Amengual Garí, and R. Mehra, "Evaluation of Sensor Self-Noise In Binaural Rendering of Spherical Microphone Array Signals," in Proc. *Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 161–165 (Barcelona, Spain) (2020 May). https://doi.org/10.1109/icassp40776.2020.9054434.

[39] D. Ackermann, J. Domann, F. Brinkmann, et al., "Recordings of a Loudspeaker Orchestra With Multichannel Microphone Arrays for the Evaluation of Spatial Audio Methods," *J. Audio Eng. Soc.*, vol. 71, no. 1/2, pp. 62–73 (2023 Jan.). https://doi.org/10.17743/jaes.2022.0059.

[40] A. Lindau, T. Hohn, and S. Weinzierl, "Binaural Resynthesis for Comparative Studies of Acoustical Environments," presented at the *122nd Conv. Audio Engineering Society* (2007 May), paper 7032. https://aes.org/publications/elibrary-page/?id=14017.

[41] Sennheiser electronic SE & Co. KG, "3D Audio Microphone AMBEO VR Mic," (2025). https://www.sennheiser.com/en-dk/catalog/products/microphones/ambeo-vr-mic/ambeo-vr-mic-507195 (accessed 18 Oct. 2025).

[42] "ZYLIA – 3D Audio Recording & Post-Processing Solutions," https://www.zylia.co/ (accessed 18 Oct. 2025).

[43] mh acoustics LLC, "Products," https://mhacoustics.com/products (accessed 18 Oct. 2025).

[44] J. Meyer and G. Elko, "A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield," in Proc. *IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1781–1784 (Orlando, FL.)(2002 May). https://doi.org/10.1109/ICASSP.2002.5744968.

[45] O. Moschner, D. Dziwis, T. Lübeck, and C. Pörschmann, "Development of an Open Source Customizable High Order Rigid Sphere Microphone Array," presented at the *148th Conv. Audio Engineering Society*, (2020 May), paper 583. https://aes.org/publications/elibrary-page/?id=20821.

[46] K. Brunnström, S. Beker, K. De Moor, et al. "Qualinet White Paper on Definitions of Quality of Experience," in Proc. *Output From the Fifth Qualinet Meeting*, vol. 5, pp. 1–25 (Novi Sad, Serbia) (2013 Mar.).

[47] M. Schoeffler, S. Conrad, and J. Herre, "The Influence of the Single/Multi-Channel-System on the Overall Listening Experience," in Proc. *55th AES Intl. Conf.* (Helsinki, Finland) (2014 Aug.).

[48] ITU-R, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," *Recommendation ITU-R BS.1534-3*, (2015 Oct.). https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en.

[49] M. Schoeffler and J. Herre, "The Relationship Between Basic Audio Quality and Overall Listening Experience," *J. Acoust. Soc. Am.*, vol. 140, no. 3, pp. 2101–2112 (2016 Sep.). https://doi.org/10.1121/1.4963078.

[50] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, S. Weinzierl, et al., "A Spatial Audio Quality Inventory (SAQI)," *Acta Acust. United Ac.*, vol. 100, no. 5, pp. 984–994 (2014 Sep./Oct.). https://doi.org/10.3813/AAA.918778.

[51] M. Geier, J. Ahrens, and S. Spors, "SoundScape Renderer," http://spatialaudio.net/ssr/ (accessed 18 Mar. 2026).

[52] M. Geier and S. Spors, "Spatial Audio With the SoundScape Renderer," in Proc. *27th Tonmeistertagung - VDT Intl. Conv.*, p. 2012 (Cologne, Germany) (2012 Nov.).

[53] A. Lindau and S. Roos, "Perceptual Evaluation of Discretization and Interpolation for Motion-Tracked Binaural (MTB) Recordings," in Proc. *26th Tonmeistertagung*, pp. 4087–4096 (Cologne, Germany) (2010 Jan.).

[54] J. Ahrens, "Ambisonic Encoding of Signals From Equatorial Microphone Arrays," Tech. Rep. v. 1 (2022 Sep.), https://arxiv.org/abs/2211.00584.

[55] L. McCormack and A. Politis, "Estimating and Reproducing Ambience in Ambisonic Recordings," in Proc. *30th European Signal Processing Conf. (EUSIPCO)*, pp. 314–318 (Belgrade, Serbia) (2022 Aug./Sep.). https://doi.org/10.23919/EUSIPCO55093.2022.9909850.

[56] S. Weinzierl, "Stereofone Aufnahmeverfahren," in S. Weinzierl (Ed.), *Handbuch der Audiotechnik*, pp. 511–536 (Springer Vieweg, Berlin, Germany, 2024). https://doi.org/10.1007/978-3-662-60357-4_21-1.

[57] G. Kailas and N. Tiwari, "An Empirical Measurement Tool for Overall Listening Experience of Immersive Audio," in Proc. *IEEE Intl. Conf. Consumer Electronics*, pp. 1–5 (Las Vegas, NV) (2021 Jan.). https://doi.org/10.1109/ICCE50685.2021.9427770.

[58] T. Walton, "The Overall Listening Experience of Binaural Audio," in Proc. *4th Intl. Conf. Spatial Audio (ICSA)* (Graz, Austria) (2017 Sep.).

[59] T. Lübeck, C. Scheer, D. Ackermann, F. Brinkmann, C. Pörschmann, S. Weinzierl, J. Ahrens, and J. M. Arend, "Supplementary Material for 'Perceptual Evaluation of Different Methods for Binaural Rendering of Recordings with Various Microphone Arrays," *Zenodo* (2026 Apr.), https://doi.org/10.5281/zenodo.17420722.

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

[61] C. F. Dormann, J. Elith, S. Bacher, et al., "Collinearity: A Review of Methods to Deal With It and a Simulation Study Evaluating Their Performance," *Ecog.*, vol. 36, no. 1, pp. 27–46 (2013 Jan.). https://doi.org/10.1111/j.1600-0587.2012.07348.x.

[62] M. Gallucci, "GAMLj: General Analyses for Linear Models," https://gamlj.github.io/ (accessed 7 May 2025).

[63] A. Politis and H. Gamper, "Comparing Modeled and Measurement-Based Spherical Harmonic Encoding Filters for Spherical Microphone Arrays," in Proc. *IEEE Workshop of Applications of Signal Processing to Audio and Acousitics*, pp. 224–228 (New Paltz, NY) (2017 Oct.). https://doi.org/10.1109/WASPAA.2017.8170028.

THE AUTHORS


Tim Lübeck


Christian Scheer


David Ackermann


Fabian Brinkmann


Christoph Pörschmann


Stefan Weinzierl


Jens Ahrens


Johannes M. Arend

Tim Lübeck received his B.Sc. degree in electrical engineering in 2017 and his M.Sc. degree in communication engineering in 2019 from TH Köln, Cologne, Germany. He completed his master's thesis in cooperation with the Division of Applied Acoustics at Chalmers University in Gothenburg. Since 2019, he has been a research fellow and working toward the Ph.D. degree at TH Köln and Technische Universität Berlin in virtual acoustics, binaural technology, auditory perception, and audio signal processing.

•

Christian Scheer is a research associate and Ph.D. candidate in the Audio Communication Group at Technische Universität Berlin. He received his B.Eng. degree in sound and video engineering from Hochschule Düsseldorf and Robert Schumann Hochschule Düsseldorf in 2021 and his M.Sc. degree in audio communication and technology from TU Berlin in 2024. His research interests include spatial audio, virtual acoustic realities, and psychoacoustics.

•

David Ackermann received his M.Sc. in audio communication and technology in 2015 and his Ph.D. (Dr. rer. nat.) in 2024 from Technische Universität Berlin. He is a professor of audio engineering at the University of Applied Sciences Duesseldorf, where his research focuses on virtual and musical acoustics and the further development of binaural reproduction methods for live streaming of 3D audio.

•

Fabian Brinkmann received an M.A. degree in communication sciences and technical acoustics in 2011 and Dr. rer. nat. degree in 2019 from the Technische Universität Berlin, Germany. He focuses on the fields of signal processing and evaluation approaches for spatial audio. Fabian is a member of the Audio Engineering Society (AES), German Acoustical Society (DEGA), and the European Acoustics Association (EAA) technical committee for psychological and physiological acoustics.

•

Christoph Pörschmann studied electrical engineering at the Ruhr-Universität Bochum (Germany) and Uppsala Universitet (Sweden). In 2001 he obtained his doctoral degree (Dr.-Ing.) from the Electrical Engineering and Information Technology Faculty of the Ruhr-Universität Bochum as a result of his research at the Institute of Communication Acoustics. Since 2004, he has been professor of acoustics at TH Köln (Germany). His research interests are in the field of virtual acoustics, spatial hearing, and the related perceptual processes.

•

Stefan Weinzierl is head of the Audio Communication Group at the Technische Universität Berlin. His research is focused on audio technology, virtual acoustics, room acoustics, and musical acoustics. With a diploma in physics and sound engineering, he received his Ph.D. in musical acoustics. He is coordinating a master program in audio communication and technology at Technische Universität Berlin and has coordinated international research consortia in the field of virtual acoustics (SEACEN) and music information retrieval (ABC_DJ).

•

Jens Ahrens has been an associate professor within the Division of Applied Acoustics at Chalmers University since 2016. He has also been a visiting professor at the Applied Psychoacoustics Laboratory at University of Huddersfield, United Kingdom, since 2018. Jens received his Diplom (equivalent to an M.Sc.) in electrical engineering/sound engineering jointly from Graz University of Technology and the University of Music and Dramatic Arts, Graz, Austria, in 2005. He completed his doctoral degree (Dr.-Ing.) at the Technische Universität Berlin, Germany, in 2010. From 2011 to 2013, Jens was a postdoctoral researcher at Microsoft Research in Redmond, Washington, USA, and in the fall and winter terms of 2015/16, he was a visiting scholar at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University, California, USA.

•

Johannes M. Arend is assistant professor at the Acoustics Lab, Department of Information and Communication Engineering, Aalto University, Finland. He leads the Technical Psychoacoustics research group, working at the intersection of technical developments in virtual acoustics and spatial audio, the perceptual evaluation of these technologies, and their application in hearing science studies on auditory and multisensory perception. Before joining Aalto University in 2024, he was a postdoctoral researcher at Technische Universität Berlin from 2022 to 2024 and a research associate at TH Köln from 2015 to 2022. He received the B.Eng. degree in media technology from Hochschule Düsseldorf, Germany, in 2011, the M.Sc. degree in media technology from TH Köln, Germany, in 2014, and the Ph.D. degree (Dr. rer. nat.) from Technische Universität Berlin, Germany, in 2022.