OPEN ACCESS Freely available online
D. Ackermann, J. Domann, F. Brinkmann, J. M. Arend, M. Schneider, C. Pörschmann, and S. Weinzierl, "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 January/February).
DOI: https://doi.org/10.17743/jaes.2022.0059

Recordings of a Loudspeaker Orchestra With Multichannel Microphone Arrays for the Evaluation of Spatial Audio Methods

DAVID ACKERMANN,[1,*] **JULIAN DOMANN,**[1] **FABIAN BRINKMANN,**[1] **JOHANNES M. AREND,**[1]
(david.ackermann@tu-berlin.de) (julian.domann@posteo.de) (fabian.brinkmann@tu-berlin.de) (j.arend@tu-berlin.de)

MARTIN SCHNEIDER,[2] **CHRISTOPH PÖRSCHMANN,**[3] **AND STEFAN WEINZIERL**[1]
(martin.schneider@neumann.com) (christoph.poerschmann@th-koeln.de) (stefan.weinzierl@tu-berlin.de)

[1]*Audio Communication Group, Technische Universität Berlin, Berlin, Germany*
[2]*Georg Neumann GmbH, Berlin, Germany*
[3]*Institute of Communications Engineering, TH Köln – University of Applied Sciences, Köln, Germany*

For live broadcasting of speech, music, or other audio content, multichannel microphone array recordings of the sound field can be used to render and stream dynamic binaural signals in real time. For a comparative physical and perceptual evaluation of conceptually different binaural rendering techniques, recordings are needed in which all other factors affecting the sound (such as the sound radiation of the sources, the room acoustic environment, and the recording position) are kept constant. To provide such a recording, the sound field of an 18-channel loudspeaker orchestra fed by anechoic recordings of a chamber orchestra was captured in two rooms with nine different receivers. In addition, impulse responses were recorded for each sound source and receiver. The anechoic audio signals, the full loudspeaker orchestra recordings, and all measured impulse responses are available with open access in the Spatially Oriented Format for Acoustics (SOFA 2.1, AES69-2022) format. The article presents the recording process and processing chain as well as the structure of the generated database.

0 INTRODUCTION

In the last 20 years, research on the analysis and synthesis of sound fields has developed a variety of methods that have the potential to provide a more immersive transmission of music and speech. In recording, either binaural or pseudo-binaural systems or different types of microphone arrays can be used for this purpose. The former are usually played back directly via headphones; the latter perform sound field decomposition, which can either be played back via loudspeaker systems or can also be binauralized by downstream signal processing using head-related transfer functions (HRTFs) or binaural room impulse responses (BRIRs) of the loudspeaker setup.

In binaural playback, only a dynamic rendering, i.e., one that takes into account the listener's head movement, is able to achieve a *plausible* synthesis of acoustic scenes [1], avoid front-back-ambiguity and localization errors [2], and provide a degree of externalization [3] close to natural hearing. If individually measured BRIRs are used, even *authentic* synthesis can be realized, i.e., one that cannot be distinguished from the original sound field even in a direct auditory comparison [4].

However, this method can only be employed if anechoic audio content is available, as well as a complete data set of transfer functions from all sources to all head orientations of the receiver. In the live broadcasting of concerts, theater performances, sports events, or panel discussions, these requirements are usually not met. A possible solution to achieve a dynamic simulation in such situations is to render and stream the binaural signals from microphone array recordings, i.e., from recordings of multiple, spatially distributed microphones. A multitude of conceptually different methods is available for this approach.

A conceptually simple approach is motion-tracked binaural (MTB) sound [5]. It allows the recording of pseudo-binaural signals by an equatorial, circular microphone array embedded in the surface of a rigid sphere. In contrast to a dummy head recording, MTB signals do not convey pinna cues but can provide dynamic cues through interpolating the microphone signals according to the positions of the

*Correspondence should be addressed to David Ackermann. E-mail: david.ackermann@tu-berlin.de

listener's ears. Through this technique, a good externalization, a highly plausible spatial sound image, and even a perception of elevated sound incidence directions can be achieved. [6].

In *parametric* spatial audio techniques, the sound field is usually recorded with multichannel microphones, and its characteristics are analyzed for different time windows and frequency bands and represented by a set of parameters [7]. This general approach includes different techniques such as first- and higher-order directional audio coding (FO/HO-DirAC) [8, 9], the spatial decomposition method (SDM) [10, 11], the high angular resolution planewave expansion method (HARPEX) [12, 13] or the multiwave sound field model [14]. Among the more recent parametric rendering approaches are COMPASS to binaural [15], the Ambisonic decoding for headphones utilizing the cross-pattern coherence algorithm [16], and the linearly and quadratically constrained least-squares (LQC-LS) decoder [17]. These approaches are often flexible in the sense that they can work with different microphone array layouts and generally require a moderate number of between four and 16 microphones.

The use of spherical microphone arrays (SMAs) offers flexible *nonparametric* methods to capture and reproduce spatial sound fields. In this case, the sound field is decomposed into higher-order spherical harmonics (SH) basis functions. This representation can either be used directly for Ambisonics rendering [18] or converted to a plane wave representation (PWD [19, 20]). Both approaches provide generic representations of the spatial sound field, for which signal processing approaches for loudspeaker and headphone reproduction are available [21, 18]. Recent advances have produced several solutions for real-time binaural rendering of SMA signals, such as the IEM Plug-in Suite[1] [22], SPARTA [23], and ReTiSAR [24, 25]. SMA-based approaches require spherical microphone arrays in the order of 16 to 64 capsules. For a comparative overview of a selection of state-of-the-art HRTF processing methods for the binaural rendering of (order-limited) Ambisonics signals, see [26].

For most of these approaches to immersive recording and playback of acoustic scenes, studies have also been conducted on the overall quality and on the qualities, i.e., the characteristic properties of the sound field synthesized in this way. A perceptual evaluation of binaural representations based on SMA captures was presented in [21, 27, 28] with respect to overall quality or with respect to specific attributes, using only one rendering approach, but various test conditions such as SH order, aliasing frequency, or complex amplitude calculation. The listening experiments conducted in [29] and [30] compared dynamic binaural renderings of SMA measurements with dynamic binaural renderings of dummy-head recordings of the same acoustic scene.

A comparative perceptual evaluation of different state-of-the-art nonparametric binaural rendering approaches for SMAs, such as the bandwidth extension algorithm for microphone arrays (BEMA) [31] that synthesizes the SH coefficients of the sound field in the higher time-frequency bands; the magnitude least-squares (MagLS) technique [32], a method for reducing the effects of truncation of order SH; the so-called spherical head filters (SHFs) [33], to compensate for the low-pass effect of SH order truncation and the method denoted as spherical harmonics tapering [34], to suppress the side lobes induced by order truncation, was presented by Lübeck et al. [35, 36].

All of the aforementioned studies are based on array room impulse responses (ARIRs). These were either numerically simulated using a room acoustic simulation framework or acquired with a virtual microphone array, equipped with only a single measurement microphone sequentially scanning the desired measurement grid. As a result, the data used in these experiments are free of microphone mismatches and were accordingly conducted under ideal conditions [30]. Furthermore, not all studies presented are based on freely available ARIRs data sets, which precludes reproducibility of the results shown.

For a comparative evaluation of parametric *and* nonparametric binaural reproduction methods for *live* performances, such as orchestral concerts, however, reproducible recordings of a complex acoustic scene with a variety of receivers are needed. At first glance, this could be realized by convolving ARIRs with anechoic audio content; however, this would make it hard to consider the signal-to-noise ratio of real recordings, which can considerably influence the results at least for some auralization approaches [37]. The direct recording of a real performance is also precluded, because not all recording devices can be positioned in the same place in order to record identical sound fields of a nonreproducible sound source. The only option is to use electroacoustic sound sources such as an orchestra of loudspeakers, fed by anechoic signals of acoustic sound sources such as singers, speakers, or musical instruments.

Although loudspeakers can only approximate the directional characteristics of musical instruments, not to mention the dynamic effects created by the pitch-dependent directionality of the instruments and the movements of musicians [38], this concept creates a reproducible stimulus that bears at least a high resemblance to the behavior of a physical orchestra. A similar approach was chosen by Pätynen et al. to compare the room acoustical conditions in different concert halls using a 34-channel loudspeaker orchestra and a six-channel vector intensity probe [39, 40]. These recordings, however, are also not accessible for subsequent use. More recent publications of freely accessible multichannel microphone array data with several sound sources include only ARIRs and no recordings of real sound fields [41–43].

In order to physically and perceptually evaluate a large variety of (pseudo-)binaural rendering methods, we recorded the sound field of a multichannel loudspeaker orchestra in two rooms and with nine different receivers. To provide an artistically convincing content, which can be used to evaluate the performance of music streaming techniques, the anechoic orchestra signals were recorded with a professional orchestra, and the sound of the loudspeaker orchestra was adjusted and balanced by professional sound

[1] https://plugins.iem.at, accessed June 2, 2022.

Table 1. Assignments of the loudspeakers to the reproduced instruments and the recording position of the microphone used in the anechoic chamber of the TU Berlin in the spherical coordinates azimuth ϕ, elevation θ, and radial distance r. A detailed description of the coordinate system used is shown in [45].

			Microphone Position		
Emitter	Loudspeaker	Instrument	ϕ in deg	θ in deg	r in m
E1	Neumann KH 120	Violin 1	−36	17	1.23
E2	Neumann KH 120	Violin 2	−67	18	1.60
E3	Neumann KH 120	Viola	−63	35	0.96
E4	Neumann KH 120	Cello	14	21	2.21
E5	Genelec 1031	Double bass	−40	10	1.19
E6	Genelec 8020	Piccolo flute	24	5	2.20
E7	Genelec 8020	Flute	24	5	2.20
E8	Genelec 8020	Oboe	0	21	2.14
E9	K+H O300D	Clarinet	22	5	2.16
E10	K+H O300D	Bass clarinet	18	16	1.98
E11	Neumann KH 120	Bassoon	−24	−16	2.05
E12	Genelec 1031	French horn	−102	−8	1.45
E13	Genelec 8331	Trumpet	−10	−8	2.87
E14	HEDD Type 07	Snare drum	32	36	1.88
E15	Genelec 8020 + KRK 12s	Bass drum	−36	15	1.98
E16	Genelec 8020	Triangle	32	36	1.88
E17	HEDD Type 07	Crash cymbals	32	30	2.18
E18	Genelec 8020	Harp	−24	5	2.42

engineers. All data are open access [44] in the Spatially Oriented Format for Acoustics (SOFA 2.1, AES69-2022) format [45].

The objective of this work was to achieve a reproducible representation of a real orchestra performance with a high degree of realism. To achieve this, the loudspeaker orchestra was mixed on site by sound engineers. Although differences to a real orchestra cannot be avoided, the emitted sound fields can be assumed to be of comparable complexity with respect to their spatio-temporal structure, which is to a large extend determined by the number of sources, the audio signals, and the response of the room. We thus considered the loudspeaker orchestra to be sufficient for comparing different microphone array based renderings under quasi-realistic conditions.

1 PRODUCTION

To enable a comparative evaluation of different spatial transmission methods, we recorded a loudspeaker orchestra as a reproducible sound source with nine different receiver systems in two different room acoustic environments. The orchestra included 18 instruments (loudspeakers), and the sound field was captured both as a live overall sound recording and as sequentially measured impulse responses for each individual loudspeaker and receiver capsule.

Following the AES69 naming convention, the microphones will be referred to as *Listeners*, and each capsule of a microphone as a *Receiver* in the following. The loudspeaker orchestra is termed the *Source* accordingly, with each loudspeaker being an *Emitter*. Metadata such as the position of an emitter or orientation of a listener are stored in fields named accordingly in the corresponding SOFA files (cf. Sec. 2).

1.1 Anechoic Recordings

An arrangement of the last movement (*Golliwogg's cake-walk*) of Claude Debussy's suite *Children's Corner* for 18 orchestral instruments (cf. Table 1) was used as source material, with a total playing time of 2:08 minutes. The piece was recorded with the Konzerthausorchester Berlin in the anechoic chamber of the TU Berlin, with a room volume of approximately 1.070 m^3 and a lower cut-off frequency of $f_c = 63$ Hz[2]. To obtain an anechoic and crosstalk-free audio recording, the musicians were placed one at a time in the chamber and recorded sequentially. To allow for musically accurate timing and intonation, a pilot track of the entire piece was initially produced and played to the subsequent instrumentalists via closed headphones using an individual monitor mix of their own direct sound and the pilot track.

A convolution reverb with a simulated binaural room impulse response suitable for orchestral performances was used to create a pleasant and familiar acoustic recording environment. To reduce latency and avoid duplicate imaging of the direct sound field, impulse responses without direct sound were convolved with the instrument signal and summed to the dry signal, which was sent directly from the mixing console with negligible latency. For recording the musical instruments, we used an NTi M2230 class 1 free-field equalized omnidirectional measurement microphone, consisting of the preamplifier MA220 and the microphone capsule MC230.

The exact microphone positions are given in spherical coordinates, i.e., in azimuth ($\phi = 0°$ pointing in positive x-direction, $\phi = 90°$ pointing in positive y-direction), elevation ($\theta = 0°$ pointing in positive x-direction, $\theta = 90°$

[2] A video documentation of the recordings is available at https://youtu.be/B9wYCZSrHTU (accessed July 27, 2022).

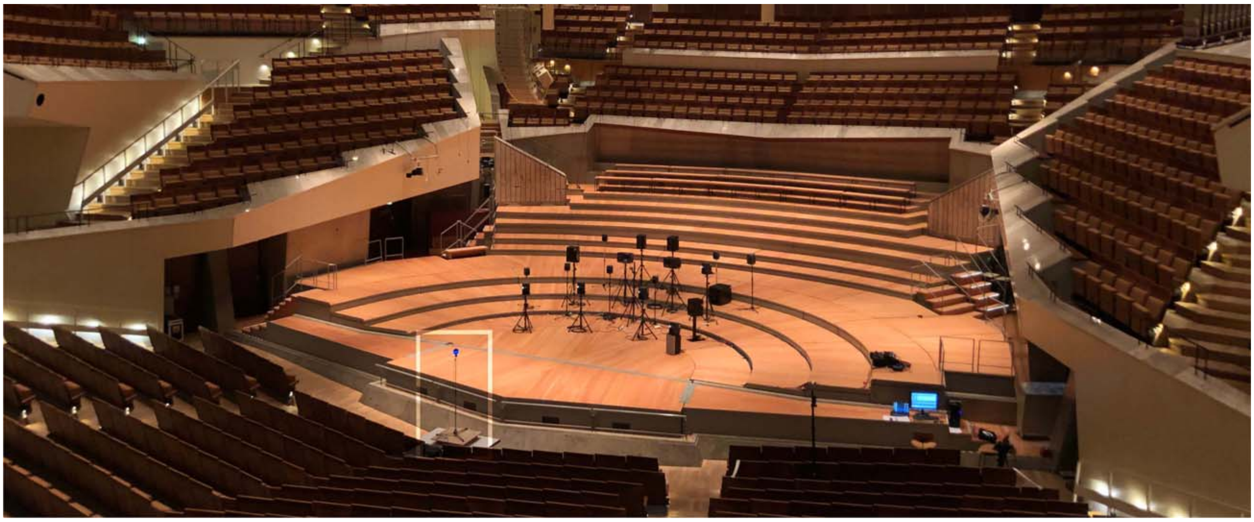


Fig. 1. Setup of the 18-channel loudspeaker orchestra on the stage in the large hall of the Berliner Philharmonie, with the slightly elevated listener position in the first row of the audience area.

pointing in positive z-direction) and distance (in meter) for each instrument (cf. Table 1). The origin of the coordinate system corresponds to the head position of the musicians. Thus, the origin is 1.2 m above the ground, the average ear position for seated people on a 0.45 m high chair [46]. The musicians' viewing direction is along the x-axis.

When processing the recorded audio, the recorded takes were edited by a professional audio engineer, correcting slight asynchronies, unclean intonation, and other inaccuracies by digital editing. When the recordings were played back in the two rooms, the balance and timbre of the single tracks were additionally adjusted to the room conditions.

1.2 Acoustic Scenes

The loudspeaker orchestra was recorded on the stage of the Berliner Philharmonie (BPH) and in the Berlin Open Lab (BOL). 3D models of the two scenes in SketchUp format with the absolute positions of all loudspeakers and the listener in the venues are contained in the database and described in SEC. 2.

1.2.1 BPH

The large hall of the BPH (Fig. 1) is a venue for classical concerts and the home of the Berliner Philharmoniker. The shape of the hall has originated the now-popular type of the vineyard architecture, with the stage moved a bit to the center, and terraces rising all around for the audience.

The hall has 2,250 seats and a room volume of $V_{BPH} \approx$ 21,000 m^3. A reverberation time of $T_{20,\ \mathrm{mean,\ BPH}} = 1.95$ s was measured at the listener position, averaged over the 18 loudspeakers on stage (Fig. 2). All measurements were carried out in the empty hall.

1.2.2 BOL

The Mixed Reality Design Lab in the BOL is a research and project space at the Berlin University of Arts (Universität der Künste). It is not an original performance space

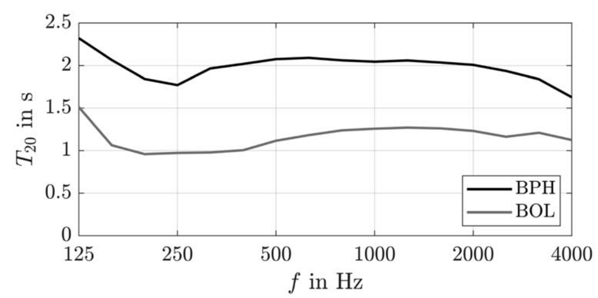


Fig. 2. Reverberation time T_{20} over frequency for the octave bands from 125 Hz to 4,000 Hz averaged over the 18 emitter positions and the listener position, measured with the NTI MA2230 system for the Berliner Philharmonie (BPH, black) and the Berlin Open Lab (BOL, gray).

for music; the size, however, is typical for a small chamber music environment, with a room volume of $V_{\mathrm{BOL}} \approx 1{,}800$ m^3 and an empty floor area of $A_{\mathrm{BOL}} = 320$ m^2. The measured reverberation time averaged over all 18 loudspeakers at the listener position is $T_{20,\ \mathrm{mean,\ BOL}} = 1.15$ s (Fig. 2).

1.3 Source Description

The loudspeakers used to represent individual musical instruments were selected on the basis of their directivity and frequency bandwidth, even though this can only be an approximation and, in particular, cannot reproduce the time-dependent changes in directivity caused by musicians' movements and the pitch-dependent change in radiation pattern [38]. The similarity of the directional characteristics of the musical instrument and loudspeaker was estimated on the basis of the directivity indices, whose values for the musical instruments were determined based on recordings with a 32-channel surrounding spherical microphone array [47].

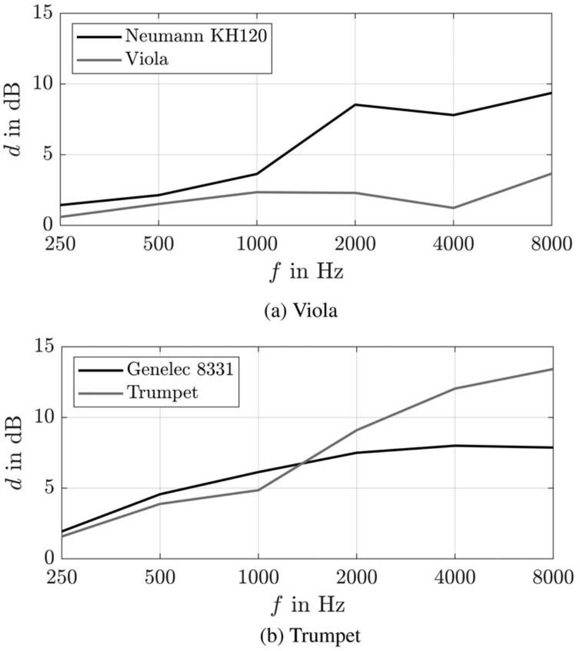


Fig. 3. Directivity indices *d* for octave bands from 200 Hz to 8,000 Hz for two musical instruments (Viola (a), Trumpet (b), gray) and the corresponding loudspeaker used (black).

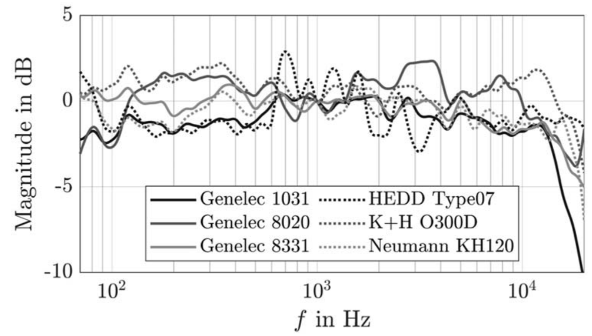


Fig. 4. On-axis free-field frequency response of the loudspeakers, measured at a distance of 4 m with the free-field equalized omnidirectional NTi M2230 measurement microphone.

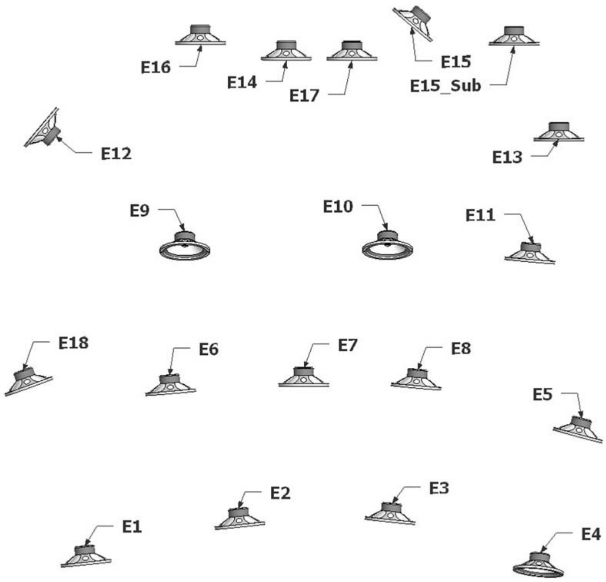


Fig. 5. Top view of the loudspeaker placement on the stage of the Berliner Philharmonie. For the assignment of the individual instruments to the respective emitters E1–E18, see Table 1.

The directivity indices for the loudspeakers were obtained from freely available data.[3] Although loudspeakers always exhibit a more or less gradual increase in directivity toward high frequencies, the behavior of musical instruments is less predictable in certain cases (Fig. 3). For example, the directivity index of the trumpet increases continuously with frequency, from 2 dB up to 14 dB at 8 kHz. Up to 2 kHz, the values are similar to that of the three-way coaxial Genelec 8331 loudspeaker. The viola's directivity index, on the other hand, fluctuates between 1 dB and 4 dB without a notable increase with frequency.

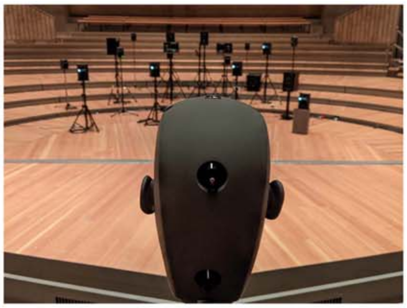
The two-way Neumann KH 120 was chosen for most of the strings and for the bassoon, the double bass and the French horn were each played out through a three-way Genelec 1031. For the trumpet, a small sound source, the three-way coaxial speaker Genelec 8331, was chosen. Spatially extended sound sources such as clarinet and bass clarinet, on the other hand, were played out via the three-way K+H O300D. For the instruments of the percussion section, such as snare drum and crash cymbals, the two-way HEDD Type 07 was used. All remaining instruments were played with a two-way Genelec 8020, with the exception of the bass drum, where a subwoofer KRK 12s was used in addition to the speaker to reproduce the low-frequency part of the instrument with sufficient sound power. The detailed assignment to the musical instruments can be found in Table 1. The on-axis free-field frequency responses of the loudspeakers used in this study are shown in Fig. 4.

The orientations of the loudspeakers were adjusted taking into account the orientation of the musical instruments to be reproduced. For example, the loudspeaker for the French horn was rotated about 110° to the right, in order to emit the sound according to the main direction of radiation of this instrument in a natural playing position. The loudspeakers for clarinet and bass clarinet, on the other hand, were turned slightly upward towards the ceiling.

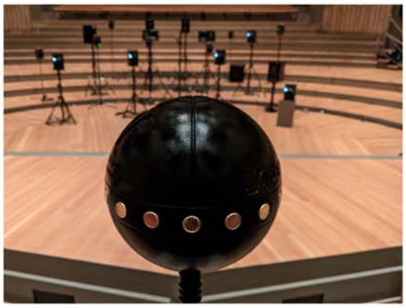
The positioning of the loudspeakers on the stage was based on the "American" seating arrangement (Fig. 5), with the 1st and 2nd violins (E1, E2) side by side in the first row on the left, viola (E3) and cello (E4) on the right, and the double bass (E5) behind them in the second row. The second and third rows contain the loudspeakers for the woodwind section (E6–E11), flanked on the left by the harp (E18) and French horn (E12) and on the right by the trumpet (E13). The loudspeakers for the percussion instruments (E14–E17) are located in the back row. The height of the loudspeakers above the floor was adjusted according to the estimated main radiation area of the associated acoustic instruments.

[3] cf. www.clfgroup.org, accessed June 2, 2022.

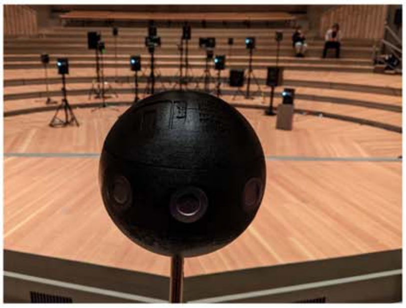
(a) Hat-and-torso simulator FABIAN (L1)



(b) Neumann KU 100 (L2)



(c) 16 Ch. Equatorial mic array (L3)

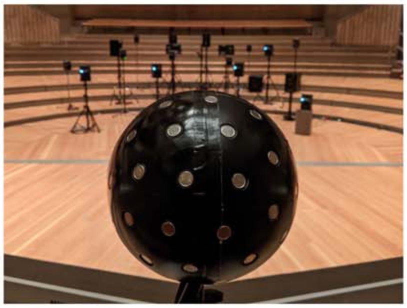


(d) 8 Ch. Equatorial mic array (L4)

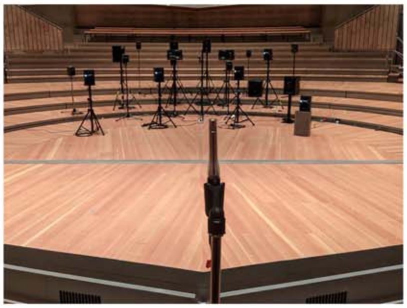
(e) Sennheiser AMBEO VR Mic (L5)

(f) Zylia ZM-1 (L6)

(g) mhacoustics em32 Eigenmike (L7)



(h) HØSMA-7N MKII (L8)



(i) NTI MA2230 (L9)

Fig. 6. The nine recording devices (”listeners”, L1–L9) in the large hall of the Berliner Philharmonie.

The positioning of the loudspeakers relative to the listener position is described in the metadata of the corresponding SOFA files (cf. Sec. 2).

1.4 Listener Description

Three kinds of microphones were selected to generate the data set: First, dummy heads were included to allow a comparison of SMA-based binaural rendering against a true binaural reference. Second, different array types (open, rigid, full-spherical, equatorial) were used to provide recordings that can be processed by a wide range of rendering approaches, and third, different commercially available arrays were included for being able to compare identical rendering methods using different recording devices. The microphone arrays are presented in the following. The exact positions of the capsules in the arrays, including their channel assignment to the recordings, is also described in the metadata of the associated SOFA files (cf. Sec. 2).

1.4.1 FABIAN (L1)

The first dummy head used for the recordings was the custom-built FABIAN head-and-torso simulator that is equipped with a high-precision neck joint to allow automated measurements for multiple head-above-torso orientations [48]. FABIAN's head and pinnae are casts of a human, while the torso and position of the head with respect to the torso were defined according to anthropometric data averaged across age and gender. FABIAN is equipped with two DPA 4060 miniature condenser microphones at the position of the blocked ear channel entrances. It was used to measure binaural room impulse responses (BRIRs) for head rotation to the left and right in the range of $\pm 45^\circ$ with a resolution of 1°.

1.4.2 Neumann KU100 (L2)

The second dummy head used was a KU100 from Georg Neumann GmbH [49] (cf. Fig. 6b). In comparison to the FABIAN dummy head, the KU100 has no torso with shoulders; hence, torso reflections do not occur. It is the third generation of this device and has two KK 83 omnidirectional microphone capsules in 4-mm-long ear canals. The frequency response of the KU 100 is diffuse-field equalized.

1.4.3 16-Channel Equatorial Microphone Array (L3)

The 16-channel MTB microphone used was developed and constructed at the TU Berlin. It is a circular microphone array on a rigid sphere whose diameter of $d = 17.5$ cm was determined by a least mean square fit between measured interaural time differences (ITDs) and spherical head ITDs calculated with the Woodworth formula [50]. The 16 hand-matched omnidirectional Sennheiser KE14 electret condenser capsules are mounted flush with the surface on the equator of the sphere. The capsules have a largely flat frequency response with a tolerance of ± 3 dB up to 16 kHz and maximum differences between capsules smaller than 2 dB. For a detailed description of the design of the microphone and its technical and perceptual evaluation, see [6].

During the recording in the BPH, an error occurred in channel 8 due to interspersed interference. This channel can therefore not be used. However, because it could be shown that there is no significant difference in the perceived reproduction quality between a 16- and an eight-channel MTB recording [6], the eight odd channels from the 16-channel recording can still be used for a valid evaluation. The MTB recording in the BOL is not affected by this error and can be used without restriction.

1.4.4 Eight-Channel Equatorial Microphone Array (L4)

Because the differences between an 8- and 16-channel MTB reproduction were shown to be small [6], a new eight-channel equatorial microphone array was designed and constructed at the TU Berlin (cf. Fig. 6d), using Neumann KM 130 modular small-diaphragm microphones and designed especially for live recording due to its superior frequency response and lower self-noise. The KM 130 combines the KM 100 output stage with the diffuse equalized omni capsule AK 30. The flush-mounted capsules are held in place by a Plastazote LD45 foam ring. The capsules are connected to the Neumann KM100 output stage placed outside the array with a 5 m LC3-KA cable. The diameter of the sphere is $d = 17.3$ cm. Even without additional filtering, the diffuse-field frequency response of the capsules when mounted in the rigid sphere is already flat with a tolerance of ± 1 dB up to about 10 kHz. In contrast, the diffuse-field frequency response of the 16-channel equatorial microphone L3 is constant only up to 1 kHz.

1.4.5 Sennheiser AMBEO VR Mic (L5)

This commercially available first-order ambisonics (FOA) microphone is built in 4-channel A-format and consists of 4 tetrahedrally arranged Sennheiser KE 14 cardioid electret condenser capsules [51]. The sound fields recorded with this type of microphone can be decomposed to first-order SH components. With the post-processing provided by the AMBEO B-format converter,[4] a largely flat diffuse field response is achieved up to around 10 kHz.

[4] https://en-us.sennheiser.com/ambeo-abconverter, accessed: 02 June 2022.

1.4.6 Zylia ZM-1 (L6)

This commercially available microphone contains 19 omnidirectional digital microelectromechanical systems (MEMS) capsules that are placed in a rigid sphere with a diameter of $d = 9.8$ cm. The higher-order ambisonics (HOA) microphone array can decompose the recorded sound field up to the third SH order. The connection to a measuring computer is realized via USB using an ASIO driver.

1.4.7 mhacoustics em32 Eigenmike (L7)

The commercially available em32 Eigenmike [52] is a 32-channel fourth-order HOA microphone. Also here, the omnidirectional electret microphone capsules are located on a rigid spherical surface with a diameter of $d = 8.4$ cm. A 24-bit A/D conversion is performed within the microphone array. A CAT-5 cable carries the digitized multiplexed audio signals from the microphone to the proprietary Eigenmike Microphone Interface Box (EMIB), where they are converted to a Firewire stream. An audio workstation connected to it recognizes the em32 as a 32-channel ASIO audio device.

1.4.8 HØSMA-7N MKII (L8)

The HØSMA-7N MKII was developed and constructed at TH Köln. It is a 64-channel rigid sphere microphone array allowing to decompose the recorded sound field up to the seventh SH order. The rigid sphere of the array was 3D printed using the fused filament fabrication technology and has a diameter of $d = 23.5$ cm. 64 omnidirectional Sennheiser KE14 electret condenser capsules are flush-mounted with the surface, arranged according to a 7th order Fliege sampling grid (cf. Figure 6h). Each capsule is operated by a corresponding converter circuit board from Sennheiser, providing a voltage divider and a converter to output a balanced signal on a standard XLR-pin connector, all fit inside a shielded microphone tube. The microphones are connected to an XLR-socket, which is soldered to the multicore cables that connect the array to the periphery [53].

1.4.9 NTI MA2230 (L9)

In addition to the multichannel microphone array recordings, the sound field was also captured using the class 1 certified free-field equalized omnidirectional measuring microphone NTi M2230, consisting of a combination of the MA220 preamplifier and the MC230 microphone capsule.

1.5 Recording Setup

The central hub of the measurement setup was the RME MADIfaceXT audio interface, connected to the measurement computer (Windows 10, Intel Core i7-4790K CPU at 4.00 GHz and 32 GB RAM) via USB 3.0 and providing 2 x MADI optical inputs/outputs (I/Os). The sample rate for all measurements was set to $f_s = 48$ kHz. The modular 32-channel PRODIGY.MC from DirectOut with integrated A/D converter was used as microphone preamplifier for both dummy heads (L1, L2), the equatorial microphone arrays (L3, L4), the 4-channel AMBEO

VR Mic (L5), and the omnidirectional measurement microphone (L9). In this device, the microphone preamplification has been set to +25 dB for all connected microphones, and the connection to the audio interface was established via a MADI fiber cable. Because of the high number of channels, the 64 capsules of the HØSMA array had to be distributed to eight PreSonus DigiMax DP88 microphone preamps.

After a synchronized A/D conversion into ADAT format, the signals were connected via eight TOSLink optical fiber cables (with eight channels each) to an RME ADI648 for further conversion into MADI format, which finally enabled the transmission of all 64 channels via one MADI fiber cable to the audio interface. The Zylia ZM-1 (L6) and the Eigenmike em32 (L7) were connected directly to the measuring computer via USB and Firewire, respectively.

The 18 channels for the loudspeakers were sent from the audio interface via a second MADI fiber cable to another RME ADI-648 format converter, which was placed on the stage side. There, the signals were converted to ADAT and routed via TOSLink optical fiber cables to two DigiMax DP88 and to one ADCON Marian D/A converter. The analog inputs of the first 16 loudspeakers (E1–E16) were connected to the eight-channel D/A converter outputs of the two Digimax, loudspeakers for cymbals and harp (E17, E18) were operated via the first two channels of the Marian Converter. To be able to run synchronized I/Os of two ASIO devices on one Windows based computer, we used Audiants's DANTE VIA software and a RME HDSPe MADI PCIe card for the measurement setup of the two microphones L6 and L7.

The synchronous playback of the anechoic audio tracks and the recording of the reproduced sound field of the loudspeaker orchestra was performed using *AKTools* [54], an open Matlab toolbox for signal acquisition, processing, and inspection in acoustics. The sweep-based impulse response measurements were performed using the same toolbox, with a sweep of order 18 (2^{18} samples $\approx$5.46 s at 48 kHz). Prior to adjusting the balance and tone color of the reproduced anechoic audio signals to the room acoustic environment, all loudspeakers were calibrated to a sound pressure level of $L_{A,S} = 87$ dB, measured in main radiation direction at a distance of 1 m with the NTi XL2 system (M2230).

To achieve an artistically satisfying sound balance at the recording location—especially for the live recording of the entire orchestra—the position of the microphone as well as the exact positions of the sources was adjusted by a professional balance engineer, as well as the timbre and loudness balance of the loudspeaker orchestra using the audio workstation REAPER and its parametric equalizers. After these adjustments, the average distance between the listener and the emitters was 8.9 m (min: 6.6 m, max: 10.8 m) in the BPH and 4.1 m (min: 2.6 m, max: 5.6 m) in the BOL. The selected recording position is for most loudspeakers slightly outside the critical distance and corresponds in the BPH to the position of the main microphone permanently used, for example, for the recordings of the Digital Concert Hall streaming platform by the in-house sound engineers.

The listener position was marked using two self-leveling cross-line lasers (Bosch Quigo, precision $\pm$0.8 mm/m) installed at the same height, with one laser indicating the frontal direction and the second laser mounted with a 90° offset. The orientation of the listeners could be determined with the two lasers based on external markings on the shells of the arrays. The position of the emitters was measured with the VariSphear [29], a fully automated robotic scanning system with a laser range finder attached. The filter and balance settings for the room adjustments were directly applied to the anechoic audio signals. They are thus contained in all recordings and also in all impulse response (IR) based auralizations in which the IRs are convolved with the anechoic audio signals.

In post-processing, the captured system latency was removed from the recordings and the measured IRs, and for optimized storage, the noise component at the end of the IRs was truncated. To remove low-frequency noise from the recordings and the IRs, presumably emitted by the room air conditioning system, a 16th-order Butterworth high-pass filter was applied at 25 Hz. Noise components on channel 8 of the eight-channel equatorial microphone array (L4) in the BPH recording caused by interference were removed using the noise reduction algorithm of the iZotope RX 10 software. In addition, the frequency responses of the two DPA 4060 miniature condenser microphones were compensated in the FABIAN (L1) recordings and BRIRs. Apart from that, no further post-processing was done on the recordings or the IRs.

2 DATABASE

An excerpt of the anechoic audio signals, the recordings of the loudspeaker orchestra, and the measured IRs are provided and are freely accessible on DepositOnce [44] under a Creative Commons share alike license (CC-BY-SA 4.0). Additionally, 3D room models and pictures of the measurement setup are provided. For a detailed description of the structure and the data format, please refer to the enclosed documentation.

The anechoic audio signals are provided as wav files, whereas the recordings and IRs are stored in the SOFA format (version 2.1). The geometry of the acoustic scenes are provided as SketchUp files, which contain the geometry of the room and the position and orientation of all objects. Additionally, overview and detail photos of the measurement setup are included in the database. The anechoic audio signals are required for an evaluation of IR-based binaural reproduction techniques. They are available separately for each room, reflecting the room specific timbre and balance adjustments made by the sound engineers.

The SOFA files can be read with a variety of APIs.[5] They contain the recorded signals and IRs along with metadata describing the data in detail. This includes the arrangement of the capsules of the microphone arrays and the position of the listeners and loudspeakers in the room.

[5] cf. www.sofaconventions.org/mediawiki/index.php/Software_and_APIs, accessed: 02 June 2022.

For the (multichannel) recordings of the loudspeaker orchestra, we chose the standardized `SingleRoomSRIR` convention. Even though this data type was originally developed for FIRs, longer time signals, such as audio recordings, can also be stored and read without restrictions. For each room and listener, there is a separate SOFA file in the database. The naming of the data follows the scheme `1_RoomName_ListenerNr_recordings.sofa`, and the recordings are stored in the Data.IR field. The field has the dimension `M×R×N`, where M (measurement) is always 1, R (receiver) corresponds to the number of capsules of the associated microphone array, and N indicates the length of the recording in samples. Because the anechoic audio signals were played back at the same time from the 18 loudspeakers, the 18 emitters are considered as one source in this convention. However, the relative position of the emitters to the source position, which was defined as the center of the loudspeaker orchestra (cf. Fig. 5), can be retrieved from the related metadata `EmitterPosition`.

The measured impulse responses, on the other hand, are stored in the `SingleRoomMIMOSRIR` convention. There is also a separate SOFA file for each room and listener, and the naming of the file corresponds to the scheme `2_RoomName_ListenerNr_RIR.sofa`. The IRs are stored in the field Data.IR with the dimension `M×R×N×E`, where M is in most cases 1, except for the listener L1 (FABIAN), where M denotes the head-above-torso orientation between -45° to 45° in 1° steps, which corresponds to $\mathrm{M} = 91$ measurements. Furthermore, the dimension R corresponds to the number of capsules of the microphone array or dummy head, N to the length of the IRs, and E (emitter) to the number of loudspeakers, which is always 18. The musical instruments assigned to the emitters can be determined from the metadata of the file. Furthermore, the metadata contain additional information about the room and the measurement conditions.

3 CONCLUSION

The presented data set includes recordings of an 18-channel loudspeaker orchestra in two different acoustic environments. As audio content, an arrangement for an 18-piece chamber orchestra, recorded in an anechoic chamber, was used. The loudspeaker orchestra was recorded by eight different multichannel microphone types, including binaural and pseudo-binaural receivers and different spherical microphone arrays. In addition, sweep-based impulse responses were recorded for each sound emitter and microphone. For the head-and-torso simulator (L1), BRIRs were measured for head rotations in an azimuthal range of $\pm 45^\circ$ and with a resolution of 1°. The use of a complex yet reproducible sound source has made it possible to place the different recording systems at the same point in the sound field without mutual interference.

These recordings can be used for comparative evaluations of different general approaches, different microphone designs, and different signal processing and rendering techniques for the immersive streaming of music. The anechoic audio signals, loudspeaker orchestra recordings, and measured IRs, along with 3D room models and a precise documentation of the data set are published with open access [44]. Because the loudspeaker orchestra in the two rooms used can be easily recovered, it is planned to add more recordings to the measurement series, which may include other array configurations as well as new recording devices, and make them available as a new version of the published data set.

4 ACKNOWLEDGMENT

We would like to thank the musicians of the Konzerthausorchester Berlin for the anechoic recording of the orchestral arrangement, Erik Brauer for sound engineering and recording supervision, and Nikolaos Alexandris, Jens Ahrens, Franz Zotter, and Georg Neumann GmbH for providing additional recording equipment, and Henning Riez for his assistance with the audio post-production. We would like to thank the Digital Concert Hall and the sound department of the Philharmonie Berlin for their conceptual and on-site support.

5 REFERENCES

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

[2] K. I. McAnally and R. L. Martin, "Sound Localization With Head Movement: Implications for 3-D Audio Displays," *Front. Neurosci.*, vol. 8, paper 210 (2014 Aug.). https://doi.org/10.3389/fnins.2014.00210.

[3] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. G. Katz, and C. de Boishéraud, "Influence of Head Tracking on the Externalization of Speech Stimuli for Non-Individualized Binaural Synthesis," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 2011–2023 (2017 Mar.). https://doi.org/10.1121/1.4978612.

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

[5] 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.).

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

[7] V. Pulkki, A. Politis, M.-V. Laitinen, J. Vilkamo, and J. Ahonen, "*First-Order Directional Audio Coding (DirAC)*," in V. Pulkki, S. Delikaris-Manias, and A. Politis, eds., *Parametric Time-Frequency Domain Spatial Audio*, pp. 89–140 (IEEE Press, Piscataway, NJ, 2018), 1st ed. https://doi.org/10.1002/9781119252634.ch5.

[8] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516 (2007 Jun.).

[9] 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 Mar.). https://doi.org/10.1109/JSTSP.2015.2415762.

[10] 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).

[11] S. V. Amengual Garí, J. M. Arend, P. T. 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.

[12] S. Berge and N. Barrett, "High Angular Resolution Planewave Expansion," in *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics* (Paris, France) (2010 May).

[13] N. Barrett and S. Berge, "A New Method for B-Format to Binaural Transcoding," in *Proceedings of the AES 40th International Conference: Spatial Audio: Sense the Sound of Space* (2010 Oct.), paper 6-5.

[14] O. Thiergart, M. Taseska, and E. A. P. Habets, "An Informed Parametric Spatial Filter Based on Instantaneous Direction-of-Arrival Estimates," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 12, pp. 2182–2196 (2014 Oct.). https://doi.org/10.1109/TASLP.2014.2363407.

[15] A. Politis, S. Tervo, and V. Pulkki, "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6802–6806 (Calgary, Canada) (2018 Sep.). https://doi.org/10.1109/ICASSP.2018.8462608.

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

[17] C. Schörkhuber and R. Höldrich, "Linearly and Quadratically Constrained Least-Squares Decoder for Signal-Dependent Binaural Rendering of Ambisonic Signals," in *Proceedings of the AES International Conference on Immersive and Interactive Audio* (2019 Mar.), paper 22.

[18] F. Zotter and M. Frank, "*Compact Spherical Loudspeaker Arrays*," in *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer Topics in Signal Processing, pp. 153–170 (Springer International Publishing, Cham, Switzerland, 2019). https://doi.org/10.1007/978-3-030-17207-7_7.

[19] B. Rafaely, "Plane-Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157 (2004 Oct.). https://doi.org/10.1121/1.1792643.

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

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

[22] F. Zotter and M. Frank, "*Higher-Order Ambisonic Microphones and the Wave Equation (Linear, Lossless)*," in *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer Topics in Signal Processing, pp. 131–152 (Springer International Publishing, Cham, Switzerland, 2019). https://doi.org/10.1007/978-3-030-17207-7_6.

[23] L. McCormack and A. Politis, "SPARTA & COMPASS: Real-Time Implementations of Linear and Parametric Spatial Audio Reproduction and Processing Methods," in *Proceedings of the AES International Conference on Immersive and Interactive Audio* (2019 Mar.), paper 111.

[24] H. Helmholz, C. Andersson, and J. Ahrens, "Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals," in *Proceedings of the Fortschritte der Akustik, DAGA*, pp. 1462–1465 (Rostock, Germany) (2019 Mar.).

[25] H. Helmholz, T. Lübeck, J. Ahrens, S. V. A. Garí, D. L. Alon, and R. Mehra, "Updates on the Real-Time Spherical Array Renderer (ReTiSAR)," in *Proceedings of the Fortschritte der Akustik, DAGA*, pp. 1169–1172 (Hannover, Germany) (2020 Mar.).

[26] I. Engel, D. F. M. Goodman, and L. Picinali, "Assessing HRTF Preprocessing Methods for Ambisonics Rendering through Perceptual Models," *Acta Acust.*, vol. 6, paper 4 (2022 Jan.). https://doi.org/10.1051/aacus/2021055.

[27] J. Nowak, K.-P. Jurgeit, and J. Liebetrau, "Assessment of Spherical Microphone Array Auralizations using Open-Profiling of Quality (OPQ)," in *Proceedings of the Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (2016 Jun.). https://doi.org/10.1109/QoMEX.2016.7498928.

[28] J. Nowak and S. Klockgether, "Perception and Prediction of Apparent Source Width and Listener Envelopment in Binaural Spherical Microphone Array Auralizations," *J. Acoust. Soc. Am.*, vol. 142, no. 3, pp. 1634–1645 (2017 Sep.). https://doi.org/10.1121/1.5003917.

[29] B. Bernschütz, *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Ph.D. thesis, TU Berlin, Berlin, Germany (2016 Feb.).

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

[31] B. Bernschütz, "Bandwidth Extension for Microphone Arrays," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), paper 8751.

[32] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Proceedings of the Fortschritte der Akustik, DAGA*, pp. 339–342 (Munich, Germany) (2018 Mar.).

[33] 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.).

[34] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. J. Tashev, "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 261–265 (Brighton, UK) (2019 Apr.). https://doi.org/10.1109/ICASSP.2019.8683751.

[35] 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.

[36] 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 *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx)*, pp. 250–257 (Vienna, Austria) (2020 Sep.).

[37] H. Helmholz, D. L. Alon, S. V. A. Garí, and J. Ahrens, "Effects of Additive Noise in Binaural Rendering of Spherical Microphone Array Signals," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 3642–3653 (2021 Nov.).

[38] D. Ackermann, C. Böhm, F. Brinkmann, S. Weinzierl and, "The Acoustical Effect of Musicians' Movements During Musical Performances," *Acta Acust. united Acust.*, vol. 105, no. 2, pp. 356–367 (2019 Mar.). https://doi.org/10.3813/AAA.919319.

[39] J. Pätynen, S. Tervo, and T. Lokki, "A Loudspeaker Orchestra for Concert Hall Studies," in *Proceedings of the Seventh International Conference on Auditorium Acoustics*, vol. 30, pp. 45–52 (Oslo, Norway) (2008 Jan.).

[40] J. Pätynen and T. Lokki, "Evaluation of Concert Hall Auralization with Virtual Symphony Orchestra," *Build. Acoust.*, vol. 18, nos. 3–4, pp. 349–366 (2011 Dec.). https://doi.org/10.1260/1351-010X.18.3-4.349.

[41] D. Satongar, C. Pike, and Y. W. Lam, "SBS-BRIR," University of Salford, V1 (2020 Feb.). https://doi.org/10.17866/RD.SALFORD.C.4851498.V1.

[42] P. Stade, B. Bernschütz, and M. Rühl, "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," Zenodo, V3 (2020 May). https://doi.org/10.5281/ZENODO.3930833.

[43] T. McKenzie, L. McCormack, and C. Hold, "Dataset of Spatial Room Impulse Responses in a Variable Acoustics Room for Six Degrees-of-Freedom Rendering and Analysis," Zenodo, V1.1 (2021 Nov.). https://doi.org/10.5281/ZENODO.6382405.

[44] D. Ackermann, J. Domann, F. Brinkmann, J. M. Arend, S. Weinzierl and, "Database: Recordings of a Loudspeaker Orchestra with Multi-Channel Microphone Arrays for the Evaluation of Spatial Audio Method," DepositeOnce: Technische Universität Berlin, V1 (2022). https://doi.org/10.14279/depositonce-15774.

[45] AES, "AES Standard for File Exchange - Spatial Acoustic Data File Format," *Standard 69-2022* (2022 Sep.).

[46] DIN, "Ergonomics - Human Body Dimensions - Part 2: Values," *Standard 33402-2* (2020 Dec.).

[47] S. Weinzierl, M. Vorländer, G. Behler, et al., "A Database of Anechoic Microphone Array Measurements of Musical Instruments," DepositeOnce: Technische Universität Berlin, V2 (2017 Apr.). https://doi.org/10.14279/depositonce-5861.2.

[48] A. Lindau, T. Hohn, and S. Weinzierl, "Binaural Resynthesis for Comparative Studies of Acoustical Environments," presented at the *122nd Convention of the Audio Engineering Society* (2007 May), paper 7032.

[49] M. Schneider, "The Development of Dummy Heads since 1970," presented at the *150th Convention of the Audio Engineering Society* (2021 May), paper 10500.

[50] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model from Anthropometry," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479 (2001 Mar.).

[51] M. Schneider, "Reale Ambisonics-Mikrophone," in *Proceedings of the Fortschritte der Akustik - DAGA*, pp. 169–172 (Munich, Germany) (2018 Mar.).

[52] G. Elko, R. Kubli, J. Meyer and, "Audio System Based on at Least Second-Order Eigenbeams," International Patent WO2003061336A1 (2003 Jul.).

[53] O. Moschner, D. T. Dziwis, T. Lübeck, and C. Pörschmann, "Development of an Open Source Customizable High Order Rigid Sphere Microphone Array," presented at the *148th Convention of the Audio Engineering Society* (2020 May), eBrief 583.

[54] F. Brinkmann and S. Weinzierl, "AKtools — An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), eBrief 309.

THE AUTHORS


David Ackermann


Julian Domann


Fabian Brinkmann


Johannes M. Arend


Martin Schneider


Christoph Pörschmann


Stefan Weinzierl

David Ackermann has received his M.Sc. degree in Audiocommunication and Technology from TU Berlin, Germany, in 2015. He has been a research associate in the Audio Communication Group at the TU Berlin since 2015, where his research is focused in the field of virtual and musical acoustics. His research includes the investigation of the time-dependent behavior of natural acoustic sound sources and their auralization in virtual acoustic environments.

•

Julian Domann received his M.Sc. degree in Audiocommunication and Technology from TU Berlin, Germany, in 2022. In 2019, he joined an engineering consultancy, working on projects including room acoustics and studio technology.

•

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

•

Johannes M. Arend received a B.Eng. degree in media technology from HS Düsseldorf, Düsseldorf, Germany, in 2011, and an M.Sc. degree in media technology from TH Köln, Cologne, Germany, in 2014. From 2015 to 2022, he was a research fellow at TH Köln and a doctoral student at TU Berlin, Berlin, Germany, receiving his Ph.D. degree (Dr. rer. nat.) from TU Berlin in 2022. Since 2022, he has been a postdoctoral researcher at TU Berlin. His research focuses on binaural technology, auditory perception, and evaluation of spatial audio.

•

Martin Schneider joined the microphone development department of Georg Neumann GmbH, Berlin in 1992, after obtaining his diploma in electrical engineering from Technical University Berlin. Since 2008, he has lectured on Electroacoustics at Hochschule für Musik, Detmold, at the Tonmeister Institute, and the Microphone Lab at Technical University Berlin, at the Institute for Audio Communication.

•

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 a 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 the 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 TU Berlin and has coordinated international research consortia in the field of virtual acoustics (SEACEN) and music information retrieval (ABC_DJ).