

Sound field synthesis for psychoacoustic research: *In situ* evaluation of auralized sound pressure level^{a)}

Matthieu Kuntz,^{b)}  Norbert F. Bischof,  and Bernhard U. Seeber 

Audio Information Processing, Technical University of Munich, 80333 Munich, Germany

ABSTRACT:

The use of virtual acoustic environments has become a key element in psychoacoustic and audiological research, as loudspeaker-based reproduction offers many advantages over headphones. However, sound field synthesis methods have mostly been evaluated numerically or perceptually in the center, yielding little insight into the achievable accuracy of the reproduced sound field over a wider reproduction area with loudspeakers in a physical, laboratory-standard reproduction setup. Deviations from the ideal free-field and point-source concepts, such as non-ideal frequency response, non-omnidirectional directivity, acoustic reflections, and diffraction on the necessary hardware, impact the generated sound field. We evaluate reproduction accuracy in a 61-loudspeaker setup, the Simulated Open Field Environment, installed in an anechoic chamber. A first measurement following the ISO 8253-2:2009 standard for free-field audiology shows that the required accuracy is reached with critical-band-wide noise. A second measurement characterizes the sound pressure reproduced with the higher-order Ambisonics *basic* decoder, with and without $max r_E$ weighting, vector base amplitude panning, and nearest loudspeaker mapping on a 187 cm × 187 cm reproduction area. We show that the sweet-spot size observed in measured sound fields follows the rule $kr \leq N/2$ rather than $kr \leq N$ but is still large enough to avoid compromising psychoacoustic experiments.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0021066>

(Received 30 April 2023; revised 17 August 2023; accepted 25 August 2023; published online 26 September 2023)

[Editor: Ning Xiang]

Pages: 1882–1895

I. INTRODUCTION

Virtual acoustic environments (VAEs) paired with sound field synthesis (SFS) methods permit the controlled auralization of a wide range of sound scenes in the free field for hearing research and beyond, from single sources often used in psychoacoustic and audiological testing to complex environments with multiple static and moving sources including reverberation ([van de Par et al., 2022](#)).

The loudspeaker-based sound reproduction allows for seamless integration of participant movement and head rotations with high fidelity, as it removes the need for measured individualized head-related transfer functions (HRTFs) and head tracking necessary for real-time headphone-based auralization ([Seeber et al., 2010](#)). Headphone-free testing is also useful for hearing research with young children ([Litovsky, 2005](#); [McCartney, 2013](#)) and for hearing aid or cochlear implant users who cannot wear headphones ([Kerber and Seeber, 2013b](#)). With free-field audiometry, i.e., realizing audiometric measurements via loudspeakers in the free field instead of with headphones, the benefit of hearing aids can be measured directly on the listener ([Shulberg, 1980](#)). The requirements to carry out those measurements are described in the [ISO 8253-2:2009 \(2009\)](#) standard.

The ability to position a sound source between two loudspeakers permits the reproduction of moving stimuli, i.e., dynamic auditory scenes, and increases the range of test positions in localization experiments, both for static and dynamic acoustic scenes ([Kolotzek and Seeber, 2020](#)), and thus offers a more natural experience for the reproduction of sound sources and their room reverberation ([Seeber et al., 2010](#)). In normal hearing listeners, VAEs can be used to study the effect of reflections on speech intelligibility ([Ahrens et al., 2019](#)), on binaural unmasking ([Bischof et al., 2023](#)), and on sound localization and the precedence effect ([Buchholz and Best, 2020](#); [Seeber and Clapp, 2020](#)). In hearing impaired listeners, VAEs can be used to assess speech intelligibility, sound localization, and the effects of sound reflections ([Kerber and Seeber, 2012, 2013a](#)) or to compare hearing aid algorithms in complex but controlled and reproducible acoustic environments, as opposed to the highly variable environments patients encounter in their daily lives ([Gomez, 2019](#); [Hendrikse et al., 2022](#)).

Beyond the acoustic advantages of loudspeaker-based reproduction, allowing participants to move freely also leads to ecological experimental designs to study human behavior in natural conversations or social situations subject to different acoustics ([Hadley et al., 2019](#)).

The auralization of rooms and buildings is often used in architectural acoustics because it enables the back-to-back comparison of the acoustics of different buildings, like concert halls ([Schroeder, 1973](#)) or classrooms ([Salanger et al.,](#)

^{a)}This paper is part of the special issue on 3D Sound Reconstruction for Virtual Auditory Displays: Applications in Buildings.

^{b)}Email: matthieu.kuntz@tum.de

2020; Pulella *et al.*, 2021). That way, investigating the influence of acoustic measures on perceptual quantities via listening experiments becomes feasible. Coupled with acoustic room simulations, VAEs can be used for virtual showcasing and planning of future rooms and buildings, integrating perceptual studies in the architectural design process (Vorländer *et al.*, 2015; Thery *et al.*, 2019).

Sound fields reproduced with SFS methods were mostly simulated and evaluated for simple sound scenes, such as those generated by a plane wave or a point source, which is why we also focus on this scenario. Simulated evaluations of SFS are quite common and were carried out multiple times by different research groups. For example, Daniel (2003) worked on the distance coding of sources auralized with higher-order Ambisonics (HOA) and introduced near-field compensation filters (NFC-HOA), which he evaluated for a single point source and a plane wave. Otani and Shigetani (2019) simulated the reproduction accuracy of the HOA $\max r_E^1$ and least norm decoders for a plane wave. These numerical simulations aimed at determining the accuracy of the different sound field reproduction methods. Other projects also ran simulations of existing or planned loudspeaker setups to assess their performance (e.g., Favrot, 2010; Grandjean, 2021). However, the idealized conditions of these simulations make it difficult to draw conclusions concerning the accuracy of these methods in experimental setups, where loudspeaker directivity, frequency response, and reflections alter the reproduced sound field.

Grimm *et al.* (2015) investigated the suitability of sound field reproduction methods for assessing hearing aid algorithms. They used head-related impulse responses (HRIRs) and hearing aid microphone HRIRs measured with an artificial head to simulate ear and microphone signals. They compared the hearing aid beam patterns and signal-to-noise ratio (SNR) obtained with HOA, vector base amplitude panning (VBAP), and nearest loudspeaker mapping (NLS) for an increasing number of loudspeakers. All three methods were deemed suitable to evaluate hearing aid performance, but different algorithms and metrics required different optimal solutions.

Frank *et al.* (2008) and Stitt *et al.* (2013, 2014) studied localization ability in the loudspeaker array center and for off-center positions. Frank *et al.* (2008) found similar performance for the HOA *basic* and HOA $\max r_E$ methods in the center and slightly better results for the HOA $\max r_E$ method at the off-center position. Stitt *et al.* (2013, 2014) verified that increasing the order of the Ambisonics reproduction method improved localization.

However, the physical accuracy of SFS achieved in real hardware setups has not been investigated much. It is unclear how the theoretical limits of SFS translate to hardware setups.

Issues such as loudspeaker frequency response, directivity, and sound reflections in the reproduction setup are usually not considered in simulations and in the formulas deriving the driving functions for the SFS. Ahrens and Spors (2009) presented an analytical derivation of SFS for

non-omnidirectional loudspeakers, which, however, does not consider variations across loudspeakers. The high number of loudspeakers required for a sufficient spatial resolution and reproduction area size also introduces unwanted reflections in otherwise anechoic environments, resulting in deviations from numerical evaluations of a given reproduction setup. A similar problem applies to visual projection systems: To free the participants from wearing head-mounted displays, screens and projectors can be added to the VAE, enabling less restraining visual display and opening the door to allow multiple participants to interact within the same VAE (Hládek and Seeber, 2023) but introducing additional hardware that might compromise the acoustic reproduction inside the VAE.

The authors only found three studies presenting measured SFS accuracy over a wider area, and only one of those presented numerical results. Weißgerber (2019) assessed the reproduction accuracy of a 128-channel wave-field synthesis setup for clinical research on cochlear implant wearers. They used a 30-channel microphone array to measure the sound pressure resulting from the reproduction of a 500 Hz tone from a focused virtual source inside the room. However, that work did not contain any numerical assessment of the measured sound pressure beyond a figure of the measured sound field of a 500 Hz tone. Grandjean (2021) evaluated a hybrid 2.5- and three-dimensional (2.5D and 3D) Ambisonics reproduction method in a 50-channel spherical loudspeaker array with a 43.75 cm \times 50 cm, 72-channel microphone grid for a 2 kHz tone and used the measurement data to extrapolate the recorded sound field to a 1 m \times 1 m grid. It is mentioned that errors in the loudspeaker positioning, microphone positioning, and test signal amplitude do not allow for a quantitative study or acceptable reproduction errors (Grandjean, 2021). Murillo *et al.* (2014) measured the reproduction accuracy of fifth-order 3D HOA with the *basic*, $\max r_E$, and *in phase* decoders for sine tones at 250 Hz, 1 kHz, and 2 kHz using a translating 29-channel linear microphone array to cover a zone of about 1.4 m \times 1.8 m in a 40-channel spherical loudspeaker array and report sound pressure amplitude, phase, and acoustic intensity direction errors. They show sound pressure level (SPL) errors of up to 5 dB inside the theoretical sweet-spot according to the $N = \lceil kr \rceil$ rule of thumb for free-field conditions from Ward and Abhayapala (2001), which gives a limit radius $r = N/k$. This rule of thumb was derived by computing the error in the spherical harmonics expansion of a plane wave for different orders of truncation. For $N = kr$, the relative truncation error is 4%, i.e., -14 dB on the surface of the sphere of radius r . While we chose to define the sweet-spot size by a different metric, this rule of thumb has been used widely to characterize the relationship between frequency and decomposition order of the sound field, which is why we also refer to it as the theoretical sweet-spot size.

In addition to our VAE consisting of a loudspeaker array in an anechoic chamber at the Technical University of Munich (Seeber and Clapp, 2017), other examples can be found at the RWTH Aachen University (Pausch *et al.*,

2020), the University of Auckland (Au *et al.*, 2021), and the Technical University of Denmark (Favrot and Buchholz, 2010).

In many recent studies, VAEs have been used as measurement devices. To ensure that the acoustic evaluations carried out using VAEs are reliable, it is crucial to evaluate the potential errors introduced by the hardware of the VAEs themselves and its impact on the measured quantity. This study investigates the strict requirements of basic audiometry measurements under free-field conditions in a loudspeaker array. Furthermore, it presents measured physical accuracy of SFS methods to establish a physically verified baseline for different SFS methods and relates them to tone-in-noise and speech intelligibility measurements.

II. METHODS

The measurements took place in the Simulated Open Field Environment (SOFE), set up in the anechoic chamber at the Technical University of Munich (Seeber and Clapp, 2017). The loudspeakers were mounted on a custom 4.8 m × 4.8 m square holding frame at a height of 1.4 m (Fig. 1). The center point of the loudspeaker array is located at a distance of 2.4 m from the front, rear, left, and right loudspeakers. Special care was taken to reduce reflections on the necessary hardware inside the anechoic chamber as much as possible. All holding frames for the loudspeakers and the projectors were wrapped in absorbing material, even below the net floor. Additional absorbers were mounted on all sides of the loudspeaker enclosures, including the solid part of their front face.

The stimuli were presented via the SOFE's 36 horizontally arranged loudspeakers (Dynaudio BM6A mkII, Dynaudio, Skanderborg, Denmark) placed in 10° spacing.

A first measurement compares the reproduced sound levels according to the requirements of ISO 8253-2:2009 (2009) for free-field audiometry. A second measurement evaluates the level errors of different auralization methods with an 8.5 cm grid across a 187 cm × 187 cm measurement area. The auralization methods studied in this work are



FIG. 1. (Color online) Loudspeaker array in a 10 m × 6 m × 4 m anechoic chamber. The loudspeakers are placed on a 4.8 m × 4.8 m frame in 10° azimuth steps and equalized in level, delay, and phase.

HOA with the *basic* and the *max r_E* decoders, VBAP, and NLS. The HOA *in phase* decoder was not measured because of previously shown high errors when reproducing a target sound field (Murillo *et al.*, 2014) and its distorted interaural cues (Kuntz and Seeber, 2020).

A. Measurement following ISO 8253-2:2009—Sound field audiometry

1. Measurement setup

The ISO 8253-2:2009 (2009) standard stipulates measuring the level of the test stimulus at the reference point (defined here as the center of the loudspeaker array) and at four side positions, ±15 cm from the reference point along the axis of the loudspeaker and perpendicularly, forming a cross of five measurement points. The 0° loudspeaker of the array was used for this measurement. A single 1/2-in. class 1 measurement microphone (MM210, Microtech Gefell, Gefell, Germany) was sequentially placed at these five positions. The integrated electronics piezo-electric (IEPE) power was supplied by the phantom power of an RME Micstasy analog-to-digital (A/D) converter (Audio AG, Haimhausen, Germany) through a 48 V-to-IEPE converter (Schalltechnik SÜD & NORD GmbH, Regensburg, Germany). The microphone signal was preamplified and digitally converted by the RME Micstasy and connected via a multichannel audio digital interface (MADI) glass fiber cable to an RME HDSPe soundcard (Audio AG) in the measurement personal computer (PC) working with a 44.1 kHz sampling frequency.

2. Stimuli

This measurement was carried out at the frequencies of 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz, commonly used in audiology. We used sine tones, frequency-modulated (FM) tones with a 4 Hz modulation frequency and a 2.5% modulation depth, and one Bark wide noise (Fastl and Zwicker, 2007) centered around the audiological frequencies. Stimuli were played back without equalization filters, with a level of 75 dB SPL at the reference point. The total signal duration was 2 s, with a rise time of 50 ms. For the level analysis, the first and last 10 000 samples (~220 ms) were truncated to remove onset effects and compensate time of arrival differences between the measurement points. The recordings were bandpassed between 100 Hz and 18 kHz using a brickwall fast Fourier transform (FFT) filter.

B. Measurement of reproduced sound fields

1. Measurement setup

To evaluate the reproduction accuracy of the different SFS methods, a custom computer-controlled movable linear microphone array with 23 1/2-in. class 1 measurement microphones (MM210, Microtech Gefell, Gefell, Germany) was used to measure the reproduced sound field within the loudspeaker array. The microphones were equally distributed on a 2 m long holding bar with an outer diameter of 2 cm.

They were mounted through holes drilled into the bar, with the diaphragm about 6.5 cm away from the holding bar. The microphones were placed in 8.5 cm steps, resulting in a measurement span of 1.87 m (Fig. 2). All microphones were connected to a PAK measurement system (PAK mkII, Müller-BBM VAS GmbH, Planegg, Germany) equipped with two ICM42 modules, which were used as a microphone preamplifier and as a power supply for the IEPE microphones. The measurement range was set to 0.1 V and an internal gain of +40 dBV. The monitoring outputs of the ICM42 modules were connected to three RME Micstasy microphone preamplifiers with A/D converters (Audio AG, Haimhausen, Germany) running at 24-bit resolution, which were connected via MADI glass fiber cables to an RME HDSPe soundcard (Audio AG, Haimhausen, Germany) in the measurement PC working with a 44.1 kHz sampling frequency.

2. Stimuli

To measure the reproduction methods *in situ* in a loudspeaker array from a psychoacoustic point of view, different test signals were used. First, sine tones at the commonly used audiological frequencies (125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz) were used as a strict and uncompromising measurement criterion and because of their role in a wealth of classical psychoacoustic experiments, e.g., detection experiments (Grantham, 1986). In addition to discrete frequencies, critical-band-wide noise was used. White noise was bandpass-filtered with a brickwall FFT filter, resulting in one Bark wide narrowband noise centered around the audiological frequencies mentioned above. The cutoff frequencies for these narrowband noise stimuli were derived by adding and subtracting 0.5 Bark from the center frequency, following the formula by Traummüller (1990).

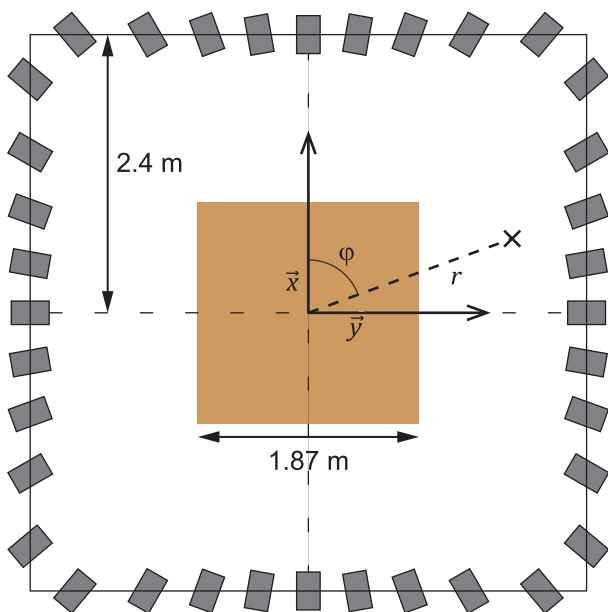


FIG. 2. (Color online) Horizontally arranged loudspeakers. The shaded area indicates the 23 × 23 grid span covered by the movable microphone array.

All signals were 732 ms long (2^{13} samples at a sampling frequency of 44.1 kHz), including 10 ms rise and fall Gaussian envelopes. The stimuli were generated to create a SPL of 70 dB SPL at the center of the loudspeaker array. The SPL was calculated on a 500 ms snippet in the middle of the 732 ms recording, removing the onset and offset of the stimuli from the analysis.

In this work, two-dimensional (2D) sound fields are considered. The sound field of a virtual point source located at a radial distance of 4 m and an azimuth angle of 13° from the center of the loudspeaker array was auralized using four different reproduction techniques, described below. That particular source position was chosen to fulfill several criteria: (1) Its azimuth angle had to lie between loudspeakers, to investigate the between-loudspeaker playback; (2) its distance had to differ from the physical distance of the loudspeakers, as this is usually the case in virtual environments [see the three environments described in van de Par *et al.* (2022) for an example]; (3) its position along the y axis should not be too far from the center to be able to show the parallax shift from the source distance mismatch.

The reference sound field used for error computation was defined as the theoretically computed sound field radiated by the same point source at 4 m distance and 13° azimuth, including the time and amplitude change with distance. The reference sound source level was set to 70 dB SPL in the center of the loudspeaker array. Time differences of arrival were compensated for to ensure the first wavefront of the reference and of the reproduced stimuli reached the center of the loudspeaker simultaneously.

3. Loudspeaker equalization

Finite-impulse response (FIR) equalization filters of length 1024 taps (time-shifted in a 2048-tap filter) were used during playback of the critical-band-wide noise stimuli to compensate for each loudspeaker’s frequency and phase response and the time difference of arrival at the center point. The equalization filters were measured with the middle microphone of the linear microphone array, positioned in the center of the loudspeaker array. Their frequency response was computed on a 23 ms windowed impulse response, short enough to discard the strongest reflections coming from the loudspeakers on the opposite side of the array.

To equalize the loudspeakers for sine tone measurement, separate equalization filters for discrete frequencies were computed by measuring the amplitude and phase of each tone in the steady state, which was compensated by scaling and resampling a sinc pulse.

C. Sound field reproduction techniques

1. NLS

For the NLS method, the source signal was played from the loudspeaker with the smallest azimuthal difference to the sound source. For the test source at 13°, this corresponds to the loudspeaker at 10° in the current sound field

reproduction environment. The resulting mapping error of 3° is slightly larger than the expected average mapping error of 2.5° for uniformly distributed sound sources with a 10° loudspeaker spacing.

2. VBAP

The VBAP method (Pulkki, 1997) uses the two loudspeakers closest to the virtual sound source (except when the virtual sound source shares its position with a loudspeaker, in which case only the relevant loudspeaker is used). The unit vectors L_1 and L_2 point toward the two loudspeakers, and the unit vector p points toward the virtual source. It can be considered as a linear combination of the loudspeaker vectors (Pulkki, 1997): $p = g_1 L_1 + g_2 L_2$, with g_1 and g_2 being the loudspeaker gains. Solving the equation gives

$$[g_1, g_2] = p^T [L_1, L_2]^{-1}. \quad (1)$$

The loudspeaker gains were then normalized according to Eq. (2) (Pulkki, 1997),

$$g_{norm} = \frac{g}{\sum_i g_i^2}. \quad (2)$$

3. HOA

In the HOA method, the sound field is decomposed into a set of circular harmonics coefficient signals, which describe the directional information in the sound field. This is typically accomplished by representing any source signal by the set of signals that would be picked up by circular harmonic microphone directivities. The Ambisonics order N refers to the order of the circular harmonics decomposition. In this work, we use an Ambisonics order of $N = 17$. For a single sound source $s(t)$, located at the azimuth angle φ_S , this is done by multiplying the encoding vector (Daniel, 2001),

$$y(\varphi_S) = \left[1, \sqrt{2}\cos(\varphi_S), \sqrt{2}\sin(\varphi_S), \dots, \sqrt{2}\cos(N\varphi_S), \sqrt{2}\sin(N\varphi_S) \right]^T, \quad (3)$$

with the source signal $s(t)$, yielding the circular harmonics signals $\chi(t) = y(\varphi_S) \times s(t)$. This format can be used to reproduce the sound field with various loudspeaker setups, which is done in a subsequent decoding stage. The circular harmonic functions are sampled at the loudspeaker positions to compute the loudspeaker gains by multiplying the circular harmonics signals $\chi(t)$ with a decoding matrix $D = (1/L) \times [y(\varphi_{LS,1}), \dots, y(\varphi_{LS,L})]^T$. L represents the number of loudspeakers and φ_{LS} their respective azimuth angle. Combining encoding and decoding yields the loudspeaker signals

$$S_{LS,basic} = D \times \chi(t). \quad (4)$$

This approach is referred to as the *sampling* or the *basic* decoder. Another popular decoder is the so-called *max r_E*

decoder, designed to maximize the energy vector \vec{r}_E of the reproduced sound field (Gerzon, 1992). The energy vector is considered a predictor for sound localization at higher frequencies. The computation of its loudspeaker signals differs from the *basic* decoder through an added order-dependent weighting matrix $W = \text{diag}(w_0, w_1, w_1, \dots, w_N, w_N)$, where $w_n = \cos(n\pi/(2(N+1)))$, leading to the loudspeaker signals

$$S_{LS,maxre} = D \times W \times \chi(t). \quad (5)$$

D. Measurement procedure

After the equalization process was finished, the array was moved from -0.935 m to $+0.935$ m in the x -direction (see Fig. 2), relative to the center of the loudspeaker array, in 8.5 cm steps using two stepper motors (PD42-x-1240, TRINAMIC Motion Control GmbH & Co. KG, Hamburg, Germany) mounted on gear tracks. This results in 529 measurement positions arranged in a 23×23 measurement point grid with a spatial resolution of 8.5 cm, leading to an approximate spatial aliasing frequency of 2 kHz. A scale drawing of the loudspeaker array and the measurement area is shown in Fig. 2. At each position of the movable array, all seven frequency conditions, two stimuli conditions, and four reproduction methods were recorded with the linear microphone array. After the recordings were finished, the microphone array was automatically moved to the next measurement position.

E. Simulation setup

We implemented a numerical simulation to compute the reference sound field and an ideal baseline of the different SFS techniques. The loudspeakers were simulated as perfect point sources. Loudspeaker directivity, residual deviations after a non-ideal loudspeaker equalization (see Sec. II B 3), and reflections on the equipment were not considered. The simulated impulse responses were then convolved with each individual stimulus. Like in the measurement, onset and offset were windowed out to ensure steady-state behavior and avoid introducing level differences due to differences in time of arrival between loudspeakers and measurement points. The simulation results are used as a reference to derive the errors in the measured sound fields.

III. RESULTS

A. Measurement following ISO 8253-2:2009—Sound field audiometry

The measured SPL difference between each of the four side positions and the reference point in the center is shown in Fig. 3. At 500 Hz, all measured levels appear shifted upward, with the level for the front position exceeding the 1 dB limit for sine and FM tones. At 1 kHz, all side positions except the front show levels below the -1 dB line. At 2 kHz, levels appear shifted downward, and the back position is the only one strictly meeting the requirements for the

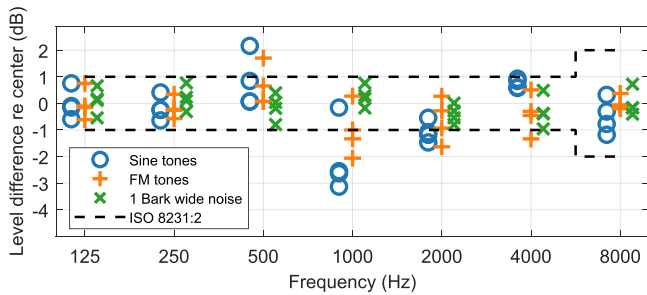


FIG. 3. (Color online) Results of ISO-style measurements. The difference between each of the four side positions and the center is plotted. Marker shapes distinguish the different stimuli: Blue circles show the results for sine tones, orange plus signs for FM tones, and green crosses for the Bark wide noise. The dotted line indicates the tolerance defined in ISO 8253-2:2009 (2009).

sine tone, while it is the only position falling outside of the requirements for the FM tones. At 4 kHz, the level at the back position is too low for the FM tones. The ISO 8253-2:2009 (2009) standard requires a difference of less than ± 1 dB for all stimuli for frequencies up to 4 kHz and ± 2 dB above. The requirements are met for critical-band-wide noise for all measured frequencies. The sine tones meet the requirements for 125 Hz, 250 Hz, 4 kHz, and 8 kHz, the FM tones for 125 Hz, 250 Hz, and 8 kHz. A second requirement by ISO 8253-2:2009 (2009) is that the difference between the left and right measurement positions does not exceed 3 dB for frequencies of 4 kHz and above, which is met by all stimuli. The final requirement is that the difference in measured level between front and back positions does not deviate from the theoretically expected value due to the $1/r$ distance law by more than ± 1 dB for any stimulus. For the loudspeaker distance of 2.4 m to the reference point, the difference between front and back measurement points should be 1.09 ± 1 dB. Here, most stimuli

comply with the requirements, except sine tones at 1 and 2 kHz and FM tones at 1 kHz.

B. Measurement of reproduced sound fields

1. SPLs

Figure 4 shows the measured SPL over the measurement grid. For both stimuli, the target level of 70 dB SPL at the center of the loudspeaker array is reached within 1 dB for NLS, HOA *basic*, and HOA *max r_E*. VBAP shows a level of around 72.7 dB. VBAP and HOA *max r_E* show quick level drops toward the side, perpendicularly to the 13° wavefront direction, reaching levels consistently 4 dB below the level at the center at 60 and 50 cm distance from the center, respectively. This is not the case with NLS and HOA *basic*, where the SPLs are almost constant laterally.

The amplitude decay with distance observed in a theoretical point source (represented in light gray lines in Fig. 4) can also be observed for the reproduced sound fields, but to a lower extent. Along the 13° axis, the level of the reference sound field decays by 10 dB over the whole reproduction area. VBAP, NLS, and HOA *max r_E* show a decay of 7 dB, HOA *basic* 8 dB.

The measured SPL for the sine tone (top row) shows some interference patterns, where the measured level locally deviates from the surrounding values by up to 2 dB. This effect is much reduced for the critical-band-wide noise, simply because of its wider bandwidth.

2. SPL errors

Figure 5 shows the absolute error in SPL, computed as the level difference between the measured sound fields and

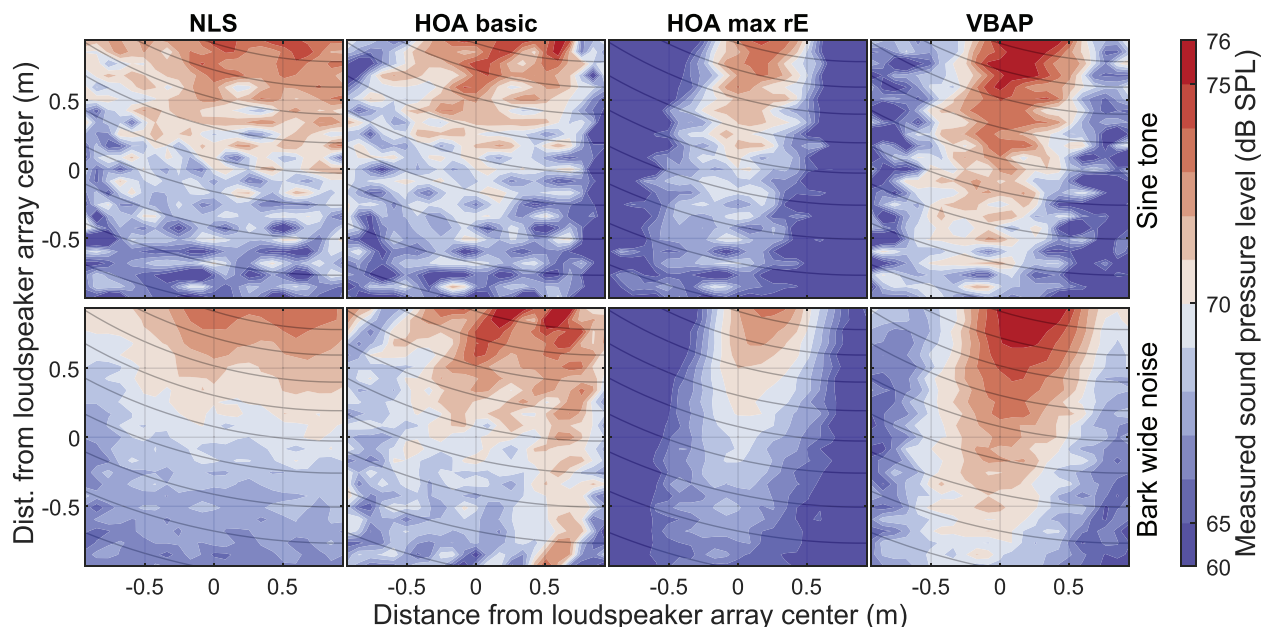


FIG. 4. (Color online) Measured SPL at 1 kHz measured with a sine tone (top row) and a critical-band-wide noise (bottom row) for the auralization of a point source at 4 m distance and 13° azimuth. The reproduction methods are NLS (left column), HOA with the *basic* decoder (second column), HOA with *max r_E* decoding (third column), and VBAP (right column). The light gray lines indicate the reference SPL contours in 1 dB steps.

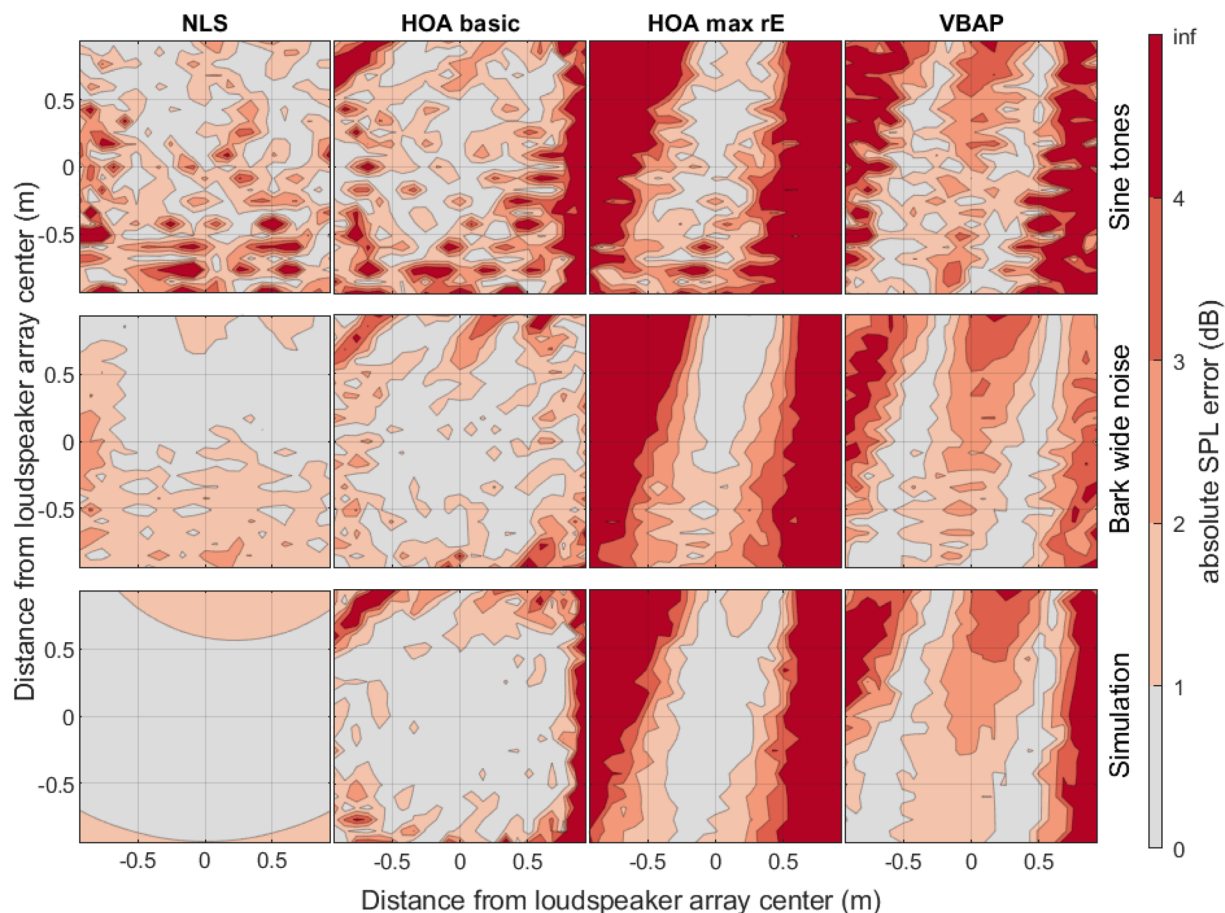


FIG. 5. (Color online) Absolute error of SPL at 1 kHz measured with a sine tone (top row), a critical-band-wide noise (middle row), and for the simulated reproduction of a sine tone (bottom row), for the auralization of a point source at 4 m distance and 13° azimuth. The reproduction methods are as presented in Fig. 4. The error is computed relative to the theoretical SPL of a point source at 4 m distance and 13° azimuth.

the reference sound field at 1 kHz. For the reproduction of a sine tone with NLS, the error is <1 dB (gray coded) and <2 dB (light red) at most measurement points around the array center for critical-band-wide noise. Striking are mostly individual measurement points where error increases to about 4 dB. This behavior is similar for HOA *basic*, although local error maxima appear reduced and further away from the center of the loudspeaker array. HOA *max r_E* shows even less individual variability but exhibits errors above 4 dB over a good part of the measurement area, providing a fairly narrow band of correct reproduction. The same trend is also visible with VBAP, except for the level offset of 2–3 dB at the center.

Looking at the critical-band-wide noise, the error for NLS is below 1 dB in a large part and generally <2 dB in the whole measurement zone. For HOA *basic*, the center area shows similarly low errors <1 dB, but errors increase to more than 3 dB at the edge of the measured area. For HOA *max r_E*, the error is already above 4 dB when moving 50 cm to the side but stays below 2 dB on-axis. A similar behavior is observed for VBAP, again with an offset in the error. The absolute error in the center is between 2 and 3 dB, going down to below 1 dB around 40 cm to the side of the center, before going up strongly to the sides. The error on-axis is between 1 and 4 dB.

The sine tone reproduction yields higher errors and more local variations than the critical-band-wide noise. Since the general error patterns of the simulated and measured tone reproductions are similar, we attribute the random deviations observed, particularly in the sine tone measurement, to the reproduction setup, as they do not appear as such in the simulation. More specifically, they are due to interferences between the direct sound and reflections, which can vary strongly across positions. The loudspeaker directivity does not explain these fluctuations, as discussed in Sec. IVC 3. The general distribution of the errors reflects the individual reproduction techniques, which is why we limit the rest of this work to the measured data.

3. Radial distribution of SPL errors

For VBAP and HOA *max r_E*, a listener moving away from the center notices two trends, depending on the direction in which they are moving. Going toward or away from the sound source location, SPL errors are lower than for a movement perpendicular to the virtual source direction, where the errors due to interferences between neighboring loudspeakers appear. This trend is different for HOA *basic*, where errors are relatively independent of direction relative

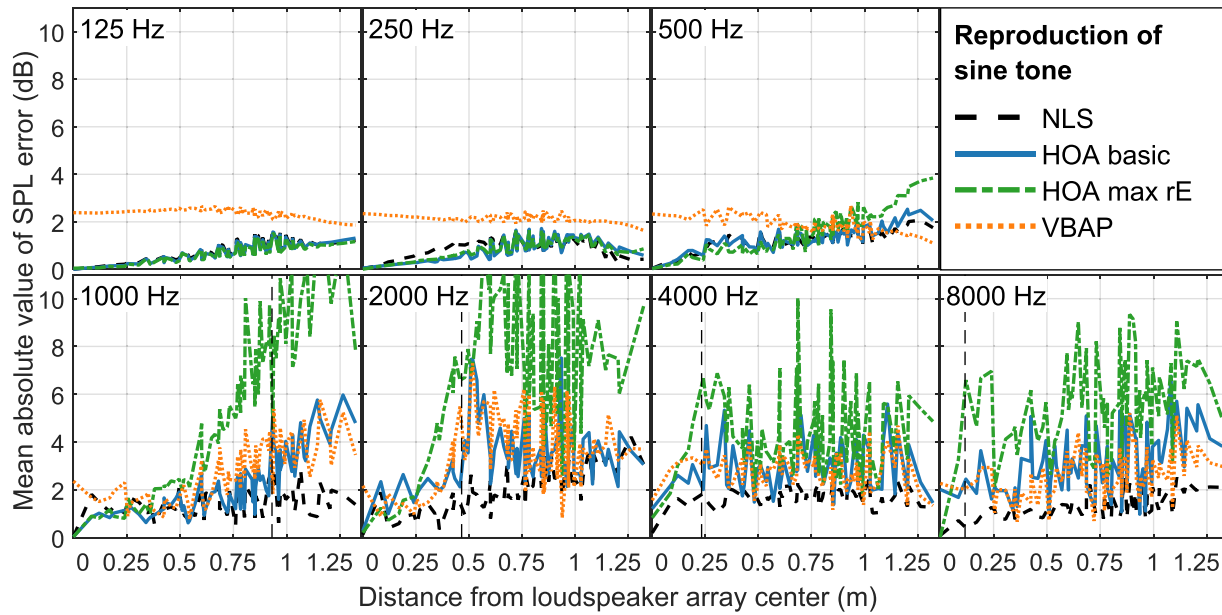


FIG. 6. (Color online) Mean absolute error (MAE) of the measured sine tone SPL. MAEs are computed relative to the theoretical sound field of a point source at 4 m and 13° azimuth and across measurement positions at equal distance from the center of the loudspeaker array. Different lines indicate the different sound field reproduction methods. Different panels show the results for the audiometric frequencies; the vertical dashed black line shows the sweet-spot radius $r = N/k$ at the given frequency. Note that for frequencies below 1 kHz, the sweet-spot radius is larger than 1.30 m and does not appear on the plot.

to the sound source, and for NLS, where side-to-side movements show lower errors than front-back movements. In more realistic environments, sources and reflections surround the listener, making a source direction-dependent description of a sweet-spot difficult. Hence, we opt for a source direction-independent description of the reproduction error. To compare different SFS methods, we computed the mean value of the absolute SPL error for all measurement positions at the same distance from the center of the loudspeaker array, deriving a metric only dependent on the listener's distance from the center of the loudspeaker array.

We chose this metric since the absolute SPL error is sensitive to destructive interference, which is easily heard and very annoying. Figures 6 and 7 summarize these results for sine tones and critical-band-wide noise, respectively.

The mean absolute SPL errors for both HOA methods increase with higher distance from the center of the loudspeaker array, as expected for Ambisonic synthesis. This increase in error is especially pronounced at higher frequencies. The error for the NLS method also increases, but at a lower rate than the HOA methods. VBAP shows a less prominent error increase with distance, for some conditions

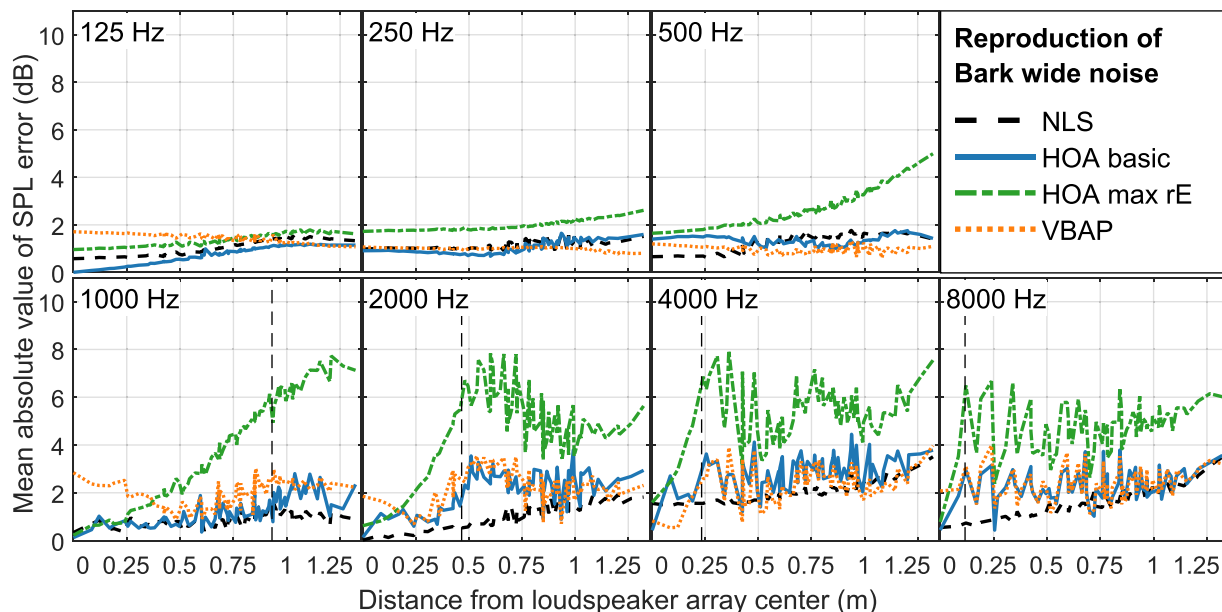


FIG. 7. (Color online) MAE of the measured critical-band-wide noise SPL, presented as in Fig. 6.

showing a rather constant azimuthally averaged error, although starting from a higher value.

For frequencies of 500 Hz and above, HOA $max r_E$ shows the highest errors, both for sine tones (Fig. 6) and critical-band-wide noise (Fig. 7), while NLS performs best. The difference between HOA *basic* and VBAP is less apparent and depends on the distance from the center of the array. There is a tendency for VBAP to show lower errors than HOA *basic* for distances above 1 m. When comparing the measured SPL errors for sine tones and critical-band-wide noise (Figs. 6 and 7), the errors are larger for sine tones, which is especially visible for higher frequencies. One exception to this trend is the HOA $max r_E$ method at 500 Hz, where the errors are slightly larger for critical-band-wide noise. The measured sine tone errors also exhibit a more chaotic course than those for critical-band-wide noise.

In the center of the loudspeaker array, at which loudspeakers have been equalized, the SPL error is ideally expected to be zero for both HOA methods and NLS and 3 dB for VBAP (further discussed in Sec. IV C 4). For sine tones, the errors are very close to 0 dB for frequencies below 4 kHz. At 4 kHz, HOA $max r_E$ gives an error of 0.8 dB in the center, and HOA *basic* 1.2 dB. At 8 kHz, the error for HOA *basic* goes up to 2 dB, while HOA $max r_E$ drops back to 0 dB. The errors are higher for critical-band-wide noise, with HOA $max r_E$ reaching 1.7 dB at 250 Hz, NLS 1.6 dB at 4 kHz, and HOA *basic* 1.4 dB at 500 Hz. VBAP deviates from the expected 3 dB by up to 1.4 dB for a sine tone and 2.2 dB for critical-band-wide noise, both at 4 kHz.

The measured sweet-spot size, which we defined here as the radius for which the azimuthally averaged errors are below 2 dB, depends on frequency, reproduction method, and source stimulus. Figures 6 and 7 show that the sweet-spot size $r \leq N/k$ needs to be halved to give a good approximation of the measured sweet-spot size. For example, the theoretical sweet-spot radius at 1 kHz is $r = N/k = 0.93$ m. The measured sweet-spot for a sine tone at 1 kHz lies at 0.4 m for HOA $max r_E$, 0.6 m for HOA *basic*, and 0.5 m for NLS. For a critical-band-wide noise around 1 kHz, it lies at 0.4 m for HOA $max r_E$, 0.9 m for HOA *basic*, and beyond the maximal measurement distance of 1.3 m for NLS. In the case of VBAP, the sweet-spot size cannot be estimated that way, as the level errors in the center already surpass 2 dB. However, the general SPL shape is quite close to HOA $max r_E$, indicating that their sweet-spots would be of similar sizes if the VBAP level offset in the center was normalized out or corrected for. Conversely, a target radius of 0.5 m corresponds to an upper frequency bound of around 800 Hz for tones and 400 Hz for critical-band-wide noise with HOA $max r_E$, 1.2 kHz for tones and 2 kHz for critical-band-wide noise with HOA *basic*, and 1 kHz for tones and above 8 kHz for critical-band-wide noise with NLS.

IV. DISCUSSION

The evaluation presented above assesses the accuracy of sound reproduction over the 36 horizontal loudspeakers

of a 61-channel array in an anechoic chamber. The first measurement investigated accuracy for free-field audiology according to ISO 8253-2:2009 (2009) with pure tones, FM tones, and critical-band-wide noise. In a second measurement, the sound field generated by four SFS methods was measured every 8.5 cm across a 187 cm \times 187 cm wide area with tones and critical-band-wide noise at audiometric frequencies between 125 Hz and 8 kHz. Considering a 2 dB error as the limit for a sweet-spot, we found the measured sweet-spot to be about half as large as the $r \leq N/k$ approximation. We found that NLS delivers the largest sweet-spot, followed by HOA *basic*, while the error quickly increases away from the center for HOA $max r_E$ and VBAP. The factors affecting the reproduced sound field, the differences between the SFS methods, and the effect of stimuli are discussed next.

A. ISO requirements

Figure 3 shows that the requirements published in ISO 8253-2:2009 (2009) are met for critical-band-wide noise, while sine tones and FM tones fall outside of the tolerated range for frequencies of 500, 1000, and 2000 Hz and 500, 1000, 2000, and 4000 Hz, respectively. The deviations observed for the narrowband stimuli can be attributed to destructive interferences and reflections on the loudspeaker array, as discussed in Sec. IV B. Because of their high susceptibility to large sound pressure variations, the use of sine tones for audiometric measurements is not recommended (Dillon and Walker, 1982). While the requirements are, strictly speaking, not met, the highest variation in level observed was 3.1 dB, which is lower than the 5 dB steps with which hearing thresholds are usually measured and was shown to be clinically insignificant (British Society of Audiology, 2019). These values are also close to the standard deviation in the frequency response of commonly used headphones measured on an artificial head by Hirahara (2004), indicating no loss of precision compared to headphone-based audiometry. We can conclude that measurements with tonal stimuli are accurate to 3 dB, and in most cases much more accurate. In contrast, local level errors average out as soon as stimulus bandwidth increases, to achieve an accuracy better than 1 dB for the energy of noise within a critical band. The MAEs across side positions and frequencies are 0.9 dB for sine tones, 0.6 dB for FM tones, and 0.4 dB for critical-band-wide noise.

B. Accuracy of sound field reproduction techniques

When comparing the mean absolute SPL error of the reproduced sound field across distance from the center (Figs. 6 and 7), we observe that the errors for sine tones are higher and show more variation than the errors for critical-band-wide noise, especially at higher frequencies and higher distances from the center of the loudspeaker array. This behavior was expected, since level deviations are more pronounced due to the narrowband nature of tones, whereas they average out for critical-band-wide noise.

In general, the mean absolute SPL errors start to increase before the $r = N/k$ sweet-spot radius is reached, showing that the ideal performance of a HOA reproduction could not be met in real loudspeaker arrays. Notably, the error increase happens at a substantially smaller distance from the theoretical $r \leq N/k$ sweet-spot radius, at about half the distance from the center. Murillo *et al.* (2014) measured the level error for 3D, fifth-order HOA *basic* and HOA *max r_E*. While they used fifth-order HOA in 3D, vs the 2D 17th-order HOA used in this study, we can still compare the errors inside the sweet-spot across measurements. Observed SPLs deviate inside the sweet-spot up to 3 dB for tones at 250 Hz, 5 dB at 1 kHz, and 4 dB at 2 kHz for HOA *basic* and up to 4 dB at 250 Hz, 5 dB at 1 kHz, and 5 dB at 2 kHz for HOA *max r_E*. Note that the color bars in their plots do not depict errors beyond ± 5 dB. We observe MAEs below 2 dB inside the sweet-spot for a 250 Hz tone and comparable errors around 4 dB for 1 and 2 kHz tones (Fig. 6).

The size and shape of the zones where the sound field reproduction resembles that of a point source vary strongly across decoders (Fig. 4). Interestingly, the HOA *basic* decoder performs similarly to the NLS method. This can be explained by the dominance of the loudspeaker at 10° in the HOA *basic* method (7 and 13 dB above the neighboring loudspeakers), which is also used for the NLS method. The remaining loudspeakers help to shape the sound field and reduce the errors close to the center of the loudspeaker array, but their influence is too little to compensate for the larger errors at specific points introduced by the dominant loudspeaker (e.g., high reproduction error at $[x, y] = [-0.2, 0.2]$).

The HOA *max r_E* method leads to the highest sound field reproduction errors, which is contrary to previous studies, both for measured and simulated results (Murillo *et al.*, 2014; Otani and Shigetani, 2019), and goes against the label “high-frequency decoder” often seen in the literature (e.g., Gerzon, 1992; Daniel, 2001). This can be explained by stronger destructive interferences between the loudspeakers. The *max r_E* weighting increases the directional spread across the loudspeaker gains, reducing the level differences between the loudspeakers with the highest playback level to only 4 dB, thereby increasing destructive interferences. In non-anechoic environments, where destructive interferences are reduced due to room reverberation, *max r_E* was found to improve localization at off-center positions (Frank *et al.*, 2008) and reduce coloration fluctuation for moving sources (Zotter and Frank, 2019).

The lowest errors are achieved using the NLS method (Figs. 6 and 7), followed by the HOA *basic* method. For the HOA *basic* method, the large level differences between the dominant loudspeaker and the others lead to a reduced impact of the other loudspeakers, leading to a smaller error in SPL where destructive interferences occur. This explains why the methods that yield higher level differences between the dominant loudspeaker and the others (HOA *basic* and NLS) perform better for greater distances to the center of the loudspeaker array. Zotter *et al.* (2023) recently

employed this effect by purposefully introducing amplitude differences in neighboring loudspeakers to reduce comb filtering in anechoic conditions. Overall, there seems to be a trade-off between reducing the local variations in the sound field and minimizing interferences between loudspeaker signals and various reflections on the hardware.

C. Reproduction errors due to hardware setup

1. Mismatch between virtual and physical sound source distance

The unequal repartition of the level error around a circle of a given radius in the reproduction zone was expected since the virtual source was intentionally placed at a distance of 4 m but was played back with physical sound sources about 2.4 m away (most of the energy was radiated by loudspeakers at 0° , 10° , and 20°). This leads to a mismatch according to the $1/r$ -law, where the distance decay of the reproduced sound field is stronger than the one expected from the virtual source since the virtual source is located further away. This does not affect the level in the center, where the loudspeakers are calibrated and the SPL is referenced to, but leads to different level changes as participants move toward or away from the virtual sound source. This mismatch appears for all reproduction methods studied here, every HOA-based reproduction method, VBAP, and NLS. One approach to decrease this effect is using NFC-HOA (Daniel, 2003), which accounts for the distance difference between loudspeakers and virtual sources. One of the drawbacks of this method is the filter implementation, which requires a precise low-frequency response, which is difficult for a real-time FIR-based acoustic simulation and rendering setup (Ahrens, 2012).

2. Destructive interferences

Any time two or more sources play coherent signals, interference patterns can appear and create deep, narrow notches in the frequency spectrum. When using equalized loudspeakers that are driven with opposite phase signals (as is the case with HOA *basic* and HOA *max r_E* decoders), destructive interferences easily appear. When considering a measurement point not in the center of the loudspeaker array, the time of arrival differences between loudspeaker signals create a phase difference. The amount of comb filtering depends on the amplitude and the phase of the different loudspeaker signals. The resulting series of wavefronts can be understood as a dominant sound (coming from the loudspeaker with the highest playback level) and a series of leading or lagging reflections. Figure 8 shows the SPL change induced by adding a single reflection to a tone. Even low relative levels are enough to induce a substantial change: For instance, a reflection of -19 dB at 180° phase shift is already enough to reduce the level of the dominant sound by 1 dB, and a reflection of -18 dB in phase raises the level of the dominant sound by 1 dB. While predicting the effect of several reflections on the overall level is more difficult, their effect is at least as important as for a single reflection.

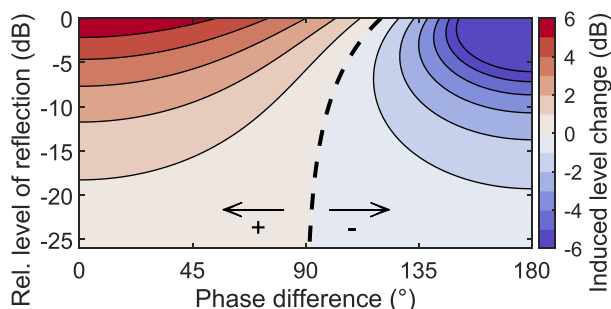


FIG. 8. (Color online) Level difference introduced by adding a reflection with a different phase and level to a tone. Contour lines are drawn in 1 dB steps, the thicker dashed line showing 0 dB.

Unlike level errors, which can at least be compensated for on average, minimizing phase errors and, thus, interferences over a wider area is more difficult.

Destructive interferences also occur due to reflections on the loudspeakers themselves. While great care was taken to place absorbers on loudspeaker baffles and suspensions, the reflections cannot be completely absorbed. Furthermore, the symmetrical arrangement of the loudspeakers can cause stronger interferences at the center point, since reflections from opposite sides of the loudspeaker array arrive in phase there and add up. This effect is visible in the ISO-style measurement (Fig. 3), particularly for sine and FM tones at 1 kHz, where the center and front measurement point levels are about 2 dB higher than the back, left, and right measurement point levels. Here, again, the addition of several reflections can amplify their effect. Pausch *et al.* (2020) give a good example by comparing the frequency response of an equalized loudspeaker measured in a hemi-anechoic chamber with its frequency response measured in the SCaLAR loudspeaker array. They observed deviations for frequencies above 300 Hz, with narrowband deviations of up to 6 dB.

3. Loudspeaker directivity

Another error source is the directivity pattern of the loudspeakers. Deviations from the on-axis response are especially pronounced at high frequencies. In our measurement setup, the highest angle deviation from the loudspeaker axis is 33°, reached for the top-left corner position of the measurement area and the loudspeaker at 0°. For this angle, directivity measurements of the loudspeakers show a drop of 4–5 dB around 8 kHz. Simulations show that even for 8 kHz, the SPL deviations from an ideal loudspeaker simulation are below 1 dB for radii of up to 50 cm for all SFS methods when considering critical-band-wide noise. NLS and VBAP show the same deviations for sine tones, but the HOA *basic* and HOA *max r_E* methods show deviations slightly above 1 dB at individual measurement points. At the corner of the measurement setup, the deviations increase to 4 dB.

These small deviations in most of the measured area let us conclude that while the loudspeaker directivity does have an effect that should not be ignored for strongly off-center

positions, their effect closer to the center is very low and is likely to be insignificant compared to the errors introduced by SFS methods and unwanted reflections inside the VAE.

4. Loudspeaker equalization

The reproduction methods used for this measurement are based on different assumptions. While HOA assumes phase coherent loudspeakers and normalizes the sum of the amplitude of the loudspeaker signals to achieve the target level, VBAP uses a power normalization to compute the loudspeaker gains (Pulkki, 1997). This results in a theoretical 3 dB error for VBAP at the center of the loudspeaker array. For sine tones, the level error in the center is about 2 dB, even at higher frequencies (Fig. 6). The error is also around 2 dB for critical-band-wide noise, except for 250 Hz and 4 kHz. For the other decoders, where an error of 0 dB was expected at the center, the sine tone reproduction yields good results, except for the HOA *basic* method above 4 kHz and the HOA *max r_E* method at 4 kHz.

In the case of the critical-band-wide stimulus, there are some residual errors in the center, up to 1.6 dB for HOA *basic* and NLS at 4 kHz and 2 dB for the HOA *max r_E* method. This shows that the loudspeaker equalization based on short 1024-tap FIR filters was not sufficient to ensure perfect phase alignment and amplitude compensation between the loudspeakers for narrowband stimuli such as critical-band-wide noise, which is about as narrow at low frequencies as an FFT-bin of the used filter.

D. The use of SFS for psychoacoustic research

1. Common psychoacoustic experiments

When running tone-in-noise detection experiments, both the SPL of the target and of the masker directly influence the measured thresholds. As shown in Fig. 5, the level errors for tones and noise are different, even when centered at the same frequency. When taking the difference of the levels of tones and critical-band-wide noise, a threshold estimation error can be defined. Considering a 1 kHz tone and the smallest sweet-spot estimation of 40 cm, the deviation lies between -1.1 and 3.4 dB for NLS, between -1.2 and 3.0 dB for HOA *basic*, between -1.6 and 1.7 dB for HOA *max r_E*, and between -2.0 and 1.0 dB for VBAP. In short, the accuracy for determining the tone-in-noise ratio is better than 3 dB, in most situations better than 2 dB. If participants are not at a fixed position, but within the sweet-spot area, we can—due to the symmetry of the errors—expect that these errors average out, but more trials will be needed to determine a stable threshold.

According to the profile analysis theory (Green, 1983), intensity discrimination is performed by comparing spectral intensity differences between critical bands across trials, rather than by directly comparing intensities of the relevant critical band across frequencies. If the listener stays in the same position for a single stimulus, the frequency spectrum does not change, and discrimination performance should not be affected.

On the one hand, it is difficult to quantify the influence of spectral changes on speech reception thresholds, as intelligibility is based on integration across auditory filters and strongly benefits from binaural unmasking. On the other hand, results show that reproduced critical-band levels are very stable for all reproduction methods. Since the intelligibility of speech is derived from a wideband signal, the average reproduced speech level will be even more immune to local spectral deviations than critical-band-wide signals. A simulation by Kuntz and Seeber (2023) also shows that the binaural benefit to speech intelligibility modeled following Jelfs *et al.* (2011) is generally reproducible in the SOFE loudspeaker array, with errors below 1 dB in a radius of 1 m. Considering measurement errors of 1–2 dB around the center when using HOA *basic* or NLS, we can conclude that the speech intelligibility error is small and can be expected to be 1–2 dB. This is also supported by the findings of Hládek *et al.* (2021), who showed that speech intelligibility measures from binaural recordings played back over headphones match those measured from the modeled environment and reproduced over the SOFE loudspeaker array with less than 2 dB deviation.

2. Listener movement

In many psychoacoustic experiments in the free field, the participants sit or stand in one place inside the reproduced sound field. Since the participants are static, the level errors they experience should be low, although some subjects would hold their head up more than others, leading to position shifts of 10–20 cm. Head rotations also introduce a translation of the ear positions of around 10 cm, depending on rotation angle and head size. Figure 9 shows the SPL gradient at each measurement point, computed from adjacent measurement points and related to a value in dB per 10 cm

translation, representing the expected level change for head turns or slight translations. When reproducing tones with NLS, the gradient around the center is about 1 dB larger than for the other reproduction methods, for which it can reach 3 dB. The picture is inverted when reproducing critical-band-wide noise: Local variations drop below 1 dB for NLS but remain up to 2 dB for the other methods. This indicates a larger impact of acoustic reflections on NLS than on the other methods, while NLS on its own yields lower errors for small translations. Overall results show that the considered methods are reasonably robust against head turns for static participants, especially for broadband stimuli.

When allowing the listener to move freely inside the loudspeaker array, the movement introduces position-dependent spectral changes, also perceived as timbre changes, as the reproduction accuracy at a given point is frequency-dependent. Since the gradient of SPL remains low around the center, slow movements should also not affect psychoacoustic research with all but tonal stimuli. Changes in critical-band level are very small and below the just noticeable level difference (<1 dB/10 cm; Fig. 9, bottom row), i.e., the critical-band level remains mostly constant. Only when the subject moves 0.5 m from the center, SPL errors increase substantially and should become critical for psychoacoustic research.

V. CONCLUSION

This work investigated the practically achievable SPL accuracy for free-field sound reproduction via a loudspeaker array in an anechoic chamber. More particularly, it discusses the effect of sound reflections and imperfections that cannot be avoided in physical reproduction setups on the accuracy of SFS methods. A first measurement verified the requirements for free-field audiometry as stated in the ISO

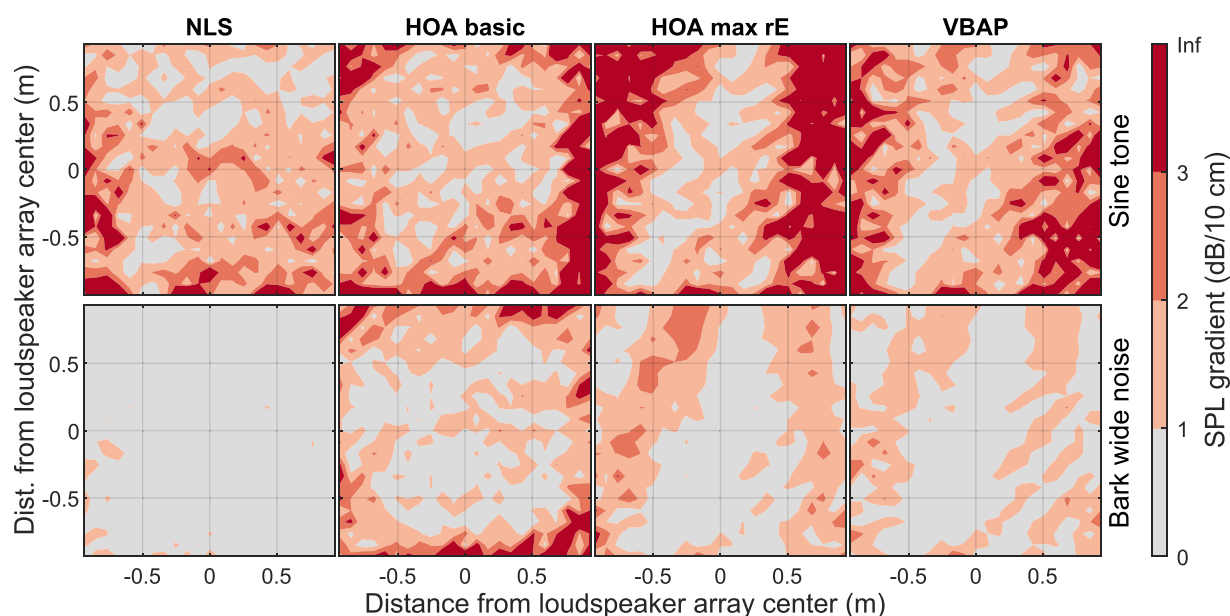


FIG. 9. (Color online) Gradient of the measured SPL for a sine tone (top row) and critical-band-wide noise (bottom row) at 1 kHz, reproduced with different auralization methods.

8253–2:2009 (2009) standard. We found that the loudspeaker array of the SOFE is suitable for free-field audiometry via the frontal loudspeaker for critical-band-wide noise with errors <1 dB at all audiometric frequencies. Pure tones fall outside the 1 dB-error requirement for frequencies of 500 Hz, 1 kHz, and 2 kHz, but the deviation remains below 2 dB, except for 1 kHz, where it is up to 3 dB. For FM tones, the level falls outside the tolerated 1 dB range between 500 Hz and 4 kHz but remains <2 dB at all frequencies. Since the deviations observed are much smaller than those tolerated in clinical settings and within the error range of common psychometric methods and headphone frequency responses, basic audiology measurements are, in practice, feasible in this particular VAE. Furthermore, our results raise the question whether the requirements defined by ISO 8253–2:2009 (2009) might be too strict to be applied in VAEs with many loudspeakers.

When reproducing sound fields over a wide spatial area, level differences between the target sound field and the measured reproduced sound field are smallest for NLS and HOA using the *basic* decoder. In general, the $r = N/k$ theoretical sweet-spot size needs to be halved to yield a good estimate of the measured sweet-spot size. We found that the HOA with *max* r_E weighting performed the worst in anechoic conditions, which goes against its description as a decoding method best suited for high frequencies. This is due to a reduced level difference between adjacent loudspeakers, creating stronger destructive interferences.

When deviating from the center, the overall SPL error is dominated by the reproduction technique. We showed that the directivity of the loudspeaker only affects the sound field for larger distances from the center. While local variations in the sound field can be attributed to the interferences of unwanted reflections in the VAE, the general distribution of the errors is determined by the sound field reproduction method used.

Using HOA with the *basic* decoder up to 2 kHz and the NLS above 2 kHz seems like a good option to reduce level errors at high frequencies and remedy the shortcomings of Ambisonics at high frequencies, while providing a more accurate, direction-dependent sound field at low frequencies.

Overall, measurements demonstrate that the theoretical $r = N/k$ sweet-spot size should be halved to give a realistic prediction of a measured sweet-spot. For instance, level errors below 2 dB can be achieved in a sweet-spot of 50 cm radius in each critical band of broadband sounds by using HOA with the *basic* decoder for frequencies up to 2 kHz and NLS above that. These low errors, coupled with a low spatial rate of level change, recommend the SOFE VAE for psychoacoustic measurements.

ACKNOWLEDGMENTS

We thank Philipp Hortig for help with building the linear array measurement system. We thank two anonymous reviewers and the editor for their helpful comments. This study was funded by TUM and by the Deutsche

Forschungsgemeinschaft (DFG, German Research Foundation), Project ID 352015383–SFB 1330 C5. The rtSOFE system was funded by BMBF Grant No. 01GQ1004B (Bernstein Center for Computational Neuroscience Munich). The data that support the findings of this study are available upon request. The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

¹The *max* r_E weighting can be applied to higher-order Ambisonics coefficients to reduce the sidelobes of the Ambisonic panning function.

- Ahrens, A., Marschall, M., and Dau, T. (2019). “Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments,” *Hear. Res.* **377**, 307–317.
- Ahrens, J. (2012). *Analytic Methods of Sound Field Synthesis* (Springer, Berlin), p. 180.
- Ahrens, J., and Spors, S. (2009). “Sound field reproduction employing non-omnidirectional loudspeakers,” in *Proceedings of the 126th Convention of the Audio Engineering Society*, May 7–10, Munich, Germany (Audio Engineering Society, New York).
- Au, E., Xiao, S., Hui, C. T. J., Hioka, Y., Masuda, H., and Watson, C. I. (2021). “Speech intelligibility in noise with varying spatial acoustics under Ambisonics-based sound reproduction system,” *Appl. Acoust.* **174**, 107707.
- Bischof, N. F., Aublin, P. G., and Seeber, B. U. (2023). “Fast processing models effects of reflections on binaural unmasking,” *Acta Acust.* **7**, 11.
- British Society of Audiology (2019). *Practice Guidance: The Acoustics of Sound Field Audiometry in Clinical Audiological Applications* (British Society of Audiology, Seaford, UK).
- Buchholz, J. M., and Best, V. (2020). “Speech detection and localization in a reverberant multitalker environment by normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **147**, 1469–1477.
- Daniel, J. (2001). “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia” (“Acoustic field representation, application to transmission and reproduction of complex scenes in a multimedia context”), Ph.D. dissertation, Université Paris 6, Paris, France.
- Daniel, J. (2003). “Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format,” in *Proceedings of the 23rd International Conference of the Audio Engineering Society*, May 23–25, Copenhagen, Denmark (Audio Engineering Society, New York).
- Dillon, H., and Walker, G. (1982). “Comparison of stimuli used in sound field audiometric testing,” *J. Acoust. Soc. Am.* **71**, 161–172.
- Fastl, H., and Zwicker, E. (2007). *Psychoacoustics—Facts and Models* (Springer, Berlin), p. 159.
- Favrot, S. (2010). “A loudspeaker-based room auralization system for auditory research,” Ph.D. dissertation, Technical University of Denmark, Kgs. Lyngby, Denmark.
- Favrot, S., and Buchholz, J. M. (2010). “LoRA: A loudspeaker-based room auralization system,” *Acta Acust. united Acust.* **96**, 364–375.
- Frank, M., Zotter, F., and Sontacchi, A. (2008). “Localization experiments using different 2D Ambisonics decoders,” in *Proceedings of 25th Tonmeisteragung—VDT International Convention*, November 13–16, Leipzig, Germany (Verlag VDT, Bergisch-Gladbach, Germany).
- Gerzon, M. A. (1992). “General metatheory of auditory localization,” in *Proceedings of the 92nd Convention of the Audio Engineering Society*, March 24–27, Vienna, Austria (Audio Engineering Society, New York).
- Gomez, G. (2019). “Consolidating natural spatial perception and improved SNR in hearing aids: Jackrabbit, a new method,” Ph.D. dissertation, Technical University of Munich, Munich, Germany.
- Grandjean, P. (2021). “Captation 3-D de métriques directionnelles et reproduction de champs sonores par la combinaison d’Ambisonies circulaires et sphériques” (“3D capture of directional metrics and reproduction of sound fields by combining circular and spherical Ambisonics”), Ph.D. dissertation, Université de Sherbrooke, Sherbrooke, Canada.

- Grantham, D. W. (1986). "Detection and discrimination of simulated motion of auditory targets in the horizontal plane," *J. Acoust. Soc. Am.* **79**, 1939–1949.
- Green, D. M. (1983). "Profile analysis: A different view of auditory intensity discrimination," *Am. Psychol.* **38**, 133–142.
- Grimm, G., Ewert, S. D., and Hohmann, V. (2015). "Evaluation of spatial audio reproduction schemes for application in hearing aid research," *Acta Acust. united Acust.* **101**, 842–854.
- Hadley, L. V., Brimijoin, W. O., and Whitmer, W. M. (2019). "Speech, movement, and gaze behaviours during dyadic conversation in noise," *Sci. Rep.* **9**, 10451.
- Hendrikse, M. M. E., Eichler, T., Hohmann, V., and Grimm, G. (2022). "Self-motion with hearing impairment and (directional) hearing aids," *Trends Hear.* **26**, 233121652210787.
- Hirahara, T. (2004). "Physical characteristics of headphones used in psychophysical experiments," *Acoust. Sci. Technol.* **25**, 276–285.
- Hládek, L., Ewert, S. D., and Seeber, B. U. (2021). "Communication conditions in virtual acoustic scenes in an underground station," in *Proceedings of 2021 Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, September 8–10, Bologna, Italy (IEEE, New York).
- Hládek, L., and Seeber, B. U. (2023). "Effects of noise presence and noise position on interpersonal distance in a triadic conversation," in *Proceedings of InterNoise22*, August 21–24, Glasgow, UK (Institute of Noise Control Engineering, Reston, VA), pp. 5326–5331.
- ISO 8253-2:2009 (2009). "Acoustics—Audiometric test methods Part 2: Sound field audiometry with pure-tone and narrow-band test signals" (International Organization for Standardization, Geneva, Switzerland).
- Jelfs, S., Culling, J. F., and Lavandier, M. (2011). "Revision and validation of a binaural model for speech intelligibility in noise," *Hear. Res.* **275**, 96–104.
- Kerber, S., and Seeber, B. U. (2012). "Sound localization in noise by normal-hearing listeners and cochlear implant users," *Ear Hear.* **33**, 445–457.
- Kerber, S., and Seeber, B. U. (2013a). "Localization in reverberation with cochlear implants: Predicting performance from basic psychophysical measures," *J. Assoc. Res. Otolaryngol.* **14**, 379–392.
- Kerber, S., and Seeber, B. U. (2013b). "Psychoacoustic abilities of CI users in relation to speech understanding and localization in a reverberant room," in *Fortschritte der Akustik–DAGA '13*, March 18–21, Merano, Italy [Deutsche Gesellschaft für Akustik e.V. (DEGA), Berlin, Germany], pp. 227–230.
- Kolotzek, N. F., and Seeber, B. U. (2020). "Localizing the end position of a circular moving sound source near masked threshold," in *Proceedings of Forum Acusticum*, December 7–11 [European Acoustics Association (EAA), Madrid, Spain], online, pp. 1925–1928.
- Kuntz, M., and Seeber, B. U. (2020). "Wiedergabe von Schallfeldern: Analyse der interauralen Merkmale" ("Sound field reproduction: analysis of interaural cues"), in *Fortschritte der Akustik–DAGA '20*, March 16–19 [Deutsche Gesellschaft für Akustik e.V. (DEGA), Berlin, Germany], online, pp. 9–12.
- Kuntz, M., and Seeber, B. U. (2023). "Spatial audio for interactive hearing research," in *Proceedings of InterNoise22*, August 21–24, Glasgow, UK (Institute of Noise Control Engineering, Reston, VA), pp. 5120–5127.
- Litovsky, R. Y. (2005). "Speech intelligibility and spatial release from masking in young children," *J. Acoust. Soc. Am.* **117**, 3091–3099.
- McCartney, D. A. (2013). "Development of the AnimalSeek method to evaluate the localisation ability of children under five," Ph.D. dissertation, University of Nottingham, Nottingham, UK.
- Murillo, D. M., Fazi, F. M., and Shin, M. (2014). "Evaluation of Ambisonics decoding methods with experimental measurements," in *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, April 3–5, Berlin, Germany (Universitätsbibliothek Technische Universität Berlin, Berlin, Germany), pp. 33–40.
- Otani, M., and Shigetani, H. (2019). "Reproduction accuracy of higher-order Ambisonics with Max-rE and/or least norm solution in decoding," *Acoust. Sci. Technol.* **40**, 23–28.
- Pausch, F., Behler, G., and Fels, J. (2020). "SCaLAr—A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments," *Acta Acust.* **4**, 19.
- Pullella, P., Hládek, L., Croce, P., and Seeber, B. U. (2021). "Auralization of acoustic design in primary school classrooms," in *Proceedings of the 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, September 7–10, Bari, Italy (IEEE, New York), pp. 1–6.
- Pulkki, V. (1997). "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.* **45**, 456–466.
- Salanger, M., Lewis, D., Vallier, T., McDermott, T., and Dergan, A. (2020). "Applying virtual reality to audiovisual speech perception tasks in children," *Am. J. Audiol.* **29**, 244–258.
- Schroeder, M. R. (1973). "Computer models for concert hall acoustics," *Am. J. Phys.* **41**, 461–471.
- Seeber, B. U., and Clapp, S. W. (2017). "Interactive simulation and free-field auralization of acoustic space with the rtSOFE," *J. Acoust. Soc. Am.* **141**, 3974.
- Seeber, B. U., and Clapp, S. W. (2020). "Auditory room learning and adaptation to sound reflections," in *The Technology of Binaural Understanding*, edited by J. Blauert and J. Braasch (Springer, Cham, Switzerland), pp. 203–222.
- Seeber, B. U., Kerber, S., and Hafter, E. R. (2010). "A system to simulate and reproduce audio-visual environments for spatial hearing research," *Hear. Res.* **260**, 1–10.
- Shulberg, M. (1980). "Free field pure tone audiometry as a comparative system for the assessment and fitting of hearing aids," *J. Biomed. Eng.* **2**, 197–200.
- Stitt, P., Bertet, S., and Van Walstijn, M. (2013). "Perceptual investigation of image placement with Ambisonics for non-centred listeners," in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, September 2–5, Maynooth, Ireland (DAFX, Maynooth, Ireland).
- Stitt, P., Bertet, S., and Van Walstijn, M. (2014). "Off-centre localisation performance of Ambisonics and HOA for large and small loudspeaker array radii," *Acta Acust. united Acust.* **100**, 937–944.
- Thery, D., Boccaro, V., and Katz, B. F. G. (2019). "Auralization uses in acoustical design: A survey study of acoustical consultants," *J. Acoust. Soc. Am.* **145**, 3446–3456.
- Traumüller, H. (1990). "Analytical expressions for the tonotopic sensory scale," *J. Acoust. Soc. Am.* **88**, 97–100.
- van de Par, S., Ewert, S. D., Hládek, L., Kirsch, C., Schütze, J., Llorca-Boff, J., Grimm, G., Hendrikse, M. M. E., Kollmeier, B., and Seeber, B. U. (2022). "Auditory-visual scenes for hearing research," *Acta Acust.* **6**, 55.
- Vorländer, M., Schröder, D., Pelzer, S., and Wefers, F. (2015). "Virtual reality for architectural acoustics," *J. Build. Perform. Simul.* **8**, 15–25.
- Ward, D. B., and Abhayapala, T. D. (2001). "Performance bounds on sound field reproduction using a loudspeaker array," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 24, New Paltz, NY (IEEE, New York), pp. 187–190.
- Weißgerber, T. (2019). "Ein Wiedergabesystem mit Wellenfeldsynthese zur Simulation alltäglicher Hörumgebungen" ("A sound reproduction system using wave field synthesis to simulate everyday listening conditions"), *HNO* **67**, 265–271.
- Zotter, F., and Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Springer, Cham, Switzerland), p. 66.
- Zotter, F., Frank, M., Bayer, O., Pinkas, J., and Müller, G.-J. (2023). "Amplitude-decorrelation avoiding spectral artifacts of sound field synthesis in the anechoic chamber or free sound field," in *Proceedings of InterNoise22*, August 21–24, Glasgow, UK (Institute of Noise Control Engineering, Reston, VA), pp. 5568–5577.