Original Article

Spatial Resolution of Late Reverberation in Virtual Acoustic Environments

Trends in Hearing
Volume 25: 1–17
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23312165211054924
journals.sagepub.com/home/tia
SAGE

Christoph Kirsch[1], Josef Poppitz[2], Torben Wendt[1,2], Steven van de Par[2], and Stephan D. Ewert[1]

Abstract

Late reverberation involves the superposition of many sound reflections, approaching the properties of a diffuse sound field. Since the spatially resolved perception of individual late reflections is impossible, simplifications can potentially be made for modelling late reverberation in room acoustics simulations with reduced spatial resolution. Such simplifications are desired for interactive, real-time virtual acoustic environments with applications in hearing research and for the evaluation of hearing supportive devices. In this context, the number and spatial arrangement of loudspeakers used for playback additionally affect spatial resolution. The current study assessed the minimum number of spatially evenly distributed virtual late reverberation sources required to perceptually approximate spatially highly resolved isotropic and anisotropic late reverberation and to technically approximate a spherically isotropic sound field. The spatial resolution of the rendering was systematically reduced by using subsets of the loudspeakers of an 86-channel spherical loudspeaker array in an anechoic chamber, onto which virtual reverberation sources were mapped using vector base amplitude panning. It was tested whether listeners can distinguish lower spatial resolutions of reproduction of late reverberation from the highest achievable spatial resolution in different simulated rooms. The rendering of early reflections remained unchanged. The coherence of the sound field across a pair of microphones at ear and behind-the-ear hearing device distance was assessed to separate the effects of number of virtual sources and loudspeaker array geometry. Results show that between 12 and 24 reverberation sources are required for the rendering of late reverberation in virtual acoustic environments.

Keywords

hearing aid evaluation, psychoacoustics, interaural coherence, room acoustics simulation, behind-the-ear devices

Received 8 April 2021; Revised 14 September 2021; accepted 5 October 2021

Introduction

Room acoustics simulation enables the well-controllable and repeatable presentation of reverberant acoustic environments, whether they exist in reality or just in the virtual domain. While artificial reverberation has historically been used as a tool for artistic expression (Välimäki et al., 2012), room acoustics simulations are also utilized for planning of spaces (e.g., concert halls and classrooms; Rindel, 2001). Recently, room acoustics simulations and virtual acoustic environments (VAEs) have gained interest as tools for psychoacoustic research and hearing aid development (Ahrens et al., 2019; Grimm et al., 2019; Huisman et al., 2019; Pausch & Fels, 2020; Pausch & Fels, 2020; Rindel, 2001; Schutte et al., 2019; Seeber et al., 2010), particularly in the context of ecological validity, where the results of speech-in-noise and hearing test with simplified "synthetic" stimuli have been questioned with regard to real-world outcomes (Cord et al., 2007; Jerger, 2009; Keidser et al., 2020; Miles et al., 2020).

Loudspeaker-based rendering of VAEs enables psychoacoustic tests with free head movement as well as the operation of (hearing aid) microphone arrays within the simulated sound field. The use of VAEs for hearing research and development and evaluation of hearing devices requires knowledge about perceptual and technical limitations of the rendered sound field: For perception and binaural hearing aid algorithms (e.g., Dörbecker & Ernst, 1996; Kollmeier et al., 1993; Srinivasan, 2008), the typical ear distance of

[1]Medizinische Physik and Cluster of Excellence Hearing4All, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
[2]Akustik and Cluster of Excellence Hearing4All, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Corresponding Author:
Christoph Kirsch, Carl von Ossietzky Universität Oldenburg, Fk.6 DMPA, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Germany.
Email: christoph.kirsch@uol.de

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

receivers is relevant in addition to the closer spacing of multi-microphone arrangements typically used to achieve spatial directivity in behind-the-ear (BTE) hearing devices (e.g., Doclo et al., 2015; Thiemann et al., 2016; Yousefian & Loizou, 2011).

Given that acoustic communication often takes place in enclosed spaces, the simulation and rendering of late reverberation and sound reflections at boundaries of these spaces are important for VAEs. While the direct sound and distinct early reflections are typically rendered as individual sound sources (e.g., Schröder, 2011; Wendt et al., 2014), diffuse late reverberation results from a superposition of many densely spaced reflections that are spatially more or less evenly distributed. In rooms, diffuse reverberation typically dominates the after the third reflection order (Kuttruff, 1995). Depending on room geometry and the spatial distribution of sound absorption at room boundaries, the resulting late reverberation can be considered approximately spherically isotropic (Hodgson, 1996) or anisotropic when containing limited spatial directivity (Alary et al., 2019a; Lachenmayr et al., 2016; Luizard et al., 2015; Romblom et al., 2016). For simplicity, cases with a spatially highly uniform incidence of late reverberation are assumed to approximate properties of a diffuse sound field and are referred to as isotropic here, although the reverberant sound field does not become isotropic in a strict sense (Polack, 1993; Jeong, 2016; Nolan et al., 2018).

In principle, the assumption of isotropy or anisotropy with limited spatial directivity of late reverberation allows for a reduced spatial resolution of the late reverberation rendering, relevant for computational efficiency in interactive, low-latency real-time VAEs. Considering hearing research and hearing devices evaluation in VAEs, the minimum required spatial resolution of late reverberation, depending on the number of (virtual) sound sources or loudspeakers, is of interest from both a perceptual and technical perspective.

Perceptually, localization of distinct sound sources or specular reflections in space is related to interaural level differences (ILDs) at higher frequencies above 1 kHz (see, e.g., Moore, 2003), and interaural time differences ITDs for lower frequencies with sensitivity sharply decreasing at about 1.5 kHz (e.g., Klumpp & Eady, 1956). For diffuse late reverberation, incoherent ear signals are observed and the auditory spatial impression involves a combination of several binaural cues like the lateral distribution of energy impinging on the listener in different frequency bands and interaural coherence (see, e.g., Blauert and Lindemann, 1986; Hidaka et al., 1995) in static or dynamic scenarios (e.g., Pöntynen et al., 2016). In anisotropic late reverberation, ILDs can occur, caused by more reverberant energy impinging from certain directions, while (long term) ILDs cannot occur in isotropic conditions. Additionally, fluctuations of ILDs and ITDs play a role in the in the perception of interaural incoherence (Goupell & Hartmann, 2006, 2007) for narrow-band signals (for an overview, see also Pulkki & Karjalainen, 2015). When considering applications in VAEs, dynamic cues introduced by head movement may also be relevant.

Overall, coherence at ear distance is considered relevant for psychoacoustic processes (e.g., Faller & Merimaa, 2004; Grosse et al., 2015) and has been suggested to assess the reproduction of diffuse sound fields (e.g., Hiyama et al., 2002; Walther & Faller, 2011). Hereby the auditory system is relatively insensitive to discriminate changes in coherence close to zero (Culling et al., 2001; Robinson & Jeffress, 1962; Walther & Faller, 2013). The spatial coherence between omnidirectional microphones, often used as a technical measure, is known to differ from that at the ear canal entrance (Lindevald & Benade, 1986) and thus does not represent exact psychoacoustic cues: Still, it can be assumed that a correct reproduction of coherence may be a necessary condition for the accurate reproduction of psychoacoustic cues.

Regarding the number of loudspeakers required to perceptually approximate the spatial impression of a diffuse sound field, Hiyama et al. (2002) found that a specific horizontal arrangement of four loudspeakers separated by 90° can already be sufficient. With such a low number of loudspeakers, however, the results were strongly dependent on the rotation of the loudspeaker array with regard to the listener. In VAE applications, where listeners can freely rotate their head, such rotation dependency is problematic. Regarding the perceptual impression of diffuse sound in three-dimensional arrangements ITU-R, BS.2159-8 (2019) recommends two vertically offset rings consisting of eight loudspeakers each. Based on informal listening tests with binaural reproduction, Laitinen and Pulkki (2009) reported a number of 12–20 virtual loudspeakers to be adequate for the reproduction of diffuse sound.

Grimm et al. (2015) evaluated different spatial loudspeaker reproduction methods in a simulated two-dimensional circular loudspeakers array. It was found that eight loudspeakers are sufficient for localization of a sound source in the horizontal plane according to an auditory model (Dietz et al., 2011) that estimates ILDs and ITDs. Furthermore, it was shown that a number of 8 up to 72 loudspeakers was required for specific hearing aid algorithms using microphone arrays in diffuse background noise.

Oreinos and Buchholz (2014) found similar signal-to-noise-ratio (SNR) benefits for an adaptive, binaural, correlation-based beamformer in a comparison between a real environment and an auralization via a 41 loudspeaker spherical array, driven by simulation and mixed-order ambisonics (Favrot et al., 2011) recordings. In Oreinos and Buchholz (2016), they additionally observed similar performance in the VAE compared to the real environment for speech intelligibility and acceptable noise level for hearing-aid equipped listeners.

Taken together, these studies provide recommendations for the rendering of diffuse sounds and demonstrate the general function of hearing aid algorithms in VAEs.

However, further systematic investigation of the minimal spatial resolution required for rendering of late reverberation in the context of VAEs is required, given that: (i) In contrast to earlier studies focusing on approximated isotropic sound fields, in VAEs late reverberation occurs in conjunction with direct sound and early reflections. (ii) Earlier studies often focused on cylindrically isotropic (two-dimensional) sound fields with horizontal loudspeaker arrangements (Grimm et al., 2015; Hiyama et al., 2002) in contrast to spherically (three-dimensional) isotropic or anisotropic sound fields occurring in the late reverberation. (iii) Rotation of the listener in the loudspeaker array has been shown to be critical for a very low number of loudspeakers (Hiyama et al., 2002). (iv) A connection of technical limitations of the reproduced sound field in a loudspeaker array and perception may help to guideline their design.

In the current study, spatial resolution of late reverberation was assessed in different simulated rooms by varying the number of spatially evenly distributed virtual reverberation sources (VRSs) rendered in a three-dimensional 86-channel loudspeaker array. For the perceptual evaluation, two psychoacoustic experiments with normal-hearing listeners were conducted using speech and transient stimuli: (1) The first experiment used VAEs with homogenous boundary conditions, resulting in approximately isotropic late reverberation. (2) For the second experiment, inhomogeneous sound absorption properties were assigned to the boundaries of the virtual room in order to create an anisotropic late reverberant field. (3) In a technical evaluation, the ability to reproduce the coherence between omnidirectional receivers at typical ear and BTE microphone distance in an isotropic sound field (e.g., Cook et al., 1955) in the loudspeaker array was investigated and related to the psychoacoustic findings. Limitations imposed by number of VRSs and array geometry are investigated by comparison of simulated sound fields to the analytic reference coherence function for the ideal spherically isotropic sound field. Additionally, for an example test case representing a real-world environment with long reverberation time and containing direct sound and early reflections, measured binaural room impulse responses (BRIRs) and simulated BRIRs with different spatial resolution of the late reverberation were compared in terms of coherence between the microphones of a dummy head and channels of typical BTE hearing aid devices.

The method used for simulation and rendering of the VAEs is freely available at www.razrengine.com.

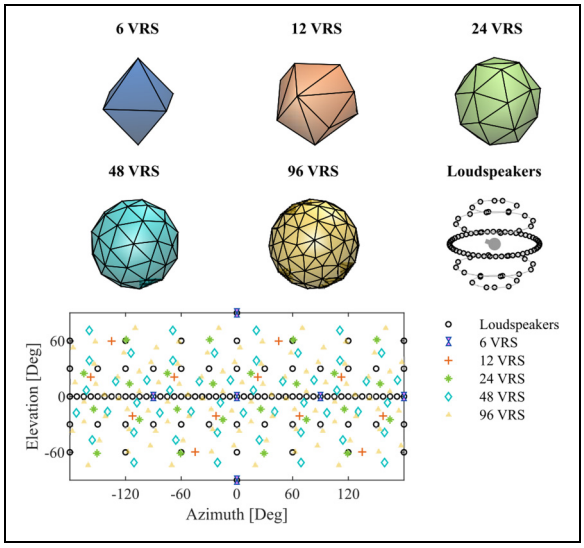

Figure 1. Number and spatial distribution of virtual reverberation sources (VRSs) and arrangement of the 3D 86-channel loudspeaker array used for playback. The vertices of the polyhedra at the top indicate the position of the VRSs. The right-hand plot in the second row shows a 3D projection of the loudspeaker positions, where the gray lines aid the visual representation of the loudspeaker arrangement in five rings. Polyhedra and loudspeaker array depiction are perspectively aligned. The bottom panel shows the directions of the loudspeakers (indicated as circles) and the VRS (other symbols, depending on number of VRSs) in a 2D projection of the spherical coordinates relative to the default listener orientation as indicated by the nosed ball in the 3D projection of the loudspeaker positions.

Methods

Room Acoustics Simulation

To vary the spatial resolution of late reverberation, the room acoustics simulator (RAZR; Wendt et al., 2014) was used. RAZR generates perceptually plausible room impulse responses (RIRs) using a geometrical acoustics-based image source model (ISM; Allen & Berkley, 1979; Borish, 1984), a shoebox approximation of room geometry and a computationally efficient feedback delay network (FDN; Jot & Chaigne, 1991) for late reverberation. Perceptual plausibility of the resulting room acoustics simulations was demonstrated by favorable performance in comparison to other state-of-the-art approaches in (Brinkmann et al., 2019, see their Figure 8). Early (specular) reflections up to the third reflection order were generated with the ISM, including effects of wall-dependent acoustic absorption as well as air absorption accounted for by a perceptually motivated low-pass filter, depending on the distance between image source and receiver. The late reverberation tail was generated by a 96-channel FDN which is fed by the last order of reflections from the ISM. The FDN parameters were derived to account for the (measured) frequency-dependent reverberation time. The output channels of the FDN were used as VRSs that were spatially evenly distributed around the listener (see Wendt et al., 2014 for details).

To adjust the spatial resolution of late reverberation, pairs of FDN output channels were added together, resulting in "downmixes" for a reduced number of 48, 24, 12, and 6 VRSs in addition to 96 VRSs. Average correlation coefficients

< 0.03 between the VRS signals enable the approximation of isotropic late reverberation as superposition of spatially distributed incoherent sound sources (see also Jacobsen & Roisin, 2000). By using the fixed, high number of 96 channels in the FDN, independent of the number of VRSs, the spatial resolution of the late reverberation can be adjusted while maintaining the spectro-temporal characteristics and thus avoiding timbre changes in the resulting reverberant tail (see, e.g., Schlecht & Habets, 2015, 2017). Polyhedra centered around the listener were used to determine the directions for spatialization of the VRSs, where the number of vertices corresponded to the number of VRSs (see Figure 1). The polyhedra were directionally aligned with the (shoebox) room boundaries and were optimized for geometrical sphericity (Wadell, 1935), quantifying to what extent the shape of the polyhedra formed by the VRS directions resembles that of a perfect sphere (for which sphericity equals 1, while it is 0.81 for a cube). Sphericity values of the current polyhedra ranged from 0.86 for 6 VRSs to 0.99 for 96 VRSs. For 6 VRSs, the resulting directions are orthogonal to each other and for 12 VRSs, they correspond to points lying on the diagonals of a room aligned cube. Directions for VRS numbers of 24 and above were derived from a combination of one, two, and four snub cuboctahedra (snub cubes) resulting in 4, 8 and 16 VRSs assigned to each of the six surfaces of the room. The maximum number of 96 VRSs are assumed to be a sufficiently high number to serve as a reference condition for diffuse late reverberation.

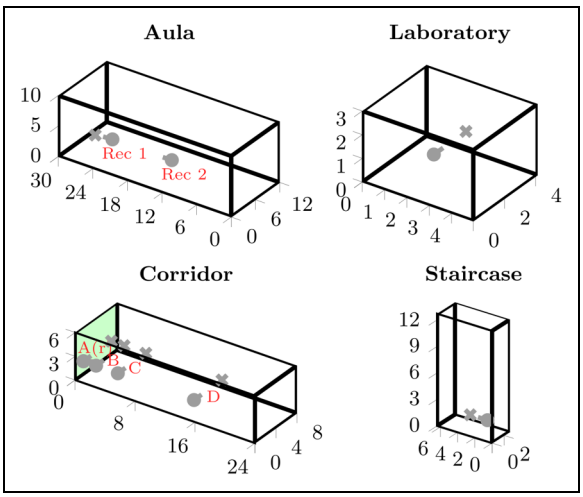

Figure 2. Overview of the virtual rooms. Source positions are denoted by crosses, receiver positions and orientation by the nosed balls. The rooms are shown at a different scale as indicated by the dimensioning provided in m. The Aula and Laboratory (top row) were used in experiment 1, the corridor (bottom left) was used in experiment 2, and the staircase (bottom right) was used for the comparison with measured binaural room impulse responses (BRIRs). The highly absorbing surface in the corridor (bottom left) is shaded.

In the case of a homogeneous distribution of acoustic absorption on all surfaces, resulting in approximately isotropic late reverberation, all VRSs radiate with the same power. To render anisotropic late reverberation in case of inhomogeneous distribution of acoustic absorption, the output of each VRS is scaled to represent the mean acoustic absorption for the solid angle covered by the VRS (see Poppitz et al., 2018 for more details). Thus, with a reduced number of VRSs, the spatial sampling of anisotropic features of the late reverberation is reduced.

Discrete early reflections and the VRSs were rendered using an 86-channel loudspeaker array and vector base amplitude panning (VBAP, Pulkki, 1997). VBAP is a three-dimensional panning technique that utilizes the closest three loudspeakers to render virtual sound sources in between the loudspeakers. The simulated RIRs that have been rendered for the loudspeaker array are referred to as multichannel room impulse responses in the following. Due to path differences between the loudspeakers participating in the rendering of a virtual sound source and the listener's ears, spectral coloration artefacts can occur when using VBAP. These were accounted for with filtering based on a statistical approach described by Laitinen et al. (2014).

Virtual Rooms

A selection of four virtual, simulated rooms (denoted as aula, laboratory, corridor, and staircase) was used with multiple source–receiver combinations in some of the rooms. An overview is provided in Figure 2.

For experiment 1, two source–receiver combinations were set in a 12 m × 10 m × 30 m room (aula, Figure 2, top left) with a reverberation time RT60 ranging from 7.2 s at 125 Hz to 1.5 s at 8 kHz. The virtual sound source was spaced either 2.72 m from the listener position (Rec 1) or 13.01 m (Rec 2). A third condition was set in a 4.97 m × 4.12 m × 3 m room (laboratory, Figure 2, top right) with a RT60 of 0.4 s in the 125 Hz–8 kHz range and a source–receiver distance of 1.7 m. In all three conditions, the receivers were oriented towards the source and all room boundaries had identical frequency-dependent absorption coefficients, resulting in approximately isotropic late reverberation.

For experiment 2, anisotropic late reverberation was assessed using a corridor (Figure 1, bottom left) with the dimensions 24 m × 8 m × 6 m and inhomogeneous absorption coefficients: One of the small surfaces (shaded) at the end of the corridor was highly absorbent ($\alpha = 0.99$ for all frequencies), while all other surfaces were quite reflective ($\alpha =$ 0.01–0.11 from 125 Hz to 8 kHz). The resulting RT60 ranged from ∼1.3 to ∼0.8 s in the frequency range from 125 Hz to 8 kHz. Several source–receiver–position combinations were considered to vary the solid angle (or field of view, FOV) occupied by the highly absorbing wall and thus the spatial features of the resulting anisotropic late reverberation. Source and receiver were always aligned on an axis parallel

to the highly absorbent surface, and were located at a height of 1.8 m above the floor and 1.33 m from the sidewalls, resulting in a fixed source–receiver distance of 5.33 m. There were four source–receiver combinations at different distances h to the highly absorbing wall, so that a wide range of solid angles with a horizontal FOV γ of the absorbent wall is obtained. The exact distances and FOV angles are denoted in Table 1. The receiver was always oriented towards the sound source (azimuth angle β was 0°) for all combinations (denoted A to D, see also Figure 2) except for Ar, where the receiver was rotated by 60° towards the direction of the corridor.

As part of the technical evaluation, a highly reverberant room (staircase, Figure 1, bottom-right) with dimensions 2.98 m × 6.83 m × 12.71 m was used, where most of the surfaces are painted concrete. In this room, BRIRs have been measured (see Apparatus and Procedure) to obtain realistic reference coherence function estimates. The source and receiver were located at the ground floor at a height of 1.72 m, spaced 2 m apart. The measured and simulated reverberation time in the staircase ranged from 5.3 s at 125 Hz to 2.6 s at 8 kHz.

Listeners

For experiment 1, nine self-reported normal hearing listeners (two female, seven male) aged between 26 and 32 years old were recruited. For experiment 2, 10 listeners (five female, five male) were recruited amongst students at the University of Oldenburg. They were between 21 and 28 years old at the time of testing. All were tested for normal hearing by means of pure tone audiometry (hearing threshold ≤ 25 dBHL between 125 Hz and 8 kHz). All subjects reported varying amounts of experience with previous listening tests.

Stimuli

Three types of stimuli were used: (i) A deterministic transient stimulus (referred to as pulse or pink pulse) of 500 ms duration was generated (sampling rate 44.1 kHz, decay from 0 dBFS to −60 dBFS in 36 ms) in the frequency domain by defining a pink spectrum between 50 Hz and the Nyquist frequency with minimum phase and transformation to the time domain. To reject temporal aliasing artifacts in the resulting periodic impulse, only the first half the impulse was cut and a raised-cosine off ramp was applied to the second half of that cut. This transient signal provides the listener with an opportunity to listen to the decay of the reverberation tail and is well suited for subjectively rating room acoustic qualities of a space (similar to using hand clapping). Given its deterministic nature, it is assumed that the pink pulse offers a high sensitivity to any change in the rendering of the RIRs. (ii) A semi-transient pink noise burst. The bursts were generated from 50 ms of white Gaussian noise that has subsequently been filtered by using the pink pulse as FIR coefficients to achieve a pink spectrum. A different noise token was generated for each burst to avoid identical (deterministic) source signals. As for the pink pulse, the relatively short pink burst resulted in a clearly audible decay, however, with random variation in the coloration, minimizing the availability of coloration cues for the discrimination task. (iii) A continuous speech stimulus composed of sentences from the Oldenburger Satztest (OLSA) corpus with different talkers (Hochmuth et al., 2015; Kuehnel et al., 1999). Different sentences were randomly selected to avoid deterministic stimuli and the availability of coloration cues, leaving subjects with only spatial cues. This condition is referred to as speech (mixed).

Table 1. Source–Receiver Geometry in the Corridor.

Position	h (m)	γ (deg)	β (deg)
A	0.18	170	0
Ar	0.18	170	60
B	1.87	110	0
C	4.77	70	0
D	14.93	30	0

h denotes the distance of the source–receiver pair from the highly absorbing surface (see Figure 1), γ describes the FOV angle and β corresponds to the azimuthal listener orientation in relation to the sound source with positive angles indicating a clockwise rotation.

All stimuli were convolved with multichannel room impulse responses for a specific source-receiver combination in a given virtual room. Five versions were generated, deviating in the number of VRSs (6–96), but being identical in terms of early reflections, where the same 62 image sources were always rendered in addition to the same direct sound.

In experiment 1, the pink pulse and speech stimulus were used. The reverberated pulse resulted in signals of about 2 s (including a 50 ms fadeout) and 860 ms in duration for the aula and laboratory rooms, respectively. For the laboratory, a pause interval of 500 ms was added between consecutive stimuli. For the speech (mixed) stimulus, different sentences were randomly selected out of a subset of six sentences from two different talkers (one male, one female). The overall duration of the stimuli was different for each of the sentences and ranged from about 3.7 s to 4.5 s including 2 s of the reverberant tail before being faded out. No extra silence interval separated the presentation of consecutive stimuli. It should be noted that for the aula with a reverberation of 7.2 s at 125 Hz, reverberant energy decayed by only about 17 dB during the observed 2 s of the reverberant tail. However, based on preliminary tests, longer stimulus durations were not feasible with regard to the overall duration of the experiment as well as for the perceptual comparison. Moreover, it can be expected that the main features of the reverberant decay are also covered in the low frequency region in the perceptually prominent part up to 20 dB decay. Furthermore, for frequencies around 500 Hz already

at least 30 dB decay were covered and more than 40 dB decay for 1 kHz and above.

In experiment 2, the pink pulse, the pink (noise) bursts, and the speech stimulus were used. For the reverberated transient stimuli, the signal duration was restricted to 1.2 s including a 100 ms fadeout and followed by a 300 ms silence (pause) interval. For the speech (mixed) stimulus, four different talkers (two male, two female) were used and different sentences were randomly selected out of a set of 100 sentences from each talker. In addition to the mixed condition, a single sentence was used repeatedly, referred to as speech (identical). Circular convolution was applied in order to mimic reverberation from previous sentences during an ongoing conversation. A silence interval of 100 ms was added before and after each sentence prior to circular convolution with the multichannel room impulse response. After convolution, a fade in and fade out over the duration of the silence interval was applied. The overall duration of the stimuli was different for each of the sentences and ranged from ~1.7 s to ~3 s. A 300-ms silence (pause) interval separated the presentation of consecutive stimuli.

For the pulse and burst conditions, the A-weighted, impulse-weighted sound pressure level at the listening position over the duration of 4.5 s containing three stimuli (followed by decay and silence) was ~65 $dB_{A,I}$ SPL. For the speech conditions, the equivalent A-weighted sound pressure level at the listening position over the duration of a presentation of three sentences (about 7.5 s) was ~65 dB_{Aeq} SPL. All signals were digitally generated and processed at a sampling rate of 44.1 kHz.

Apparatus and Procedure

The listeners were seated on a fixed (non-rotating) chair in the center of an 86-channel 3D loudspeaker array (Genelec 8030 c/b) mounted in a 7 m × 9 m × 7 m anechoic chamber with 0.75 m foam wedges. The loudspeaker array (see Figure 1) is approximately spherical with a radius of the main (horizontal) loudspeaker ring of about 2.5 m. The loudspeakers are inhomogeneously arranged in five rings at −60°, −30 °, 0°, 30°, and 60° elevation and two additional loudspeakers below and above the center point (−90°, 90° elevation). The azimuthal spacing of loudspeakers is 7.5° in the horizontal ring and 30° and 60°, respectively, in the rings outside of the horizontal plane.

A computer monitor was placed straight ahead of the listeners in a distance of 2.5 m in order to inform them about the progress of the experiment and provide a direction to point their gaze at. The listeners' head movement was neither constrained nor monitored, enabling natural head movements. Participants used a wireless keyboard on their lap so that responses could be given without looking at the controls.

For the listening tests, an ABX paradigm was used. A reference rendering with maximum number of VRSs (96) and the rendering under test (with a lower number of VRSs) were presented randomly as either A and B of the sequence. X was randomly chosen to be either the reference or the rendering under test. Test subjects had to determine, whether X was perceived to be similar to A or B in terms of spatial properties. Depending on the condition, the source material (prior to application of the multichannel room impulse response) was not necessarily the same for A, B, and X: For the pink pulse, source material was identical throughout the entire measurement and for speech (identical) throughout each trial, while it differed for pink burst and speech (mixed). The latter two stimuli forced listeners to solely rely on binaural spatial cues, while potential other cues like coloration changes became unreliable.

In both experiments, the procedure was separated into runs that covered a particular combination of room condition and stimulus. Subjects were given the opportunity to take short breaks in between runs. Experiments 1 and 2 consisted of six and nine runs, respectively. A total of 20 presentations per number of VRSs resulted in 80 ABX trials per run. Only exception to this was the speech (mixed) condition in experiment 2, where only 6 VRSs versus the reference have been tested, resulting in 20 ABX trials per run. One experimental run took about 10 min (pulse, burst) or 15 min (speech) to complete. Subjects were not provided feedback on the correctness of their responses. Additionally, audiometry prior to the listening test and a familiarization phase were added for experiment 2. The latter was comprised of the presentation of four ABX trials for a pink pulse, a pink burst with 6 and 12 VRSs, and four trials with mixed speech and 6 VRSs. During the familiarization phase, subjects received feedback on the correctness of their responses. For experiment 2, the experimental procedure, except for the audiometry, was repeated in an additional session on a different day resulting in overall 40 presentations per VRS number.

Technical Evaluation

The focus of the technical evaluation was on the ability to reproduce (i) interaural (head-size spaced) signal properties relevant for binaural perception and binaural hearing aid algorithms, and additionally on (ii) closely spaced inter-microphone signal properties relevant for multi-microphone hearing aid signal processing in an approximated spherically isotropic sound field given the number and positions of the VRS. The technical evaluation focused on the spherically isotropic case for which a clearly defined target coherence function (reference) exists.

Coherence was used as technical measure to assess the quality of the reproduction of an isotropic sound field, generated by independent Gaussian noises as output of the VRS. All frequency-dependent coherence estimates $C_{xy}(f)$ between the signals x and y have been calculated according to the following

equation:

$$C_{xy}(f) = \Re\left(\frac{G_{xy}(f)}{\sqrt{G_{xx}(f)G_{yy}(f)}}\right), \quad (1)$$

where f denotes the frequency, $\Re$ is the real-part operator, and G represents the spectral density estimate according to Welch. The calculations were performed for consecutive windows with a 75% overlap and a length of 512 samples at 96 kHz sampling rate to obtain an average coherence estimate. The Gaussian noise VRS signals were 30 s in duration.

The coherence was estimated for omnidirectional receivers, omitting the effect of the listener's head in favor of a clearly defined reference. The (reference) coherence $\boldsymbol{\Gamma}$ between two omnidirectional receivers spaced with the distance d in an ideal spherically isotropic sound field can, according to, for example, Cook et al. (1955), can be described as follows:

$$\Gamma_{sph}(\omega) = \mathrm{sinc}\left(\frac{\omega d}{c}\right), \quad (2)$$

where sinc is the unnormalized sinc function, ω is the angular frequency and c is the speed of sound.

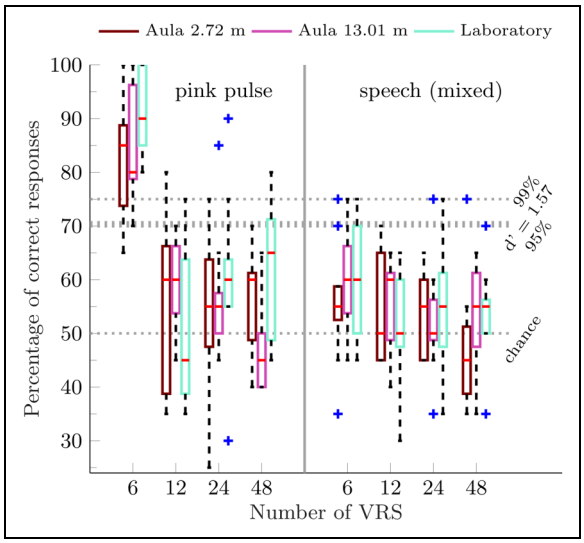

Figure 3. Percentage of correct responses for conditions with isotropic late reverberation. Boxes represent the 25%–75% percentile range and the horizontal red lines within the boxes represent the median. Whiskers denote highest or lowest value within 1.5 times the interquartile range from the box edge. Any values outside of this range were classified as outliers, represented by a blue cross. The different virtual room conditions are indicated on the top and by different colors of the boxes. Along the x-axis, the number of virtual reverberation sources (VRSs) is denoted. The three horizontal lines indicate the chance level and the 95% and 99% confidence range of a binomial distribution with $p = .5$. The d'-value of 1.57 almost overlaps with the 95% confidence range here.

In the simulation, two receivers were symmetrically positioned near the center of the array on an axis orthogonal to the 0° azimuth and elevation directions as defined in Figure 1. The spacing between the receivers for the simulations was 170 mm, approximating ear distance and the width of a human head, and 15.6 mm, corresponding to the distance between the front and rear microphones of the BTE hearing aids used in the dummy head recordings of the staircase room.

Idealized IRs have been derived for the paths from sound sources to the receivers for three arrangements: (i) for the current 86-channel array to the receivers, (ii) for a different array geometry with a similar number of loudspeakers arranged in a homogenous spacing across the spherical surface according to a Fibonacci lattice (González, 2009), and (iii) for the VRS positions, representing a direct spatialization with loudspeakers at the VRS positions. In all three cases, a sphere diameter of 2.5 m was assumed. The effect of air absorption was neglected, meaning that each idealized IR consisted of a delay and attenuation resulting from the distance between the sound source and the receiver. Receiver signals have been generated by convolving independent Gaussian noises with the IRs and summing the resulting signals for both receivers.

In addition to the systematic coherence assessment for the spherically isotropic sound field, the staircase room served as an example scenario to assess coherence in the VAE with simulated BRIRs using different numbers of VRSs in comparison to the measured BRIRs in a real room. To generate the simulated BRIRs, a set of head-related impulse responses (HRIRs) has been measured for all loudspeakers in the 86-channel array (sampling rate 44.1 kHz) using a G.R.A.S. KEMAR type 45BM head and torso simulator equipped with BTE hearing aid microphone dummies (see Kayser et al., 2009). Each hearing aid dummy had three microphones, resulting in a total of eight measured channels (eardrum, BTE front, mid and rear, for each the left and right side of the head). For each loudspeaker in the array, the respective channel of the simulated multichannel room impulse response was convolved with the according HRIRs for the two eardrum microphones and the BTE front and rear channels (right side) and summed over all loudspeakers resulting in simulated BRIRs. These simulated BRIRs allow for a direct comparison of the rendering in the loudspeaker array to the according real-world BRIRs which were measured using the same equipment on the ground floor of the staircase room (see Figure 2, bottom right).

Perceptual Evaluation

Experiment 1: Isotropic Late Reverberation

Figure 3 shows the percentage of correct responses (ability to discriminate the auralization with reduced number of VRSs from the reference) as a function of VRSs for the pink

pulse stimuli in the left section and for the speech (mixed) stimuli in the right section. The boxplots indicate the median and interquartile ranges across listeners for all three room conditions (Aula Rec 1, Aula Rec 2, Laboratory, from left to right), grouped by the number of VRSs used for rendering. Crosses indicate outliers, defined as being more than 1.5 interquartile ranges outside of the actual interquartile range. The three dashed horizontal lines correspond to chance level (50% correct) and to the percentage of correct responses required to reject the hypothesis, that responses are drawn from a 0.5 chance binomial random variable at a confidence level of 95% or 99%, as indicated on the right-hand side. These lines serve as an orientation to estimate whether the respective scores are achieved by guessing. Additionally, to assess discriminability, the percentage correct (70.7%) according to a d'-value of 1.57 for the current ABX/BAX experiment is indicated according to Macmillan and Creelman (2004; Table A5.3.) based on the assumption of subjects employing an independent observations strategy. If a differencing strategy is assumed as argued by Hautus and Meng (2002), the d'-value increases to 1.76.

For the pink pulse and 6 VRSs (left-hand side), there is a high proportion of correct responses regardless of the virtual room configuration (median > = 80%). For higher number of VRSs and the pink pulse, the median amount of correct responses is close to chance (50% correct), even though there are outliers with up to 90% of correct responses. There are no obvious tendencies with regard to the effect of the virtual room or the further increase in the number of VRSs. For the speech (mixed) condition (right-hand section in Figure 3), even for 6 VRSs, the median correct responses range from only 55% to 60% and are otherwise also close to chance as observed for the pink pulse.

A three-way repeated measures analysis of variance (ANOVA) showed a significant main effect of the number of VRSs ($F[3, 24] = 42.4$, $p < .001$) and of the stimuli ($F[1, 8] = 22.4$, $p < .01$), but no significant main effect of the room configuration ($F[2, 16] = 1.5$, $p = .26$). A significant interaction was found only for the number of VRSs and the stimuli ($F[3, 24] = 13.2$, $p < .01$].

Post-hoc pair-wise comparisons (Bonferroni) revealed that the main effect of the number of VRSs and the interaction of VRS and stimulus can be attributed to significant differences ($p < .001$) between 6 VRSs and all other VRS numbers for the pink pulse, while there are no significant differences in detection performance between the higher VRS numbers (12 – 48) for the pink pulse and no significant differences for any number of VRSs for speech (mixed).

In summary, for isotropic late reverberation, 12 VRSs are sufficient to perceptually approximate a spherically isotropic sound field with the transient pink pulse stimulus. For speech with additional uncertainty regarding the source signal, even 6 VRSs appear to suffice.

Experiment 2: Anisotropic Late Reverberation

Figure 4 shows the average discrimination results between conditions with fewer VRSs as indicated on the x-axis and the reference condition with 96 VRSs in a similar format as shown in Figure 3. The order of the boxes and colors indicate the room conditions with different distance from the absorbing wall (left to right: A–D, close to far; Ar, rotated) as indicated in the legend. The figure is separated into three sections by vertical gray lines for pink pulse, pink burst, and speech (identical, mixed) from left to right, grouped by the number of VRSs. For pink burst and speech, only subsets of the

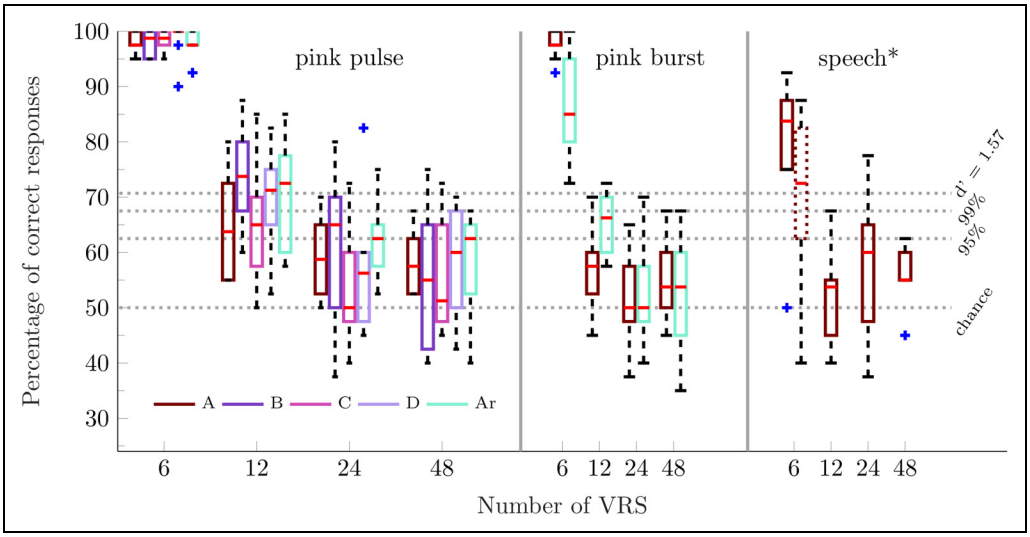

Figure 4. Percentage of correct responses for the different experimental conditions with anisotropic late reverberation, in the same style as shown in Figure 3. The order of the position inside the virtual corridor is indicated at the top and by the different shades of colors of the boxes. Along the x-axis, the number of virtual reverberation sources (VRSs) is denoted. Results are grouped by stimulus. For speech (right-hand subpanel), the dotted box shows the data measured with random speech (mixed).

conditions have been measured: Room conditions A and Ar for the pink burst and room condition A for the speech. For speech mixed (dotted box), only renderings with 6 VRSs were compared to the reference.

A fairly narrow distribution of subject performance can be observed with 6 VRSs for the pink pulse (left-hand section) on all positions and for the non-rotated (A) pink burst (center section). For these stimuli, all subjects responded correctly in at least 90% of the trials. For 12 VRSs with the pink pulse, there is a wide spread at all different positions with individual performances ranging between 50% and 87.5% correct responses. This generally wide performance range can also be observed for 24 and 48 VRSs, but with lower medians compared to 12 VRSs. A consistent dependency on the position within the room (A–D) is not immediately apparent.

For the pink burst (center section), a much wider performance range with a lower median can be observed for 6 VRSs for the rotated condition (Ar) compared to the non-rotated condition (A). While the performance median for the rotated condition is lower for 6 VRSs when compared to no rotation, it is higher relative to the non-rotated condition for 12 VRSs. Similar to the pink pulse, with the pink burst, the medians with 24 and 48 VRSs are slightly lower compared to 12 VRSs.

Opposed to the other source signals, speech-based stimuli (right-hand section) show a wide performance range even for 6 VRSs. This is especially apparent for the mixed speech stimuli (dotted box) where individual discrimination performance ranged from 40% to 90%. Here, subjects had to purely rely on spatial changes in the rendering. Renderings with 12 – 48 VRSs result in performance medians at and somewhat above 50%, with variability strongly reduced for 48 VRSs, where none of the subjects exceed 65% of correct responses.

Given that not all room conditions where measured with every stimulus, for statistical analysis, three separate ANOVAs were performed[1]:

For the pink pulse, the effect of VRS and position within the room was assessed with a two-way repeated-measure ANOVA. A significant main effect was found for the number of VRSs ($F[3, 27] = 250.26$, $p < .001$). No significant main effect was found for the position within the room ($F[4, 36] = 0.84$, $p = .51$), and no significant interaction was found ($F[12, 108] = 0.96$, $p = .46$). Post-hoc pair-wise comparisons (Bonferroni) revealed significant differences between the conditions with 6 VRSs compared to all other conditions. While for 12 VRSs, there were some significant differences compared to 24 VRSs at position C and to 48 VRSs at positions B and D, no significant difference was found between 24 and 48 VRSs.

For the pink pulse and burst, a three-way repeated-measures ANOVA was performed to assess the effect of the number of VRSs, stimulus type (pulse vs. burst) and rotation (position A vs. Ar). A significant main effect was found for the number of virtual sources ($F[3, 27] = 335.67$, $p < .001$) and for stimulus type ($F[1, 9] = 28.80$, $p < .001$). No main effect was found for rotation ($F[1, 9] = 12.66$, $p = .34$). No significant interactions were found for the number of VRSs and stimuli, the stimuli and rotation and the number of VRSs, stimuli and rotation ($p > .2$). However, there was an interaction between the number of VRSs and rotation ($F[3, 27] = 3.42$, $p = .03$).

To additionally assess the effect of stimulus type (pulse, burst, speech), a two-way repeated-measured ANOVA was performed for the number of VRSs and the different stimuli with all the results obtained at position A except for the additional measurement of speech (mixed). A significant main effect was found for the number of virtual sources ($F[3, 27] = 137.04$, $p < .001$) and for the stimulus type ($F[2, 18] = 12.29$, $p < .001$), as well as a significant interaction ($F[6, 54] = 5.12$, $p < .001$). Post-hoc pair-wise comparisons (Bonferroni) revealed a significant difference between the pink pulse and the other two stimuli, but no significant difference between the pink burst and speech (identical) condition.

Finally, a paired-samples t-test between the results obtained with 6 VRSs for speech (identical)- and speech (mixed)-stimuli showed a significant mean difference of 11.75 percentage points in score, with a standard error of 2.79, $p = .0023$.

Taken together, the results indicate that for the pink pulse stimulus 12–24 VRSs appear sufficient to render the late reverberation. If uncertainty is introduced in a transient stimulus (pink burst), already 12 VRSs are sufficient for the position closest to the wall (A). Similar results are observed for the speech (identical) stimulus. A number of 6 VRSs are distinguishable for all conditions, however, for speech (mixed) a clear decline of discrimination performance was observed, with some individuals even performing at chance level.

Discussion

Isotropic Late Reverberation. For isotropic late reverberation (experiment 1), the current finding of 12 VRSs as minimum number for rendering late reverberation support the default choice of 12 VRSs in the original RAZR implementation (Wendt et al., 2014). The result is also generally in line with earlier findings in the literature regarding the minimum number of loudspeakers required to achieve a perceptually diffuse sound field: Laitinen and Pulkki (2009) suggested 12–20 loudspeakers as sufficient for a three-dimensional arrangement. In Hiyama et al. (2002) for eight or more horizontally arranged loudspeakers, the contribution of additional loudspeakers was strongly decreased. The current results pointing to 12 VRSs were obtained for the pink pulse stimulus, which was determined to be the most critical to show effects of the number of VRSs in preliminary experiments. For this transient and deterministic stimulus, subtle differences in the reverberant decay process including spatial features but also spectral coloration

changes that might occur as an artefact of VBAP are easily revealed. With the random speech (mixed) stimulus, median percentage-correct values of only 55%–60% even for 6 VRSs were observed, and again no difference between 12 and 96 VRSs with median values around 50%–55%. Thus, depending on the stimulus and with a fixed relation between receiver orientation and VRS, even less than 12 VRSs might be sufficient for spherical (three-dimensional) isotropic cases.

Furthermore, it can be assumed that cylindrical (two-dimensional) isotropic sound fields, as occurring for late reverberation with reflections resulting mainly from vertical structures (e.g., in the case of a highly absorbent or absent ceiling) are less demanding and require at least eight horizontal loudspeakers extrapolating from 12 for the spherical isotropic case. This coincides with the findings of Grimm et al. (2015). However, it is conceivable that the anisotropic cases, as assessed in experiment 2, are more demanding, and that the number of VRSs also affects binaural parameters as

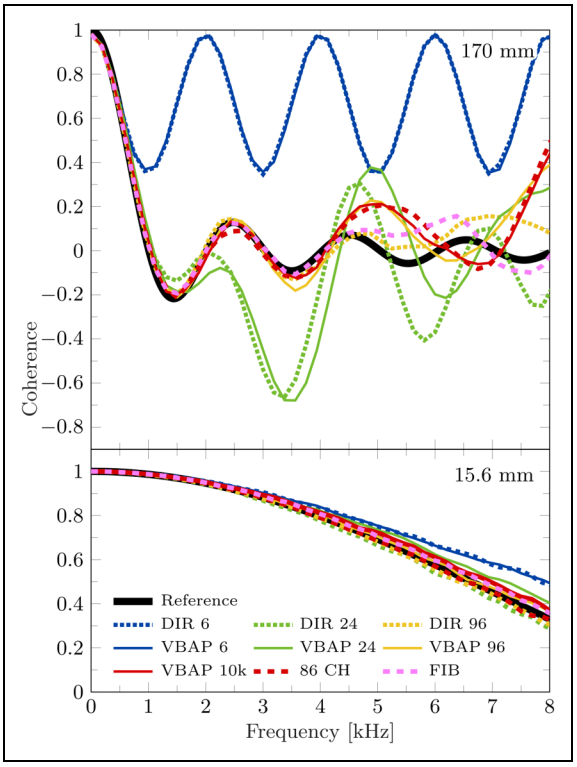

Figure 5. Coherence function estimates between two omnidirectional receivers at 170 mm (top panel) and 15.6 mm distance (bottom panel). The reference function for an ideal spherically isotropic sound field is shown as a thick solid line. The dashed lines show the coherence functions for the 86-channel loudspeaker array with weighted incoherent loudspeaker signals (86 CH) and for 87 incoherent sources distributed as a Fibonacci lattice (FIB). DIR: direct spatialization of the virtual reverberation sources (VRSs) (dotted traces), VBAP: spatialization of the VRSs via VBAP in the 86-channel loudspeaker array (solid traces). VBAP 10k refers to 9999 VRSs.

the interaural coherence and ILDs depending on the rotation of the listener in the loudspeaker array as discussed below in the context of experiment 2 and the technical evaluation.

Anisotropic Late Reverberation. For anisotropic late reverberation (experiment 2), reflecting a spatially more challenging condition, 6 VRSs could be clearly distinguished from the reference for all stimuli, with the exception of the speech (mixed)-stimulus. Here, different (random) sentences were presented for each signal in the ABX trial so that listeners had to rely solely on spatial cues. Similar to the isotropic case in experiment 1, uncertainties in the (dry) source signal resulted in considerable differences in individual discrimination performance. The results indicate that for unknown source material, or stimuli like running speech, demands for spatial resolution of late reverberation in VAEs can be lowered even for anisotropic late reverberation. Already a comparatively small number of VRS could result in a convincing rendering for the majority of listeners, which might be useful in applications where lowest computational requirements are key (e.g., spatial rendering in teleconference systems or hearing supportive algorithms in mobile applications).

A dependency of the number of required VRSs for a perceptually plausible rendering on the source material is suggested by the significant differences between the deterministic pink pulse and the other stimuli for 12 and 24 VRSs. As stated earlier, the deterministic pink pulse allows for a detailed comparison of spectral coloration differences in the decay process. For the random pink burst, a general decrease in detection performance is observed in non-rotated conditions with more than 6 VRSs compared to the pink pulse. The stochastic nature of the pink burst stimulus and its variability throughout the ABX trials likely reduced the availability of spectral artefacts as a discrimination cue.

Receiver Rotation. The effect of receiver orientation (A, Ar) in relation to the virtual sound source was assessed in experiment 2 for the pink pulse and burst stimuli. Based on the statistically significant interaction between the number of VRSs and the rotation (regardless of the non-significant main effect of receiver orientation), visual assessment of the distribution of responses allows for some speculation: The most obvious deviation occurs for 6 VRSs with the pink burst, where the rotated condition shows a considerably reduced median and a much wider spread in the distribution of responses, indicating that some listeners had difficulties to distinguish renderings in the rotated condition. The rotated condition differs from the un-rotated condition by a 60° azimuth incident angle of the virtual sound source and a listener rotated more towards the reverberant corridor. For 6 VRSs, the un-rotated condition results in a considerably larger interaural coherence compared to 96 VRSs, given that four VRSs are in the sagittal plane and produce no interaural differences and one VRS is to the right of the listener. The VRS

to the left is largely attenuated, representing the highly absorbent wall in the spatially subsampled rendering. For any higher number of VRSs, the interaural coherence is reduced, as more VRSs outside the sagittal plane introduce interaural differences. Likewise, for the rotated condition, two VRSs of the overall five active VRSs are moved out of the sagittal plane, reducing interaural coherence. It is likely that the interaural coherence changes for the rotated condition, explaining the reduced detection rates. A similar effect of rotation was observed in Hiyama et al. (2002) for rendering a diffuse sound field with only four horizontally arranged speakers. In their study, the 45° rotated array, with all loudspeakers introducing interaural differences, was preferred over the (un-rotated) arrangement with two loudspeakers (front, back) in the sagittal plane. The slight increase in detection likelihood for 12 VRSs with the rotated condition with pink bursts and the absence of such an increase for 24 and 48 VRSs might indicate that 12 VRSs overall just barely meet the requirements and a higher number of VRSs would be beneficial in some applications.

The effect of rotation on interaural coherence is also assessed in the following technical evaluation for the isotropic case.

Technical Evaluation

Isotropic Sound Field Reproduction

The top panel of Figure 5 shows coherence function estimates for two omnidirectional receivers spaced at 170 mm (ear distance) in the frequency range up to 8 kHz. The analytically derived reference function for an ideal spherically isotropic sound field is represented by a thick black solid line.

First, to separate the effect of loudspeaker array topology and VRS, the closest approximation of an isotropic sound field possible in the specific 86-channel loudspeaker array is shown as VBAP 10k (solid red) representing 9999 VRSs spatialized via VBAP, and 86 CHs (dashed red) representing 86 incoherent noises with relative power identical to the VBAP 10k case. For both, a very similar approximation of the reference function is observed up to about 4.3 kHz. For higher frequencies, deviations occur caused by the finite number of loudspeakers as well as by the inhomogeneous distribution of the loudspeakers on a sphere which will affect all VRS renderings. In comparison, for a spatially more homogeneous distribution of 87 loudspeakers on a sphere according to a Fibonacci lattice (FIB, dashed pink), smaller deviations occur for higher frequencies.

Second, to separate the effect of the number of VRSs and VBAP in the loudspeaker array, coherence function estimates for 6, 24, and 96 incoherent VRSs with VBAP-based spatialization (VBAP; solid lines) and without VBAP (DIR; dotted lines) are shown in Figure 5. The functions for DIR would be achieved if physical loudspeakers were present at the positions of the VRSs (see Figure 2) or in the case of binaural rendering to headphones using HRTFs for the respective VRS positions. For VBAP, multiple loudspeakers typically represent a single VRS, thus introducing partial coherence of the loudspeaker signals. For 6 VRSs, the VRS positions match those of the loudspeakers, resulting in identical periodic coherence function for VBAP and DIR for the 170 mm spacing in the top panel of Figure 5, beginning to diverge from the reference at about 700 Hz. For 24 VRSs, the VRS positions do not match those of the loudspeakers

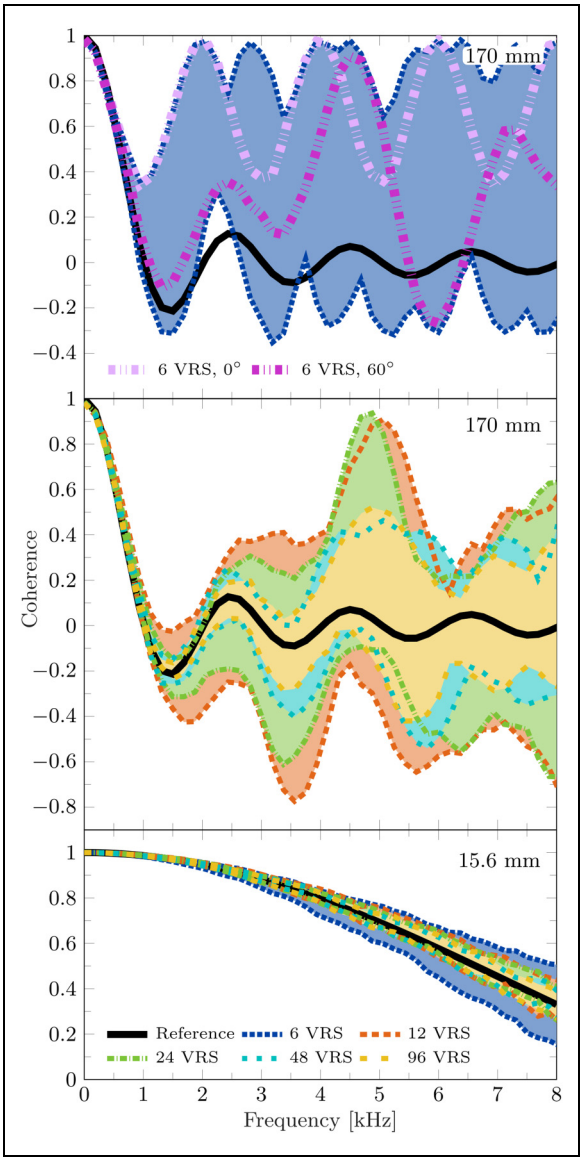

Figure 6. Coherence ranges for pairs of 170 mm (top and middle panel) and 15.6 mm (bottom panel) spaced omnidirectional receivers throughout a full azimuth rotation in relation to the 86 channel loudspeaker array for different number of virtual reverberation sources (VRSs). For an ideal reproduction of the isotropic sound field, the range would be infinitely narrow and follow the solid black reference curve. As an example, coherence functions for 6 VRSs at a receiver orientation of 0° and 60° are illustrated in the top panel.

and the VBAP and DIR coherence function estimates diverge at higher frequencies, following the reference up to a frequency of about 1.7 kHz. With 96 VRSs, 82 of the 86 loudspeakers are used for VBAP due to the inhomogeneous loudspeaker distribution and coherence starts diverging from the reference above about 4.3 kHz. Without VBAP (DIR 96, dotted), using 96 spatially uniformly distributed VRS optimized for sphericity, the reference is approximated more accurately even above 4.3 kHz.

For clarity, only the selected coherence function estimates for 6, 24, and 96 VRSs are shown in Figure 5. For all numbers of VRSs, assessment of the coherence curves for 170 mm receiver spacing yielded estimates of the upper frequency limit of correspondence (± 0.1) with the reference curve of 0.7, 1.4, 1.7, 3, 3.3, 4.3 kHz for 6, 12, 24, 48, 96, 9999 VRSs, respectively.

For the receivers at BTE-distance (15.6 mm spacing between front and rear, bottom panel of Figure 5), as expected from theory, the closer receiver spacing allows for a considerably better approximation of the reference coherence (± 0.1) up to at least 5 kHz even for 6 VRSs, and at least 8 kHz for higher number of VRSs. Here, the reference curve only decays to a coherence of about 0.4 at 8 kHz and the first zero crossing occurs at 11 kHz outside the depicted range in contrast to 1 kHz for the 170 mm case in the top panel of Figure 5.

Receiver Rotation

To investigate the effect of head rotations in VAEs, rotation in the horizontal plane ranging from 0° (as used before) to 358° in 2° steps was assessed for the 170 and 15.6 mm spaced receiver pairs for different numbers of VRSs. Figure 6 shows the maximum and minimum range of the coherence function estimates as shaded areas outlined by different line styles (see legend). In an ideal isotropic sound field, coherence would be independent of the rotation and follow the reference function (solid black). For the 170 mm spacing (upper and middle panel of Figure 6), coherence ranges are narrow and follow the reference curve independent of the number of VRSs below 700 Hz. Above 700 Hz, the coherence ranges diverge the more as the number of VRSs becomes smaller. Particularly for the lowest number of 6 VRSs (upper panel), a wide coherence range is already observed above 1 kHz. To further assess the variability of the coherence for 6 VRSs, two example curves for 0° (replotted from Figure 5) and 60°, as used in the psychoacoustic experiments (condition A, Ar) are shown. The upper frequency, for which the maximum deviation of coherence can be considered reasonably close (± 0.1) to the reference, is about 0.7 kHz for 6 VRSs, 1.1 kHz for 12 VRSs, 1.5 kHz for 24 VRSs, 2.3 kHz for 48 VRSs, and about 2.5 kHz for 96 VRSs.

For the receivers at BTE-distance (15.6 mm spacing, bottom panel of Figure 6) the effect of rotation is considerably smaller and all coherence function estimates for more than 6 VRSs are close to the reference function (± 0.1) up to 8 kHz and higher. For 6 VRSs, this limit is reached at about 5 kHz.

Comparison to Measured BRIRs

As an example of a typical VAE application, the interaural (between eardrums) and BTE coherence for measured BRIRs in the staircase are shown in Figure 7 (solid black line) in comparison to simulated BRIRs rendered via VBAP in the loudspeaker array. Since the BRIRs also contain direct sound and early reflections, the resulting signals are not a test case for the isolated reproduction of isotropic late reverberation, but better reflect conditions that might occur during a listening tests or hearing aid evaluation in a VAE. The limited set of 6, 12, and 24 VRSs was chosen for increased clarity in the figure and these numbers of VRSs are expected to show the largest effects based on the results so far.

Coherence functions between the two eardrum channels (binaural, upper panel) are conceptually comparable to those of the 170 mm spaced omnidirectional receivers in the evaluation of the isotropic late reverberation. Both simulated and measured curves show a similar behavior up to a

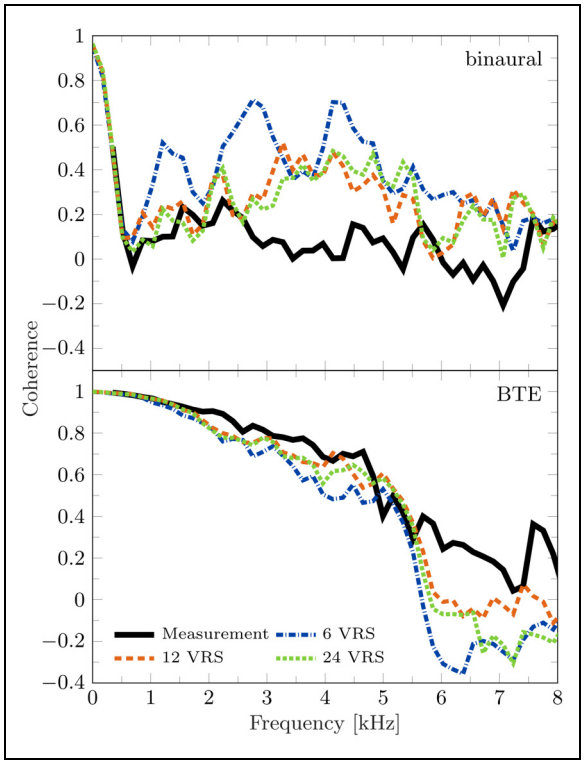

Figure 7. Coherence function estimates for simulated and measured (solid black line) binaural (top panel) and behind-the-ear (BTE) (bottom panel) impulse responses in the example staircase room.

frequency of about 700 Hz for 6, 12, and 24 VRSs. At higher frequencies, 6 VRSs result in a higher coherence than the measured reference, as can be expected from the reproduction of the isotropic diffuse field (compare to upper panel of Figure 5). The simulations utilizing 12 VRSs start to deviate from the reference at about 700 Hz while 24 VRSs follow the reference more closely up to above 1 kHz. At about 2.3 kHz, both simulations show a peak in the coherence which is not present in the measurement and then approximate the measurement again up to 2.6 kHz. Above 2.6 kHz, the simulated coherence for 12 and 24 VRSs is generally higher than in the measurement. The differences between 12 and 24 VRSs are comparatively small and no systematic behavior is apparent. For higher number of VRSs (not shown here) similar results to 24 VRSs are obtained.

For the BTE microphones (15.6 mm spacing, bottom panel of Figure 7), all coherence function estimates independent of VRS number and the measurement are very similar up to about 2 kHz. Up to 5 kHz, the simulations show a reasonably close approximation of the measured coherence, with increasing divergence from the measurement for higher frequencies. Larger errors are observed for 6 VRSs compared to 12 and 24 VRSs.

Discussion

Number of VRSs and array topology. A large effect of the number of VRSs and rotation, as well as of array topology on coherence was observed for the ear distance (170 mm) spacing: Independent of the concept of VRSs for rendering diffuse late reverberation, the inhomogeneous distribution of loudspeakers on the spherical surface of the investigated 86-channel loudspeaker array leads to a decrease in accuracy for the reproduction of spherical isotropic sound fields above 4 kHz when compared to the spatially more evenly distributed Fibonacci-lattice layout with a similar number of loudspeakers. Often, the reproduction of diffuse reverberation is, however, not the only goal. In many listening scenarios, auditory objects tend to be predominantly located in the horizontal plane, where an inhomogeneous layout can minimize rendering artefacts of the direct sound. Also, the reproduction of cylindrical isotropic sound fields (not shown here) benefits from the high spatial resolution in the horizontal plane. The comparison between renderings with and without VBAP shows a detrimental effect of the inhomogeneous array topology for larger numbers of VRSs for which the geometrical sphericity of the spatial VRS distribution is high. Nevertheless, any increase in the number of VRSs still results in an increase of the upper frequency limit for accurate reproduction for the different numbers of VRSs considered in this study.

Relation to Perception. In the perceptual findings, 12 or more VRSs were indistinguishable from the highest spatial resolution of the reverberation rendering for a majority of listeners. This coincides with an accurate reproduction of sound field coherence up to at least 1.5 kHz for the 170 mm spaced omnidirectional receivers and a reasonable approximation of coherence in that frequency range in the ear drum channels (staircase example), in line with the upper frequency limit for sensitivity to ITDs commonly found in literature (e.g., Brughera et al., 2013; Klumpp & Eady, 1956; Zwislocki & Feldman, 1956). Given that listeners are less sensitive to changes in coherence when values are close to zero (e.g., Robinson & Jeffress, 1962), it is likely that inaccuracies at higher frequencies, where coherence generally drops to low values in the simulations as well as in the reference case, are perceptually less relevant. The observed deviations in coherence below 1.5 kHz for more than 12 VRSs in the simulations considering rotation (Figure 7) as well as in the test-case scenario are well within the more recently derived narrow-band JNDs stated by Walther and Faller (2013). However, given that other cues besides interaural coherence might contribute to the auditory spatial impression, only limited conclusions can be drawn from the technical evaluation.

The strong fluctuations of coherence for rotation of the 170-mm-spaced receiver for low numbers of VRSs (see, e.g., 0° and 60° for 6 VRSs in the upper panel of Figure 7) likely played a role in the results of the perceptual evaluation for the rotated condition, although the isotropic case in the technical evaluation and the anisotropic case in the perceptual evaluation are not directly comparable. Thus, considering head rotations, the technical evaluation suggests that an increased number of VRSs might be beneficial when deviating from a static (optimal) alignment of head orientation and loudspeaker array. An additional parameter that may become relevant in practical applications is the relative orientation of the VRSs and the loudspeaker array. However, this parameter was kept fixed by exploiting rotational symmetry of the array geometry and hence kept out of the scope of this study.

Based on both perceptual and technical findings, 12 directions can thus be generally considered sufficient for the spatial rendering of late reverberation. In more critical conditions with strong inhomogeneity in late reverberation, 24 directions are required.

BTE Microphones and Hearing Aid Algorithms. The isotropic sound field reproduction for the BTE inter-microphone spacing of 15.6 mm showed a close approximation of the reference coherence up to at least about 5 kHz even for 6 VRSs. For the staircase test case an accurate reproduction of coherence up to about 4,500 Hz can be achieved with 12 or more VRSs. This covers the entire frequency range that is most commonly used for hearing aid signal processing. The performance that can be expected from any multimicrophone hearing aid signal processing algorithm would have to be evaluated individually, depending on the assumptions about the sound field (also beyond coherence) and error tolerance of that particular algorithm.

In line with the current findings, Grimm et al. (2015) showed that for a horizontal loudspeaker array, the function of a monaural adaptive differential microphone showed no further improvement for more than eight loudspeakers in the frequency range up to 4 kHz. For a binaural noise reduction algorithm, using the larger ear distance, at least 18 loudspeakers were required, in agreement with the higher number of VRSs required in the current study to approximate coherence for the ear distance. Considerably higher numbers of up to 72 loudspeakers were only required in Grimm et al. (2015) for off-center positions in the loudspeaker array, which were not considered here.

The current findings are also in general agreement with relatively faithful reproduction of interaural cross-correlation observed in a spherical 29-loudspeaker loudspeaker array by Cubick and Dau (2016), and the faithful performance of binaural hearing aids in a spherical 41-loudspeaker array as observed by Oreinos and Buchholz (2016) which is broadly comparable to the 48 VRSs conditions for the reproduction of an isotropic sound field. However, their ambisonics-based loudspeaker driving functions and conditions differ from the current study, making a more detailed comparison difficult.

Application to Real-Life Scenarios and Limitations. The simulation of the highly reverberant staircase example case showed a reasonably accurate reproduction of the interaural coherence for the perceptually relevant frequency range up to 1 – 1.5 kHz. The similarities in the observations for both the simulated isotropic sound field and the example case suggest that the proportion of energy in the late reverberation in comparison to the energy in the total impinging sound is large enough to play a significant role in the coherence of the receiver signals. With shorter reverberation times, this proportion is expected to be smaller and the relevance of spatial reproduction of late reverberation might be reduced. However, a relaxation of the rendering requirements could not be confirmed by the seemingly room-independent perceptual results for rooms with isotropic reverberation as presented in Experiment 1: Isotropic Late Reverberation.

While RAZR was used as specific implementation to find the parameters that determine the minimum number of VRSs, the current considerations are universally applicable to other implementations of late reverberation rendering, provided they generate incoherent VRS signals. Similarly, the loudspeaker array geometry that has been the basis for the investigations is not universal, but not atypical in terms of layout and number of loudspeakers and provides an orientation for applications in other arrays.

Conclusions

For simulating virtual acoustic environments, three-dimensional late reverberation is often approximated by a limited number of discrete directions, from which incoherent late reverberation signals are rendered (virtual reverberation sources). A low number of VRSs reduces computational complexity of the underlying simulation and rendering. Considering human perception and physical properties of the reproduced sound field in VAEs, the following conclusions can be drawn:

- Perceptual evaluation shows that the ability of subjects to discriminate a rendering with a lower number of VRSs from the reference condition depends on the source material. Deterministic transient stimuli reveal more differences than random speech tokens, which can be assumed to better represent applications using running speech.
- For approximately spherically isotropic late reverberation of transient and speech stimuli, presented in a setting with unconstrained head movements, 12 spatially evenly distributed VRS are indistinguishable from higher number of VRSs. Anisotropic reverberation requires 12–24 VRSs, for the most critical case using a deterministic pink pulse. In the technical analysis, these numbers coincide with sufficiently close approximation of coherence at ear distance up to 1.5 kHz or more.
- Variations in coherence at ear distance caused by rotation of the head relative to the VRSs are particularly critical for the lowest number of 6 VRSs assessed in the current study. Although such a low number might be acceptable for applications limited to running speech and isotropic reverberation, it can generally not be recommended for applications where the head can be freely moved in relation to the VRSs.
- At ear distance, reproduction of sound field coherence is affected by both the arrangement of a particular loudspeaker array and the number of VRSs. In the current spherical 86-channel loudspeaker array using VBAP for rendering of the VRSs, interaural coherence for the spherically isotropic case was reproduced up to 3.3 kHz for the highest number of 96 VRSs and for a real-life test scenario of a simulated staircase room up to at least 1.5 kHz for 12 and more VRSs.
- For a typical BTE microphone spacing, coherence can be well reproduced up to at least 5 kHz with 12 VRSs for all conditions considered here. For higher number of VRSs, no further improvements are observed in the considered frequency range up to 8 kHz.

Acknowledgments

The authors thank Sabine Hochmuth for providing additional OLSA recordings used in the listening test and Florian Denk for support for the HRTF measurements.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—[Project-ID 352015383]—SFB 1330 C 5.

ORCID iD

Christoph Kirsch https://orcid.org/0000-0002-7270-4715

Note

1. A Gaussian distribution of the individual scores per experiment condition can only be assumed for the majority of the data according to Shapiro-Wilk testing, given that the results for 6 VRSs are at the upper range of the measured scale. Sphericity was (Mauchly's test) was not violated.

References

Ahrens A., Lund K. D., Marschall M., & Dau T. (2019). "Sound source localization with varying amount of visual information in virtual reality," (M. S. Malmierca, Ed.). *PLoS ONE*, *14*(3), e0214603. https://doi.org/10.1371/journal.pone.0214603

Alary B., Masse P., Valimaki V., & Noisternig M. (2019). *Assessing the Anisotropic Features of Spatial Impulse Responses*. EAA Spatial Audio Signal Processing Symposium, Sep 2019, Paris, France. pp.43-48, (hal-02275194). https://doi.org/10.25836/SASP.2019.32

Allen J. B., & Berkley D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, *65*(4), 943–950. https://doi.org/10.1121/1.382599

Blauert J., & Lindemann W. (1986). Auditory spaciousness: Some further psychoacoustic analyses. *The Journal of the Acoustical Society of America*, *80*(2), 533–542. https://doi.org/10.1121/1.394048

Borish J. (1984). Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, *75*(6), 1827–1836. https://doi.org/10.1121/1.390983

Brinkmann F., Aspöck L., Ackermann D., Lepa S., Vorländer M., & Weinzierl S. (2019). A round robin on room acoustical simulation and auralization. *The Journal of the Acoustical Society of America*, *145*(4), 2746–2760. https://doi.org/10.1121/1.5096178

Brughera A., Dunai L., & Hartmann W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America*, *133*(5), 2839–2855. https://doi.org/10.1121/1.4795778

BS.2159-8 (2019). *Multichannel sound technology in home and broadcasting applications*. International Telecommunication Union. https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BS.2159-8-2019-PDF-E.pdf.

Cook R. K., Waterhouse R. V., Berendt R. D., Edelman S., & Thompson M. C. (1955). Measurement of correlation coefficients in reverberant sound fields. *The Journal of the Acoustical Society of America*, *27*(6), 1072–1077. https://doi.org/10.1121/1.1908122

Cord M., Baskent D., Kalluri S., & Moore B. (2007). Disparity between clinical assessment and real-world performance of hearing aids. *Hearing Review*, *14*, 22–26.

Cubick J., & Dau T. (2016). Validation of a virtual sound environment system for testing hearing aids. *Acta Acustica United with Acustica*, *102*(3), 547–557. https://doi.org/10.3813/AAA.918972

Culling J. F., Colburn H. S., & Spurchise M. (2001). Interaural correlation sensitivity. *The Journal of the Acoustical Society of America*, *110*(2), 1020–1029. https://doi.org/10.1121/1.1383296

Dietz M., Ewert S. D., & Hohmann V. (2011). Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication*, *53*(5), 592–605. https://doi.org/10.1016/j.specom.2010.05.006

Doclo S., Kellermann W., Makino S., & Nordholm S. E. (2015). Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones. *IEEE Signal Processing Magazine*, *32*(2), 18–30. https://doi.org/10.1109/MSP.2014.2366780

Dörbecker M., & Ernst S. (1996). Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverberation. 1996 8th European Signal Processing Conference (EUSIPCO 1996), Trieste, Italy, 1996, pp. 1–4.

Faller C., & Merimaa J. (2004). Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *The Journal of the Acoustical Society of America*, *116*(5), 3075–3089. https://doi.org/10.1121/1.1791872

Favrot S., Marschall M., Käsbach J., Buchholz J., & Welle T. (2011). Mixed-order Ambisonics recording and playback for improving horizontal directionality. 131st Audio Engineering Society Convention 2011, 641–647.

González Á (2009). Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, *42*(1), 49. https://doi.org/10.1007/s11004-009-9257-x

Goupell M. J., & Hartmann W. M. (2006). Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects. *The Journal of the Acoustical Society of America*, *119*(3), 3971–3986. https://doi.org/10.1121/1.2200147

Goupell M. J., & Hartmann W. M. (2007). Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models. *The Journal of the Acoustical Society of America*, *122*(2), 1029–1045. https://doi.org/10.1121/1.2734489

Grimm G., Ewert S., & Hohmann V. (2015). Evaluation of spatial audio reproduction schemes for application in hearing Aid research. *Acta Acustica United with Acustica*, *101*(4), 842–854. https://doi.org/10.3813/AAA.918878

Grimm G., Luberadzka J., & Hohmann V. (2019). A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology. *Acta Acustica United with Acustica, 105*(3), 566–578. doi: 10.3813/AAA.919337

Grosse J., Hungar F., Klockgether S., & van de Par S. (2015). "Wahrgenommene Quellbreite einer Lautsprecheranordnung in Abhängigkeit der physikalischen Quellbreite," (Perceived source with of a loudspeaker arrangement depending on the physical source width). Fortschritte der Akustik, presented at the DAGA 2015 Nürnberg.

Hautus M. J., & Meng X. (2002). Decision strategies in the ABX (matching-to-sample) psychophysical task. *Perception & Psychophysics*, *64*(1), 89–106. https://doi.org/10.3758/BF03194559

Hiyama K., Komiyama S., & Hamasaki K. (2002, October 1). *The Minimum Number of Loudspeakers and its Arrangement for*

Reproducing the Spatial Impression of Diffuse Sound Field. Gehalten auf der Audio Engineering Society Convention 113. http://www.aes.org/e-lib/browse.cfm?elib=11272

Hidaka T., Beranek L. L., & Okano T. (1995). Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. *The Journal of the Acoustical Society of America*, *98*(2), 988–1007. https://doi.org/10.1121/1.414451

Hochmuth S., Jürgens T., Brand T., & Kollmeier B. (2015). Talker- and language-specific effects on speech intelligibility in noise assessed with bilingual talkers: Which language is more robust against noise and reverberation? *International Journal of Audiology*, *54*(Suppl 2), 23–34. https://doi.org/10.3109/14992027.2015.1088174

Hodgson M. (1996). When is diffuse-field theory applicable? *Applied Acoustics*, *49*(3), 197–207. https://doi.org/10.1016/S0003-682X(96)00010-2

Huisman T., Piechowiak T., Dau T., & MacDonald E. (2019). Audio-visual sound localization in virtual reality. *Proceedings of the International Symposium on Auditory and Audiological Research*, *7*, 349–356. Retrieved from https://proceedings.isaar.eu/index.php/isaarproc/article/view/2019-40

Jacobsen F., & Roisin T. (2000). The coherence of reverberant sound fields. *The Journal of the Acoustical Society of America*, *108*(1), 204–210. https://doi.org/10.1121/1.429457

Jerger J. (2009). Ecologically valid measures of hearing aid performance. *Starkey Audiology Series*, *1*(1), 4. Retreived from https://starkeypro.com/pdfs/sas/Starkey_Audiology_Series_v1i1.pdf

Jeong C. -H. (2016). Kurtosis of room impulse responses as a diffuseness measure for reverberation chambers. *The Journal of the Acoustical Society of America*, *139*(5), 2833–2841. https://doi.org/10.1121/1.4949365

Jot J.-M., & Chaigne A. (1991). "Digital delay networks for designing artificial reverberators.,". Proceedings of the 90th Convention on the Audio Engineering Society, Paris, France.

Kayser H., Ewert S. D., Anemüller J., Rohdenburg T., Hohmann V., & Kollmeier B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing*, *2009*(1), 1–10. https://doi.org/10.1155/2009/298605

Keidser G., Naylor G., Brungart D. S., Caduff A., Campos J., Carlile S., & Smeds K. (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear and Hearing*, *41*, https://doi.org/10.1097/AUD.0000000000000944R

Klumpp R. G., & Eady H. R. (1956). Some measurements of interaural time difference thresholds. *The Journal of the Acoustical Society of America*, *28*(5), 859–860. https://doi.org/10.1121/1.1908493

Kollmeier B., Peissig J., & Hohmann V. (1993). Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *Journal of Rehabilitation Research and Development*, *30*(1), 82–94.

Kuehnel V., Kollmeier B., & Wagener K. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache I: Design des oldenburger satztests. *Zeitschrift für Audiologie*, *38*(1), 4–15. Retreived from https://www.researchgate.net/publication/266735660_Entwicklung_und_Evaluation_eines_Satztests_fur_die_deutsche_Sprache_I_Design_des_Oldenburger_Satztests

Kuttruff H. (1995). A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries. *The Journal of the Acoustical Society of America*, *98*(1), 288–293. https://doi.org/10.1121/1.413727

Lachenmayr W., Haapaniemi A., & Lokki T. (2016, Mai 26). *Direction of Late Reverberation and Envelopment in Two Reproduced Berlin Concert Halls*. Gehalten auf der Audio Engineering Society Convention 140.

Laitinen M.-V., Vilkamo J., Jussila K., Politis A., & Pulkki V. (2014). Gain normalization in amplitude panning as a function of frequency and room reverberance. In AES 55th International Conference, Helsinki, Finland, August, 2014 (pp. 1-6).

Laitinen M., & Pulkki V. (2009). Binaural reproduction for Directional Audio Coding. 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 337–340. https://doi.org/10.1109/ASPAA.2009.5346545

Lindevald I. M., & Benade A. H. (1986). Two-ear correlation in the statistical sound fields of rooms. *The Journal of the Acoustical Society of America*, *80*(2), 661–664. https://doi.org/10.1121/1.394061

Luizard P., Katz B. F. G., & Guastavino C. (2015). Perceptual thresholds for realistic double-slope decay reverberation in large coupled spaces. *The Journal of the Acoustical Society of America*, *137*(1), 75–84. https://doi.org/10.1121/1.4904515

Macmillan N. A., & Creelman C. D. (2004). *Detection Theory: A User's Guide (2nd ed.)*. New York: Psychology Press. https://doi.org/10.4324/9781410611147

Miles K. M., Keidser G., Freeston K., Beechey T., Best V., & Buchholz J. M. (2020). Development of the everyday conversational sentences in noise test. *The Journal of the Acoustical Society of America*, *147*(3), 1562–1576. https://doi.org/10.1121/10.0000780

Moore B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). Academic Press.

Nolan M., Fernandez-Grande E., Brunskog J., & Jeong C. -H. (2018). A wavenumber approach to quantifying the isotropy of the sound field in reverberant spaces. *The Journal of the Acoustical Society of America*, *143*(4), 2514–2526. https://doi.org/10.1121/1.5032194

Oreinos C., & Buchholz J. (2014). Validation of realistic acoustic environments for listening tests using directional hearing aids. 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC), Juan-les-Pins, France, 2014, pp. 188-192, https://doi.org/10.1109/IWAENC.2014.6954004.

Oreinos C., & Buchholz J. M. (2016). Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids. *Journal of the American Academy of Audiology*, *27*(07), 541–556. https://doi.org/10.3766/jaaa.15094

Pausch F., & Fels J. (2020). Localization performance in a binaural real-time auralization system extended to research hearing aids. *Trends in Hearing*, *24*, 2331216520908704. https://doi.org/10.1177/2331216520908704

Poppitz J., Wendt T., van de Par S., & Ewert S. (2018). Required Spatial Resolution for Late Reverberation in a 3-dimensional Loudspeaker Array, Fortschritte der Akustik, Deutsche Gesellschaft für Akustik, Berlin. Presented at the DAGA 2018 München.

Polack J. -D. (1993). Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics. *Applied Acoustics*, *38*(2–4), 235–244. https://doi.org/10.1016/0003-682X(93)90054-A

Pöntynen H., Santala O., & Pulkki V. (2016). Conflicting dynamic and spectral directional cues form separate auditory images. Presented at AES 140th Convention, Paris, France, June, 2016.

Pulkki V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, *45*(6), 456–466. Retreived from http://www.aes.org/e-lib/browse.cfm?elib=7853

Pulkki V., & Karjalainen M. (2015). *Communication acoustics: An introduction to speech, audio, and psychoacoustics*. Wiley.

Rindel J. (2001). The use of computer modeling in room acoustics. *Journal of Vibroengineering*, *3*(4), 219–224.

Robinson D. E., & Jeffress L. A. (1962). Effect of noise correlation on binaural-signal detection. *The Journal of the Acoustical Society of America*, *34*(5), 727–728. https://doi.org/10.1121/1.1937229

Romblom D., Guastavino C., & Depalle P. (2016). Perceptual thresholds for non-ideal diffuse field reverberation. *The Journal of the Acoustical Society of America*, *140*(5), 3908–3916. https://doi.org/10.1121/1.4967523

Schlecht S. J., & Habets E. A. P. (2015). Time-varying feedback matrices in feedback delay networks and their application in artificial reverberation. *The Journal of the Acoustical Society of America*, *138*(3), 1389–1398. https://doi.org/10.1121/1.4928394

Schlecht S. J., & Habets E. A. P. (2017). Feedback delay networks: Echo density and mixing time. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(2), 374–383. https://doi.org/10.1109/TASLP.2016.2635027

Schröder D. (2011). *Physically based real-time auralization of interactive virtual environments, aachener beiträge zur technischen akustik*. Logos-Verlag. 206 pages.

Schutte M., Ewert S. D., & Wiegrebe L. (2019). The percept of reverberation is not affected by visual room impression in virtual environments. *The Journal of the Acoustical Society of America*, *145*(3), EL229–EL235. https://doi.org/10.1121/1.5093642

Seeber B. U., Kerber S., & Hafter E. R. (2010). A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hearing Research*, *260*(1–2), 1–10. https://doi.org/10.1016/j.heares.2009.11.004

Srinivasan S. (2008). Noise reduction in binaural hearing aids: Analyzing the benefit over monaural systems. *The Journal of the Acoustical Society of America*, *124*(6), EL353–EL359. https://doi.org/10.1121/1.2993747

Thiemann J., Müller M., Marquardt D., Doclo S., & van de Par S. (2016). Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene. *EURASIP Journal on Advances in Signal Processing*, *2016*(1), 12. https://doi.org/10.1186/s13634-016-0314-6

Välimäki V., Parker J. D., Savioja L., Smith J. O., & Abel J. S. (2012). Fifty years of artificial reverberation. *IEEE Transactions on Audio Speech and Language Processing*, *20*(5), 1421–1448. https://doi.org/10.1109/TASL.2012.2189567

Wadell H. (1935). Volume, shape, and roundness of quartz particles. *The Journal of Geology*, *43*(3), 250–280. https://doi.org/10.1086/624298

Walther A., & Faller C. (2011). Assessing Diffuse Sound Field Reproduction Capabilities of Multichannel Playback Systems. Presented at Audio Engineering Society Convention 130. Retrieved from: https://www.aes.org/e-lib/browse.cfm?elib = 15895

Walther A., & Faller C. (2013). Interaural correlation discrimination from diffuse field reference correlations. *The Journal of the Acoustical Society of America*, *133*(3), 1496–1502. https://doi.org/10.1121/1.4790473

Wendt T., van de Par S., & Ewert S. (2014). A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation. *Journal of the Audio Engineering Society*, *62*(11), 748–766. https://doi.org/10.17743/jaes.2014.0042

Yousefian N., & Loizou P. (2011). A dual-microphone speech enhancement algorithm based on the coherence function. *IEEE Transactions on Audio, Speech, and Language Processing, 5957265 20*(2), 599–609. https://doi.org/10.1109/TASL.2011.2162406

Zwislocki J., & Feldman R. S. (1956). Just noticeable differences in dichotic phase. *The Journal of the Acoustical Society of America*, *28*(5), 860–864. https://doi.org/10.1121/1.1908495