OPEN ACCESS Freely available online

C. Kirsch, T. Wendt S. Van De Par, H. Hu and S.D. Ewert,
"Computationally-Efficient Simulation of Late Reverberation for Inhomogeneous Boundary Conditions and Coupled Rooms"
J. Audio Eng. Soc., vol. 71, no. 4, pp. 186–201, (2023 April).
DOI: https://doi.org/10.17743/jaes.2022.0053

Computationally-Efficient Simulation of Late Reverberation for Inhomogeneous Boundary Conditions and Coupled Rooms

CHRISTOPH KIRSCH,[1,*†] TORBEN WENDT,[1,2,*†] STEVEN VAN DE PAR,[2]
(christoph.kirsch@uol.de) (torben.wendt@uol.de) (steven.van.de.par@uol.de)
HONGMEI HU,[1] AND STEPHAN D. EWERT[1]
(hongmei.hu@uol.de) (stephan.ewert@uol.de)

[1] *Medizinische Physik and Cluster of Excellence Hearing4all, Carl von Ossietzky Universität, Oldenburg, Germany*
[2] *Akustik and Cluster of Excellence Hearing4all, Carl von Ossietzky Universität, Oldenburg, Germany*

For computational efficiency, acoustic simulation of late reverberation can be simplified by generating a limited number of incoherent signals with frequency-dependent exponential decay radiated by spatially distributed virtual reverberation sources (VRS). A sufficient number of VRS and adequate spatial mapping are required to approximate spatially anisotropic late reverberation, e.g., in rooms with inhomogeneous distribution of absorption or for coupled volumes. For coupled rooms, moreover, a dual-slope decay might be required. Here, an efficient and perceptually plausible method to generate and spatially render late reverberation is suggested. Incoherent VRS signals for (sub-) volumes are generated based on room dimensions and frequency-dependent absorption coefficients at the boundaries. For coupled rooms, (acoustic) portals account for effects of sound propagation and diffraction at the room connection and energy transfer during the reverberant decay process. The VRS are spatially distributed around the listener, with weighting factors representing the spatially subsampled distribution of absorption on the boundaries and the location and solid angle covered by portals. A technical evaluation and listening tests demonstrate the validity of the approach in comparison to measurements in real rooms.

0 INTRODUCTION

Room acoustics simulation and virtual acoustic environments (VAEs) enable the auralization of existing or not (yet) existing spaces with applications ranging from architectural planning [1] to entertainment [2, 3]. Because of their potentially high ecological validity [4–6], VAEs have also gained interest as tools for psychoacoustic research and hearing aid development [7–10].

For many of these applications, computational efficiency is important to allow for interactive real-time updates with six degrees of freedom (6-DOF) for movement of sources and receivers. Thus, simplifications of the underlying room acoustics simulation and the rendering are desirable.

In enclosed spaces, (early) reflections at boundaries and reverberation occur in addition to the direct sound. Whereas the direction of individual early reflections might be perceivable, late reverberation results from a superposition of many densely spaced reflections that are spatially more or less evenly distributed. Depending on room geometry, the presence of coupled (sub-) volumes, and the spatial distribution of sound absorption at the boundaries, the resulting diffuse late reverberation can be considered spherically isotropic or might contain limited spatial directivity [11–15] and direction-dependent multi-slope decays (e.g., [16]). Thus, for room acoustics simulation and rendering of sound fields in VAEs, computationally-efficient methods that provide diffuse late reverberation with appropriate spatial resolution and spatio-temporal decay characteristics are of interest. Particularly, applications focusing on human perception and behavior in the VAE might allow for simplifications of the simulation and rendering of late reverberation without altering the perceived sound field.

Room acoustics simulations are often based on geometrical acoustics in the form of image-source models (ISMs) [17, 18], ray tracing [19, 20], radiosity [21], and radiance transfer [22]. Using statistical assumptions, even more com-

*To whom correspondence should be addressed, e-mail: christoph.kirsch@uol.de. Last updated: July 11, 2022.

†The first two authors contributed equally and share first authorship.

putationally efficient reverberation algorithms like feedback delay networks (FDNs; [23]) can be applied for generating reverberated signals (e.g., [16, 24]), which, however, lack in accurately modeling room geometry. In hybrid approaches, simplified statistical methods, which fulfill perceptual requirements for late reverberation, can be combined with more accurate geometrical acoustic models for early reflections. An ISM for early reflections and raytracing to calculate the spatiotemporal energy distribution of scattered reflections and late reverberation were combined in [25]. An ISM and spatialized FDN for late reverberation were combined in [26].

Perceptual evaluations in [27] showed that a high degree of perceptual plausibility can be achieved with current room acoustics simulation methods, for a variety of large and small rooms, including two rooms coupled by a door opening. However, [28] showed that many approaches cannot successfully account for dual-slope decays observed in the latter case. In addition to a dual slope decay, often a sloped onset of the room impulse response (RIR) is observed instead of a pronounced initial peak caused by the unobstructed direct sound and early reflections. In such a case, diffracted direct sound and early reflections for sources in the connected room can result in challenging daily-life communication scenarios in home environments, particular for elderly or hard-of-hearing people. Accordingly, such acoustic scenarios are of interest for ecologically valid VAEs (see, e.g., [29]).

In the room-simulation approaches discussed before, late reverberation is typically rendered using a number of (incoherent) late reverberation signals that are spatially evenly distributed around the listener, using an FDN (e.g., [26]) or incoherent decaying noises with a spectral profile matching the frequency-dependent reverberation time (e.g., [30]). Such virtual reverberation sources (VRS) can be played back on physical or virtual loudspeakers in a loudspeaker array, or alternatively, they can be spatialized through head-related transfer functions (HRTFs) on headphones. Using these methods, a diffuse sound field can be approximated by superposition of incoherent sounds from many directions (for a review of theoretical approaches see, e.g., [31]). The spatial resolution of late reverberation simulation and rendering is therefore determined by the number of (incoherent) late reverberation signals.

Regarding the perception of diffuse sound fields, [32] found that a specific horizontal arrangement of only four sound sources separated by 90° can already be sufficient to reproduce the spatial impression of a diffuse sound field. With such a low number, however, the results were strongly dependent on the position of VRS with regard to the listener. This is problematic in VAE applications, in which listeners can freely rotate their heads.

Although an isotropic late reverberant field is approximated in the case of a completely homogeneous distribution of absorption coefficients at all boundaries, a more critical case regarding spatial resolution of late reverberation occurs for inhomogenously distributed absorption coefficients and in the case of coupled (sub-) volumes with connecting openings. Such a critical test case was assessed in [33], using a spatial mapping method for inhomogeneous boundary conditions in shoebox shaped rooms [34]. They have shown that 12–24 VRS are perceptually sufficient in a single room with anisotropic late reverberation, depending on the stimulus. Although [33] provides an estimate of the perceptually required spatial resolution for anisotropic late reverberation, there is no comparable computationally efficient method available that is also applicable to more-complex room geometry and coupled rooms (or generally coupled volumes with connecting openings).

In the current study, computationally efficient methods to generate and spatially render isotropic and anisotropic late reverberation are suggested for single and coupled rooms. For simplicity, a "proxy" shoebox approximation of each single room geometry is created with the room acoustics simulator RAZR [26]. Based on the concept of "portals" originally applied for visibility determination in interactive computer graphics [35–38], and later also adopted in virtual acoustics (e.g., [25]), acoustic portals are used to handle effects of diffraction and energy transfer between volumes in a simplified and computationally efficient way.

In a first step, spatial rendering of late reverberation was addressed in a simulated room with one highly absorbing wall covering various solid angles [or field of view (FOV)] from the listener position as used earlier in [33, 34]. For this purpose, a simplistic, highly efficient spatial subsampling method is suggested to map the spatial distribution of reflection coefficients at the shoebox room boundaries to weighting factors for the VRS. An alternative method for arbitrary room geometry is suggested and compared to the shoebox room method. For technical evaluation, the coherence between the two ears of a head and torso simulator in the sound field generated by the VRS was analyzed. The number of VRS was varied from low spatial resolution (six VRS) to high spatial resolution (96 VRS). Interaural coherence (IC) is considered relevant for psychoacoustic processes (e.g., [39, 40]) and has been suggested as a suitable metric for assessing the reproduction quality of diffuse sound fields (e.g., [32, 41]). Additionally, the frequency-dependent interaural level difference (ILD) was assessed.

In a second step, two rooms connected by a typical door opening were considered. Each room was simulated by its proxy shoebox with early reflections undergoing a strongly simplified low-pass diffraction filter (see e.g., [42]) when passing the portal. For diffuse late reverberation, coupling of the room volumes through an aperture surface based on [43] was used. Late reverberation from the receiver and neighbor room is rendered using 12 VRS. Depending on the listener position relative to the aperture (door), the spatial distribution of the VRS emitting the late reverberation from the neighbor and receiver room were adapted to enable a smooth transition while passing through the door. The VRS were spatially mapped around the listener's head such that the solid angle covered by the door was correctly represented.

For the technical evaluation, energy decay curves were measured for a variety of source and receiver positions in different connected rooms and compared to the simulation. A psychoacoustic experiment was conducted to assess

differences on spatial audio quality features between the measured and simulated rooms using headphone auralization.

1 INHOMOGENEOUS BOUNDARY CONDITIONS

1.1 Basic Room Acoustics Simulation Method

RAZR [26] generates perceptually plausible room impulse responses (RIRs) using a hybrid ISM/FDN approach. To achieve high computational efficiency, the ISM uses a shoebox approximation of room geometry. By default, RAZR computes discrete early (specular) reflections up to the third reflection order in the ISM. The reverberation tail is generated by a 12-channel FDN, which is fed by the last order of reflections from the ISM. The output channels of the FDN are used as VRS that are spatially mapped to the same number of discrete locations on a listener-centered cube, aligned with the six walls of the "proxy" shoebox room (reverberation cube mapping). In this default setting, the spatial resolution of the late reverberation is a result of 12 discrete directions (two representing each wall) spatially evenly distributed around the listener. The frequency-dependent absorption coefficients of each of the six walls determines a wall-specific "absorption filter," which is applied to the respective two FDN outputs in two VRS representing each wall. Perceptual plausibility of the resulting room acoustics simulations was demonstrated by favorable performance in comparison to other state-of-the-art approaches ([27], see algorithm B in their Fig. 8).

1.2 Extension to Inhomogeneous Boundary Conditions for Shoebox Geometry

In the current study, RAZR was adapted to generate six, 12, 24, 48, or 96 VRS, equivalent to one, two, four, eight, and 16 VRS per surface of the underlying shoebox room. To ensure similar spectro-temporal characteristics of the FDN output throughout evaluation and avoiding subtle timbre changes in the resulting reverberant tail due to a variation of the number of FDN channels [44, 45], the FDN always operated with the highest number of 96 channels, independent of the number of rendered VRS. The 96 FDN outputs were directly used for 96 VRS and downmixes with 48, 24, 12, and six VRS were generated in a sequential procedure by adding pairs of the FDN output channels.

The VRS were spatially distributed according to the directions of vertices of polyhedra, centered on the listener and directionally aligned with the (shoebox) room boundaries. The polyhedra had a number of vertices equivalent to the number of VRS and were optimized for geometrical sphericity [46], which ranged from 0.86 for six VRS to 0.99 for 96 VRS. For six VRS, the resulting directions are orthogonal to each other, and for 12 VRS, they correspond to points lying on the diagonals of the surfaces of a room-aligned cube (similar to the original alignment suggested in [26]). Directions for 24 and more VRS were based on a combination of one, two, and four snub cuboctahedra (snub cubes).

In rooms with an inhomogeneous distribution of acoustic absorption, e.g., a shoebox room with one highly absorbing wall, and otherwise more reflective walls, anisotropic late reverberation will occur. The spatial distribution of late reverberation depends on the distance of the listener to the absorbing wall, affecting the solid angle (or FOV) occupied by the wall. In the extreme case of a listener being very close to one completely absorbing wall, no reverberant sound energy impinges from the hemisphere occupied by the absorbing wall. To cope with the general case of different arbitrary absorption coefficients per wall (representing the average absorption coefficient of the room geometry approximated by the walls of the proxy shoebox room), RAZR was extended by a spatial sampling routine to calculate the absorption filters applied at the output of each delay line of the FDN that maps to the VRS. Dependent on the listener's position in relation to the boundaries, resulting FOV occupied by each boundary, and direction of the VRS in relation to the boundaries, the contribution of the absorption coefficients to the absorption filter at the FDN output for the respective VRS were blended.

Fig. 1 shows a two-dimensional representation of an example case for a number of 48 VRS [Fig. 1(a)] and eight VRS [Fig. 1(b)] in the horizontal plane. Here, an elongated room is shown, and the receiver is placed close to the upper $+y$ wall. The dashed lines are along the x and y directions and coincide with the (axis-aligned) wall normals. The inner circle of crosses (color online) shows a straightforward approach in which the absorption coefficients of a respective wall are directly projected onto the VRS. Although this simplistic approach appears feasible for a high number of VRS [Fig. 1(a)], it is unsuited for lower numbers of VRS [Fig. 1(b)], in which the lower $-y$ wall would not be represented at all. Moreover, for a moving listener, the lower $-y$ wall may become represented in a discontinuous manner, which would lead to potentially audible artefacts. Therefore, an efficient spatial sub-sampling that allows for smooth transitions within interactive environments has to be devised.

To achieve a smooth sampling for arbitrary number and positions of VRS, the space around the receiver position was divided in octants (quadrants in the depicted two-dimensional case in Fig. 1, as indicated by the dashed lines). In each octant, the VRS direction vectors (from receiver to VRS) were divided elementwise by the vector from the receiver to the shoebox corner (vertex, thin dash-dotted lines) contained in the octant, and then normalized. The resulting warped VRS position of index k can be expressed as

$$\overrightarrow{\mathrm{VRS}}_{\mathrm{w,k}} = \frac{\overrightarrow{\mathrm{VRS}}_{\mathrm{k}} \oslash \vec{V}_{\mathrm{n,abs}}}{\left\| \overrightarrow{\mathrm{VRS}}_{\mathrm{k}} \oslash \vec{V}_{\mathrm{n,abs}} \right\|}, \tag{1}$$

where $\overrightarrow{\mathrm{VRS}}_{\mathrm{k}} = [x_k \; y_k \; z_k]$ is the original VRS position, $\oslash$ is the Hadamard (elementwise) division operator, and $\vec{V}_{\mathrm{n,abs}}$ is the vector of absolute values of the room vertex coordinates relative to the receiver in the octant n. This transformation warps the vectors from the receiver to the eight vertices of the shoebox to represent a receiver-centered cube. Likewise, all k VRS directions are warped in such a way that a

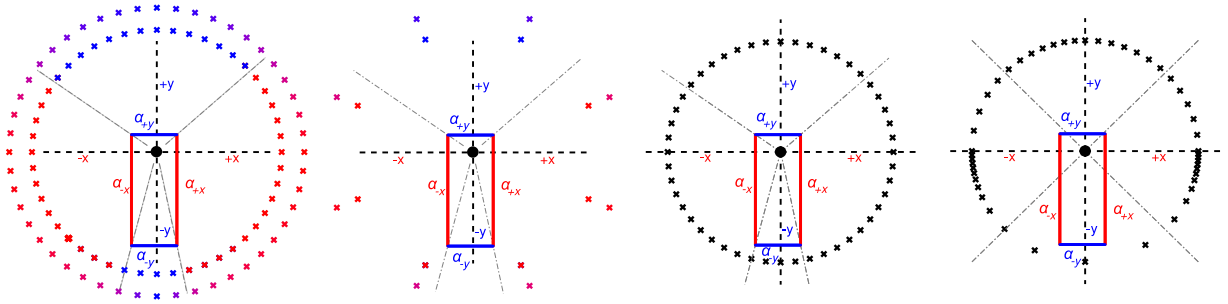


Fig. 1. Left panels: Distribution of virtual reverberation sources (VRS) (crosses) around the listener (black circle) in a room (rectangle) with 48 VRS (a) and 8 VRS (b). Colors (online) represent reflection coefficients associated with particular boundary surfaces of the room. Inner circle: simple assignment of VRS to a certain boundary based on projection. Outer circle: smooth, weighted assignment based on warping on the VRS directions. Right panels: VRS directions (black crosses) and directions of the room corners (dash-dotted lines) before warping (c) and after warping (d).

VRS on the room corner appears exactly on the according cube vertex. Figs. 1(c) and 1(d) [Fig. 1(c), before warping; Fig. 1(d), after warping] show the effect of warping in the depicted two-dimensional representation. The dash-dotted lines to the edges (vertices) appear under an angle of 45°, exactly on the diagonals between the normals of the two walls in each quadrant after warping [Fig. 1(d)].

Using the warped VRS directions, the contributions of the reflection coefficients of each of the orthogonal shoebox walls are calculated using vector base amplitude panning (VBAP; [47]) with the inverted (outside-pointing) wall normals as vector base. Eq. (18) from [47] is therefore adapted to yield the contribution $\vec{g}_\mathrm{k} = [g_1\ g_2\ g_3]$ of the absorption coefficients of the three intersecting walls in each octant with the normal vectors n_1, n_2, n_3 to the filtering applied to a particular VRS with index k:

$$\vec{g}_\mathrm{k} = \overrightarrow{\mathrm{VRS}}_{\mathrm{w,k}} \begin{bmatrix} n_{1,\mathrm{x}} & n_{1,\mathrm{y}} & n_{1,\mathrm{z}} \\ n_{2,\mathrm{x}} & n_{2,\mathrm{y}} & n_{2,\mathrm{z}} \\ n_{3,\mathrm{x}} & n_{3,\mathrm{y}} & n_{3,\mathrm{z}} \end{bmatrix}^{-1} . \tag{2}$$

For an axis-aligned shoebox room, $\vec{g}_\mathrm{k}$ simplifies to the absolute values of each element of $\overrightarrow{\mathrm{VRS}}_{\mathrm{w,k}}$. The resulting spatially (sub-) sampled absorption coefficient a_k (typically computed for multiple frequencies), which serves as a basis for designing the absorption filter for each VRS, is

$$a_\mathrm{k} = \vec{g}_\mathrm{k} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}, \tag{3}$$

where α_1, α_2, α_3 are the absorption coefficients of the respective walls in the octant.

Following this method, a VRS at a room corner receives an identical contribution from all three walls (two in the depicted two-dimensional case). In a case in which the listener is close to a highly absorbing boundary, the corresponding VRS in the FOV, defined by the "view" frustum from the receiver to the four shoebox vertices of the boundary now emits significantly less energy than the other VRS, mimicking that no or reduced sound energy in the late reverberant field can emerge from that boundary. The farther the listener moves away from the absorbing boundary, the smaller the surface's FOV and the fewer VRS are affected. Depending on the overall number of VRS, the spatial resolution for representing the absorbing boundary, as observed from the listener position, is reduced. This results in an increasingly blurred spatial representation with decreasing number of VRS. It is assumed that the highest number of 96 VRS applied here is sufficiently high to serve as a reference condition for evaluation of spatial resolution of late reverberation with all 96 FDN channels mapped to separate VRS.

1.3 Generalized Approach for Arbitrary Geometry

An alternative, computationally more expensive extension of the above approach to arbitrary room geometry is based on ray tracing and a triangulated representation to the room. In the simplest case, each wall of a shoebox room is represented by two triangles. The underlying idea is again to obtain a fast, spatially subsampled representation of the distribution of absorption coefficients, assuming that each triangle has absorption coefficients assigned. The absorption coefficients around the listener are sampled by intersecting a uniform distribution of directions originating from the listeners with the surface triangles. For non-convex geometries, the triangle with the intersection point that is closest to the listener is used to account for visibility.

Here, a deterministic ray-tracing approach was implemented, based on the 96 directions of the highest available VRS resolution. When 96 VRS are rendered, each VRS is assigned the absorption coefficient α_s of the triangle that is intersected by its respective "scanned" direction. If a smaller number of VRS is rendered, a subsampling of the scanned directions is performed, using weighting coefficients $w_{\mathrm{s,k}}$ based on a dot product between the scan direction unit vector $\vec{v}_\mathrm{s}$ and VRS direction unit vector $\vec{v}_\mathrm{k}$

$$w_{\mathrm{s,k}} = (\vec{v}_\mathrm{s} \cdot \vec{v}_\mathrm{k})^\beta . \tag{4}$$

Negative dot products are set to a value of 0, resulting in a direction-dependent weighting. The –3-dB lobe width for the resulting (directivity) weighting function is determined by β. The lobe width is adjusted for numbers of K VRS

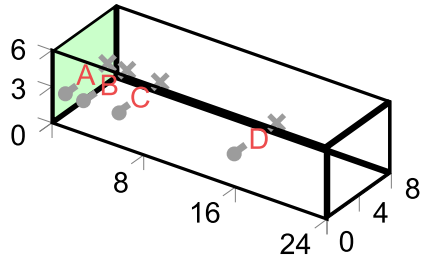


Fig. 2. Virtual corridor. Source positions are shown as crosses; receiver positions and orientation are shown by the nosed balls. The highly absorbing surface is shaded. The distances to the highly absorbing surface were chosen so that the horizontal field of views (FOV) was approximately 170°, 110°, 70°, and 30° for position A, B, C, and D, respectively.

according to $\beta = K/6$, resulting in 90° for $K = 6$. The weighted absorption coefficient for each VRS is

$$a_k = \frac{\vec{w}_{s,k} \cdot \vec{\alpha}_s}{\sum w_{s,k}}, \tag{5}$$

where $\vec{w}_{s,k}$ and $\vec{\alpha}_s$ are vectors of the weights and absorption coefficients for all scanned directions. Eq. (5) is typically calculated for absorption coefficients at multiple frequencies. For a static VRS arrangement, the weights $w_{s,k}$ remain unchanged and can be pre-calculated.

For detailed geometry, the deterministic scanning directions can result in missing surfaces that cover a small FOV from the listener, depending on the listener position. To avoid artifacts for moving listeners, temporal averaging methods can be applied. Additionally, randomized directions or triangle-based directions can be included.

1.4 Technical Evaluation

1.4.1 Virtual Room

A corridor (Fig. 2) with the dimensions 24 m × 8 m × 6 m and inhomogeneous absorption coefficients was used for the evaluation of the proposed method. One of the small surfaces (shaded) at the end of the corridor was highly absorbing (with an absorption coefficient $\alpha = 0.99$ for all frequencies), whereas all other surfaces were quite reflective ($\alpha = 0.01$ to 0.11 from 125 Hz to 8 kHz). The resulting reverberation time ranged from approximately 1.3 s at 125 Hz to 0.8 s at 8 kHz.

Four source-receiver–position combinations were considered to vary the solid angle (or 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 absorbing 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 to the highly absorbing wall, so that a wide range of different FOVs of the absorbing wall was obtained. The default receiver orientation was toward the sound source (azimuth angle of 0°) for all combinations (denoted A–D, see Fig. 2).

1.5 Results

IC and ILD were estimated using the FABIAN HRTF database [48] in the late reverberant sound field (disregarding early reflections and direct sound), rendered using the VRS directions. For this technical evaluation, instead of using the uncorrelated outputs of the FDN, independent Gaussian noises were used as output of the VRS, which received the weights according to the suggested spatial sampling method.

The frequency-dependent IC estimates $C_{lr}(f)$ between the signals l and r in both ears were calculated according to

$$C_{lr}(f) = \Re\left(\frac{G_{lr}(f)}{\sqrt{G_{ll}(f)\,G_{rr}(f)}}\right), \tag{6}$$

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 time windows with a 75% overlap and a length of 512 samples at 44.1-kHz sampling rate to obtain an average coherence estimate. The Gaussian noise VRS signals were 60 s in duration.

The ILD was calculated from the same simulated ear signals l and r as

$$\mathrm{ILD}(f) = 10\log_{10}{}^{G_{ll}(f)}\!/_{G_{rr}(f)}. \tag{7}$$

The left column of Fig. 3 shows the IC for increasing numbers of VRS (top to bottom) for position A (close to the absorbing wall) and position D (farthest away from the absorbing wall). Intermediate results are observed for positions B and C, which are not shown for clarity. The dotted trace represents the approximation of the isotropic case with the same output power assigned to all VRS instead of the suggested weighting according to the inhomogeneous distribution of absorption coefficients at the boundaries.

The results for positions A and D for 96 VRS are shown for comparison in all panels (less-saturated thick traces). Large deviations from the 96 VRS condition are observed for six VRS (top panel) above about 400 Hz for the shoebox approach (solid traces), and there is no difference between positions A and D. The generalized approach (dashed traces) deviates from 96 VRS above about 800 Hz, with only position D closely resembling the coherence indicated for the approximated isotropic condition (dotted trace). For an increasing number of VRS, the IC gradually asymptotes against that obtained for 96 VRS with deviations only occurring at increasingly higher frequencies. Table 1 shows the upper frequency limit above which clear deviations occur.

Differences between position A and D are generally small and also become increasingly similar to those observed for 96 VRS with increasing number of VRS. As shown for the "reference" case of 96 VRS, only small differences in IC for positions A and D are observed for more than six VRS, although very little sound energy impinges from one hemisphere in position A. In position D, the results already are virtually identical with those observed for the approximated isotropic case (dotted traces).

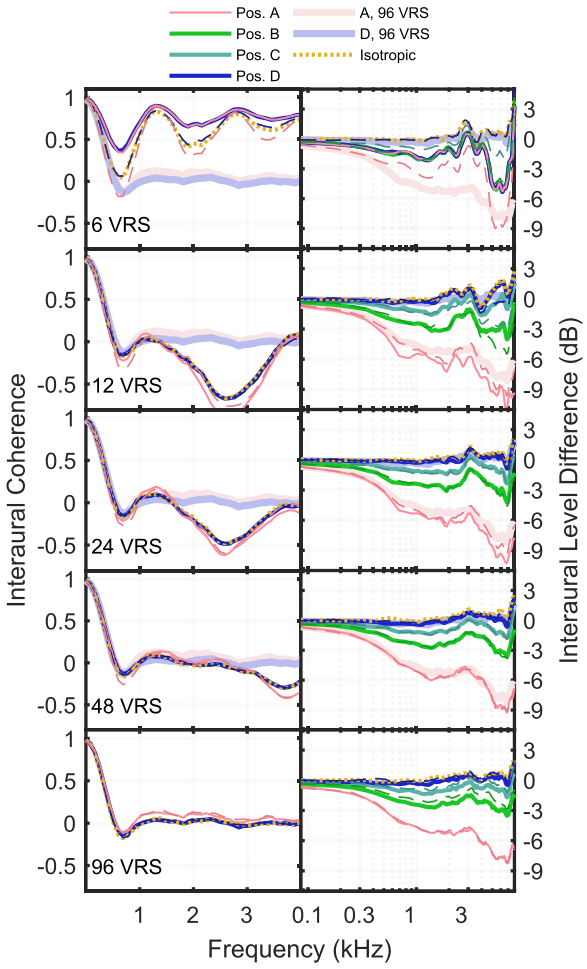


Fig. 3. Interaural coherence (left column) and interaural level difference (right column) for 6–96 virtual reverberation sources (VRS) (top to bottom) and different positions in the corridor as indicated in the legend. Solid traces (color online) correspond to the shoebox approach for mapping wall absorption coefficients to VRS; thin dashed traces indicate the results of the generalized approach. The dotted trace shows the approximation of an isotropic sound field with varying number of VRS. For reference, the results for 96 VRS for position A and D are displayed by less saturated thick lines in all panels.

Table 1. Heuristically determined frequency limit for reasonably accurate sound reproduction compared to the 96-VRS reference. All values are in Hertz. Values in parentheses are for the generalized approach if differences occurred.

	IC		ILD	
VRS	A	D	A	D
6	400 (800)	400 (800)	400	400
12	950	950	600 (5,000)	5,000
24	1,800	1,800	1,500 (>8,000)	>8,000
48	3,000	3,000	>8,000	>8,000

IC = interaural coherence; ILD = interaural level difference; VRS = virtual reverberation sources.

The right column of Fig. 3 shows the ILD for different numbers of VRS and all positions A–D. Generally, ILDs are largest for position A (close to the absorbing wall) and decrease to near zero for the farthest position D, similar to the ILD observed for the approximation of the isotropic case (dotted). Remaining small ILDs are to be attributed to the HRTF database used for simulation and the selection of the nearest-neighbor directions in the database (according to the VRS directions).

Again, large deviations from 96 VRS are mainly observed for six VRS with the shoebox approach (solid), in which the maximum ILD for position A (up to 5 dB) in the depicted 8-kHz range is considerably smaller than for 96 VRS (up to 8 dB) and results are independent of the listener position. The generalized approach (dashed) with six VRS underestimates the ILD for position A in a frequency range from about 1–3 kHz; however, a gradual decrease in ILD for positions with decreasing FOV is observed. For 12 VRS, the generalized approach matches ILDs from the 96 VRS reference more closely than the shoebox approach, which results in an overestimation (up to 10 dB). For larger numbers of VRS, both approaches gradually asymptote against the result for 96 VRS. The right-hand columns in Table 1 indicate the upper frequency limit of reasonable agreement with the 96 VRS case for A and D, which increases for increasing number of VRS. Values in parentheses are for the generalized approach if differences occurred.

In Fig. 4, IC (left column) and ILD (right column) are displayed for different receiver orientations to assess the effect of rotational movements in the VAE using the shoebox approach. The orientation of 0° (replot of traces from Fig. 3) refers to the orientation indicated in Fig. 2. Orientations of 30° (dashed) and 60° (dotted) are clockwise rotations towards the corridor. It is obvious that IC in the left column is independent of rotation up to a frequency limit which increases with the number of VRS. For six VRS, IC for different rotations starts diverging around 500 Hz, and at about 900 Hz for 12 VRS, 2,000 Hz for 24 VRS, and 3,000 Hz for 48 and 96 VRS.

The ILDs depicted in the left column of Fig. 4 show that for any number higher than six VRS, there are no major deviations in ILD depending on orientation for the near-isotropic position D. An overall similar behavior is observed for all orientations for more than six VRS at position A. For six VRS, strong deviations from 96 VRS are observed for all orientations for the shoebox approach. Here, the generalized approach is additionally shown (thin traces), resulting in smaller deviations. Traces representing the generalized approach are not shown in all other panels for clarity, given that results are very similar to the shoebox approach.

In summary, the technical evaluation shows a large deviation of the results for six VRS from those obtained with higher numbers of VRS, particularly for the shoebox approach. The IC in the perceptually relevant range up to 1.5 kHz (e.g., [49–51]) is well approximated for 24 VRS and more, whereas 12 VRS still result in deviations above about 750 Hz (especially when receiver rotation is considered). However, these deviations mainly occur for small

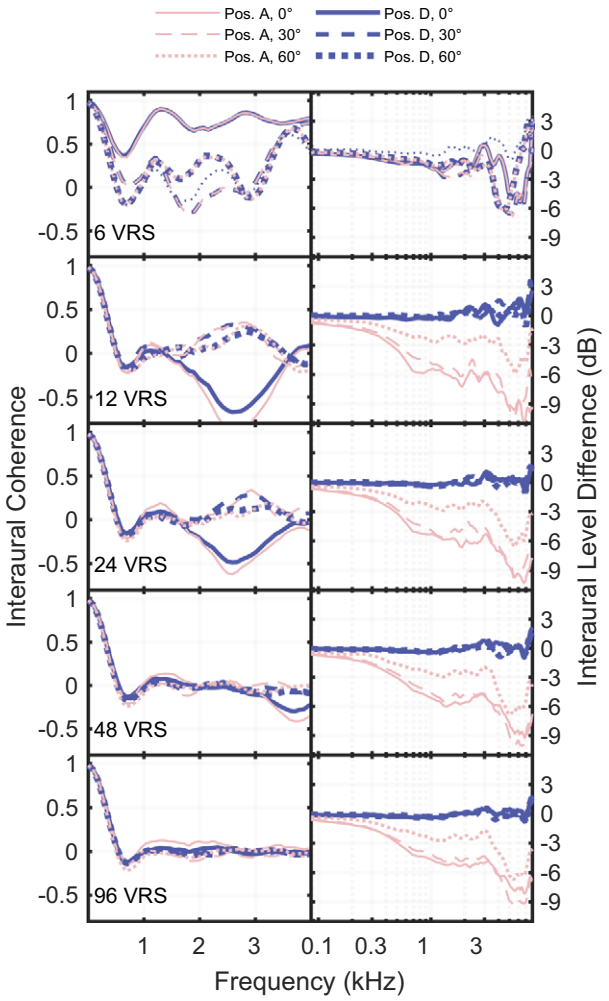


Fig. 4. Interaural coherence (left column) and interaural level difference (right column) for 6–96 virtual reverberation sources (VRS) (top to bottom) for different receiver orientations using the shoebox approach. Here, 0° refers to the orientation indicated in Fig. 2 (replot from Fig. 3), and 30° and 60° are a clockwise rotation towards the corridor. The additional thin traces in the top row correspond to the generalized approach.

coherence values, for which perceptual discrimination is poor (e.g., [52]). Regarding ILDs, a good approximation is obtained using 12 VRS with the generalized approach and with either approach for 24 VRS and more. For six VRS, considerably smaller ILDs compared to 96 VRS are observed. In addition to being compatible with arbitrary geometry, a benefit of the generalized approach over the shoebox approach is the gradually increasing ILD for six VRS and a performance for 12 VRS, which is generally more similar to 96 VRS.

The results of the current technical evaluation support the perceptual evaluation in [33] showing that 24 VRS are sufficient for rendering anisotropic late reverberation in the tested conditions.

2 LATE REVERBERATION IN COUPLED ROOMS

The same general underlying room acoustics simulation [26] as described in Sec. 1.1 was used here to simulate late

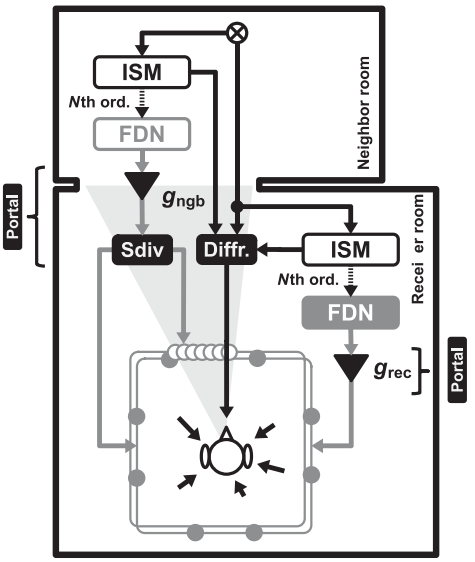


Fig. 5. Block diagram of the coupled-rooms algorithm with a sound source "⊗" in the neighbor room (ngb). Small arrows around the listener indicate image-source model (ISM) reflections arriving from their respective directions. The highest ("Nth") order image-source reflections serve as feedback delay network (FDN) inputs. All gray parts are related to late reverberation. The circles on the square outlines around the listener's head indicate the virtual reverberation sources (VRS). According to the solid angle of the door to the neighbor room (light gray), the directions of the neighbor room VRS (open gray circles) and the receiver room (rec) VRS (closed gray circles) are adapted. All dark gray filled processing steps are part of the portal.

reverberation in coupled rooms. The general approach was to divide the geometry into smaller proxy shoebox subvolumes (rooms). For each proxy shoebox, the single-room algorithm was run in parallel. The concept of portals (e.g., [38]) was used to handle (diffuse field) sound transmission between rooms, "visibility" of images sources, and spatial mapping of VRS to render late reverberation. Here, a portal is considered as a collection of signal-processing units representing the simplified effects of diffraction (for early reflections) and energy transfer (for late reverberation) through an aperture with a limited size opening area in comparison to the overall surface area of the rooms. For simplicity, the current coupled-room extension was restricted to two coupled rooms. Furthermore, the simplified proxy rooms were assumed to share an infinitely thin boundary in which the portal resides. The portal was a single rectangular (door) opening.

2.1 Extension to Coupled Rooms

A block diagram of the suggested coupled-room algorithm is provided in Fig. 5. The room in which the receiver resides, is referred to as "receiver room" and the other as "neighbor room."

2.1.1 Early Reflections

Although the focus was on late reverberation, the handling of early reflections had to be extended to account for coupled rooms, which was done in a strongly simplified

way. Image sources are subject to validity and visibility determination [18] as required for arbitrary geometry, if the sound path involves surfaces in the room that do not contain the source, i.e., if the sound path passes the portal.

Two cases were considered. If the source is in the neighbor room, all image sources from the wall containing the door are omitted. All other image-source paths through the portal that are either unobstructed or in most cases are obstructed by the "door wall" are constructed. In this latter case, a virtual diffraction path node was generated at the door outline, broadly following the concept of secondary sources (e.g., [53, 54]). The frequency-dependent attenuation caused by diffraction by an aperture was simulated in a strongly simplified manner using a high-shelving filter with progressively strong attenuation for increasing deflection angles. No high-frequency attenuation was applied if an unobstructed path existed through the portal. Higher-order low-frequency effects of diffraction were simulated by a low-frequency shelving filter. If the source is in the receiver room, image sources that would occur in the door surface are neglected.

Sound paths into the neighbor room were further simplified. A proxy source is placed at the center of the door surface, and image sources for the neighbor room are calculated originating from this position. These neighbor-room image sources are then treated exactly as if the direct source was placed inside the neighbor room. By this, for simplicity, diffraction by passing the portal is only applied once for each image source. As a further simplification, higher-order image sources created by image sources of the respective other room are neglected.

2.1.2 Late Reverberation

Reverberation in coupled rooms can be described by multiple decay processes in sub-volumes (e.g., [43]). The idea is that acoustic power is continuously transmitted between the volumes through the coupling surface. The power transfer depends on the coupling surface area and equivalent absorption areas of the rooms. [43] provides explicit formulations for energy decay curves (EDCs) for each room consisting of the sum of two exponential decay processes. Simplifying this idea, it is assumed here that each of the rooms has an independent decay process simulated with a respective FDN. Both contributions are superimposed. The relative amplitude of the two FDN outputs has to be determined in order to create a correct overall EDC. Each FDN is fed with the output of the highest-order image sources of the underlying room (like in the single-room simulation), as shown in the block diagram in Fig. 5. For simplicity, diffraction was neglected here before feeding the signals into the FDN.

Based on [43], gain factors for the FDNs were derived. In the following, S_d denotes the door surface area, and A_rec and A_ngb denote the total equivalent absorption areas in the receiver and neighbor room, respectively, averaged across frequencies. For the reverberation of the neighbor room, the gain factor g_ngb represents the amount of acoustic power transmitted through the door, normalized by the total acoustic power absorbed by all surfaces:

$$g_\mathrm{ngb} = \begin{cases} \sqrt{\frac{S_\mathrm{d}}{A_\mathrm{ngb}+S_\mathrm{d}}}, & \text{source in ngb room} \\ \sqrt{\frac{S_\mathrm{d}^2}{(A_\mathrm{ngb}+S_\mathrm{d})(A_\mathrm{rec}+S_\mathrm{d})}}, & \text{otherwise} \end{cases} \tag{8}$$

Eq. (8) represents the notion of power transmission from the receiver room to the neighbor room and back. Similarly, for the reverberation of the receiver room, the gain factor g_rec depends on losses to the neighbor room:

$$g_\mathrm{rec} = \sqrt{\frac{A_\mathrm{rec}}{A_\mathrm{rec}+S_\mathrm{d}}}. \tag{9}$$

The square roots in Eqs. (8 and 9) reflect that the factors are applied on the sound-pressure scale, rather than on the sound energy.

To correctly represent the spatial directions from which reverberation from the receiver and neighbor room can arrive, the VRS positions were spatially warped, depending on the relative position of the receiver and door. According to the direction and solid angle covered by the door, the 12 VRS directions are shifted such that they either cover directions only "outside" the door to render the late reverberation from the receiver-room or only "inside" the door to render the late reverberation from the neighbor-room.

The gray shaded area in Fig. 5 indicates the solid angle of the door, as observed from the receiver. Using such a spatial warping, a receiver in the center of the door would receive late reverberation in one hemisphere from the one room and in the other hemisphere from the other room, as desired. For larger distances of the receiver from the door, the 12 VRS of the neighbor room are collapsed to a single VRS to reduce computational cost. To account for reflections of the neighbor-room reverberation at the walls of the receiver room (contributing to perceived envelopment), a certain fraction of the neighbor-room reverberation is mixed to the receiver-room reverberation and thus rendered with the receiver-room VRS, resulting in spatial components from all directions (around the listener's head).

In realistic rooms with door frames or connecting passageways, additional inter-reflections, which are not covered by the current shoebox simplification and portal interface, might occur. To mimic the resulting increase in reflection densities, the FDN input for the neighbor room was filtered by a cascade of four all-passes based on the design suggested in [55] for artificial colorless reverberation. If the source resides in the neighbor room, such a filter was also applied to the receiver-room FDN input. The overall decay time of the all-pass cascade was set to a heuristic value of 400 ms for the coupled-room scenarios considered here. However, this effect is not expected and therefore not desired, if either the source or receiver is located close to the door. In the extreme case, this becomes equivalent to the case in which both are inside the same room. Therefore, a transition of the maximum value of 400 to 0 ms is applied, depending on the distances of the source and receiver, respectively, to the center of the door surface. Generally, this

decay time depends on the dimensions of the passageway and is set to zero as default.

2.2 Evaluation Setup

For the evaluation of the suggested method, a database of measured and synthesized binaural RIRs (BRIRs) was created. When possible, BRIR synthesis was performed using the same room and source-receiver configurations as for the real rooms.

2.2.1 Rooms

Overall, 32 BRIRs for several source-receiver configurations were measured in two environments at the University of Oldenburg. In the "staircase" environment, a corridor was connected to a staircase and a small and rather dry room. In the "office" environment, an office room was coupled with a long corridor. Five rooms were involved:

- Room S: A staircase, covering four floors in total, with uniform and very smooth varnish coat (plaster) at all side walls and tiles on the floor and stairs. Doors to adjacent corridors are made of glass. Apart from the stairs and railings, the room has approximated shoebox dimensions of $7.24 \times 3.51 \times 12.60$ m^3 ($= 320.20$ m^3). Because of the wall materials and dimensions, this room has a high reverberation time of 3.9 s. Measurement points were only at the ground floor, which is coupled to room C.
- Room C: A corridor with the same wall and floor material as room S. Doors to adjacent rooms are wooden or made of metal-framed glass. The room geometry deviates from the shoebox shape by two rectangular recesses acting as entrances to other rooms. For the shoebox approximation, the following dimensions were used: $2.65 \times 7.60 \times 2.49$ m^3 ($= 50.15$ m^3). The measured reverberation time is about 0.9 s. The room is connected to rooms C and M.
- Room M: A small room ($5.00 \times 2.41 \times 2.92$ m3 $= 35.19$ m^3) with a felt carpet and acoustic tiles at the ceiling. Walls are made of sound-absorbing modules, diffusors, and glass windows. The room contains one large wooden desk. The reverberation time is about 0.3 s.
- Room O: An office room with plaster and one large glass window, a felt carpet on the floor, and acoustic tiles at the ceiling. The room is furnished with tables, chairs, shelves, and a sideboard and has the dimensions $4.43 \times 4.50 \times 3.00$ m^3 $= 59.81$ m^3. The reverberation time is about 0.4 s.
- Room W: A long corridor, connected to room O, with plaster and wooden doors at the walls, linoleum floor, and acoustic tiles at the ceiling. The corridor is not fully straight but has a bend at approximately 45°. The shoebox approximation has the dimensions $30.00 \times 1.94 \times 2.50$ m^3 $= 145.50$ m^3. The reverberation time is about 0.5 s.

Fig. 6 shows top views of the staircase and office environments. All marked points are potential source or receiver positions. Some example positions are indicated with a room label and number. In the following, a source-receiver configuration, labeled as, e.g., "S1 → C2," indicates that the source is located at position S1 and the receiver at C2.

2.2.2 BRIR Measurements

An omnidirectional loudspeaker based on a ring-radiator [56] was used to excite the rooms. For recording, the artificial head MK2 by Cortex with the corresponding measurement amplifier Manikin MK1 was used. The excitation signal was a logarithmic sine sweep [57] ranging from 50 Hz to 18 kHz, allowing removal of nonlinear harmonic distortions of the loudspeaker [58]. For each recording, BRIRs were obtained from averaging ten measurements in the office environment and 22 measurements in the staircase environment. Finally, the BRIRs were equalized by the spectrum of the loudspeaker. For further details, see [26].

2.3 Technical Evaluation

Fig. 7 shows examples of measured and synthesized BRIRs for two coupled-room conditions. The upper panels represent the situation of a source residing in the same room as the receiver (C2 → C1), whereas in the lower panels, the source resides in the neighbor room (S2 → C1) and is not visible to the receiver. It can be observed that the overall envelope shapes are similar between measured and synthesized BRIRs. For the invisible source, a more gradual onset in the BRIR is clearly visible, which is also represented in the simulation.

Fig. 8 shows energy decay curves calculated using the Schroeder backward integration method [55]. The EDCs for measured (gray) and synthesized BRIRs (black) match well in most cases, and the expected dual-slope decay can be clearly observed in all EDCs. However, an extra bending is visible in the late-decay parts for the measured BRIRs, which is more pronounced in cases in which the effect of room S is higher, depending on the source and receiver positions (see Fig. 6). For the configuration C1 → S2, for which the receiver is in room S, the deviation is largest. This can likely be attributed to the fact that room S (staircase) itself shows a curved (multi-slope) EDC, given that its relatively complex geometry with stairs and railings behaves like multiple coupled sub-volumes, strongly deviating from the shoebox approximation.

For further comparison, early decay times (EDTs) and late decay times (LDTs) were calculated from linear fits of the EDC using the level intervals $[-10, 0]$ and $[-25, -35]$ dB, respectively. Furthermore, the definition D50 and the clarity index C80, which are of relevance for speech intelligibility and musical transparency, respectively, were calculated from the BRIRs. The results for EDT, LDT, D50, and C80 from the BRIRs of all 32 source-receiver configurations are shown as scatter plots in Fig. 9. The Pearson correlation coefficient between the measured and synthesized conditions is given in the corresponding panels. For all measures, the correlations are highly significant

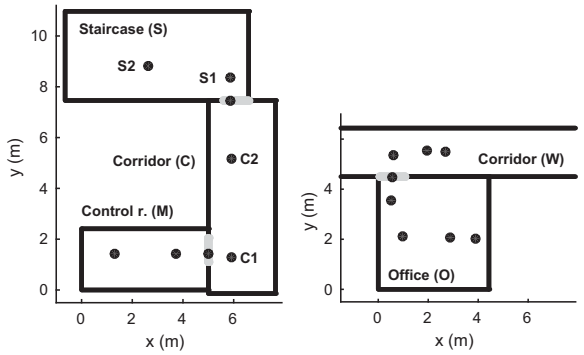


Fig. 6. Floor plans of the two coupled-room environments used for the evaluation (shoebox approximations): (a) the "staircase" environment and (b) "office" environment. Light gray bars indicate doors. Black circles indicate possible source and receiver positions. For some example source-receiver conditions in the staircase scenario, positions are labeled using room identifiers and numbers. See text for detailed room and position descriptions.

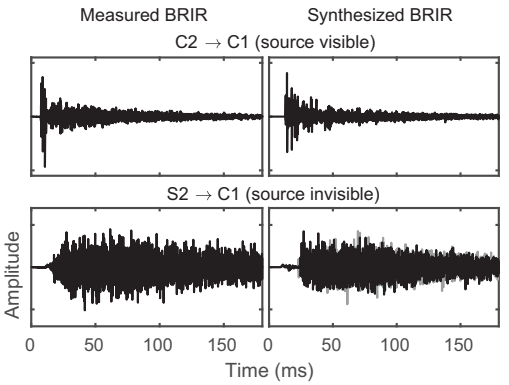


Fig. 7. Qualitative comparison of time signals of measured and synthesized binaural room impulse responses (BRIRs) for two example coupled-room conditions.

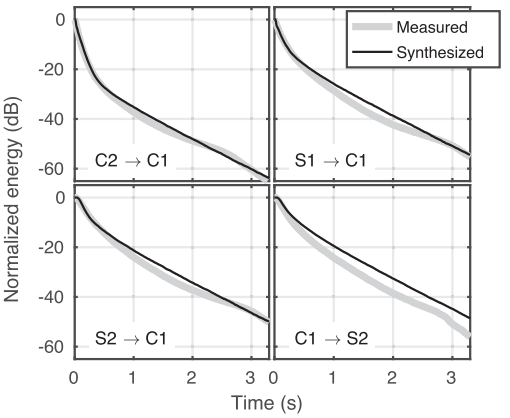


Fig. 8. Comparison of energy decay curves of measured and synthesized BRIRs for four example coupled-room conditions.

($p < 0.001$) and a good overall accordance was achieved. The data for D50 and C80 are more scattered than for EDT and LDT, however, still show a high corelation overall. Both, definition and clarity also tend to be higher for synthesized BRIRs than for the measured BRIRs. For future investigations, a comparison of binaural measures, such as IC and lateral fraction may be of interest. Here, a strong dependence on the receiver position is likely.

2.4 Perceptual Evaluation

2.4.1 Method

2.4.1.1 Listeners Eight listeners who were not Deaf or hard of hearing (one female and seven male; age 27–44 years) participated in the experiment. All participants were working in the field of (virtual) acoustics or hearing research and had experience in performing psychoacoustic experiments. Accordingly, they can be considered "expert listeners."

2.4.1.2 Procedure The participants' task was to rate perceived differences between measured and synthesized conditions for specific source-receiver configurations with respect to nine perceptual attributes selected from the Spatial Audio Quality Inventory [59, 60]: tone color, metallic tone color, distance, width, reverberation level, reverberation time, envelopment (by reverberation), loudness, and naturalness. The attributes "distance" and "width" referred specifically to the sound source itself. Each attribute consists of the name, a circumscription, and scale end labels of a bipolar scale (e.g., less pronounced–more pronounced; for further details, see [59]). Additionally, the overall difference was assessed on a unipolar scale (none–very large). The choice of the current Spatial Audio Quality Inventory subset was a trade-off between measurement time and relevance for further development.

Ratings were performed by adjusting a slider on a graphical user interface. Values ranged from 0 to 100 for the unipolar scale and −50 to 50 for bipolar scales. The acoustic material was presented in randomized order. First, the overall perceived difference between measured and synthesized conditions was rated. If it was larger than zero, the rating of the different attributes was performed. During the introduction given by the experimenter, all perceptual attributes including their original definitions were presented to the participants and potential misunderstandings were clarified. Additionally, the attribute definitions were displayed on the screen before each rating as a reminder.

2.4.1.3 Apparatus and Stimuli Participants were seated in a sound-attenuating listening booth. Sennheiser HD 650 headphones were used, driven by an RME ADI-2 soundcard at a sampling rate of 44.1 kHz. The sounds were generated and controlled using MATLAB. The headphone transfer function was equalized. The same source-receiver configurations as in the technical evaluation were used (C2 → C1, S1 → C1, S2 → C1, C1 → S2; see Fig. 7 for the geometric layout). Two different dry source signals were convolved with the BRIRs: a speech and music signal. The speech signal contained male and female spoken example sentences from the "Oldenburg Sentence Test" [61] and was approximately 30 s in duration. The music signal contained a 20-s long recording of a slapped bass guitar with wide frequency range and transient parts. For each rating, the stimuli were

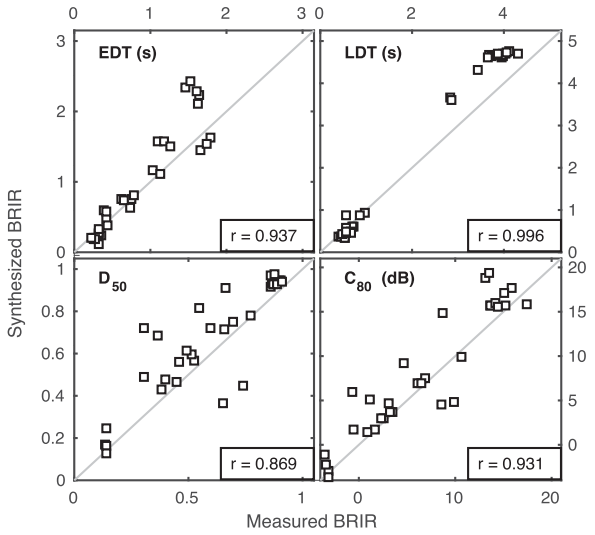


Fig. 9. Correlation between four single-value room acoustical parameters for measured and synthesized BRIRs: early decay time (EDT), late decay time (LDT), definition (D50), and clarity index (C80). See text for further details. The Pearson correlation coefficient r is given in each panel.

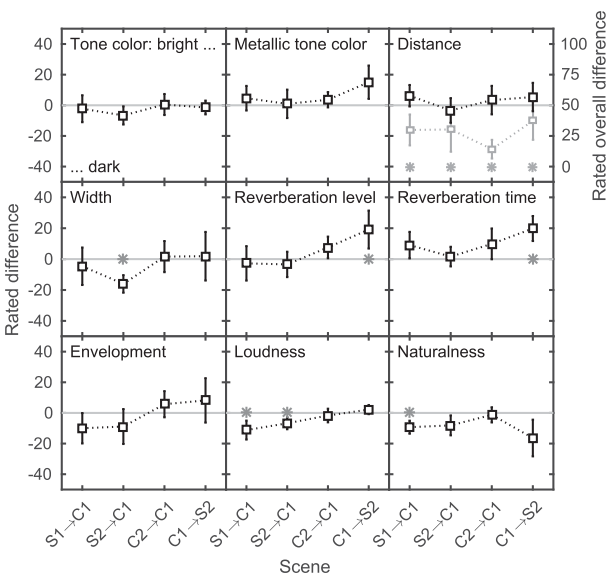


Fig. 10. Rated differences between measured and synthesized BRIRs with respect to different perceptual attributes (panels), plotted against source-receiver configurations, averaged across participants and source signals (speech and music). Error bars indicate inter-individual standard deviations. Depending on the attribute, ordinate scales range from "less pronounced" to "more pronounced" or semantically fitting descriptors. The asterisk symbols indicate significant differences from zero. The smaller gray symbols in the upper right panel show the ratings for the overall difference (using the right-hand scale) with the corresponding light gray asterisk symbols.

presented in a loop. Circular convolution ensured seamless looped playback. Participants could instantly switch between the stimuli during playback.

2.4.2 Results

Fig. 10 shows the results for the rated perceived differences between the measured and synthesized conditions, averaged across participants and source signals (speech and music). In each panel, mean values are plotted against source-receiver configurations with error bars indicating the inter-individual standard deviations. Overall differences (indicated by the smaller gray symbols in the upper right panel) were perceived but rated all within the lower half of the range (indicated on the right-hand scale). The maximum average score was 37.4 for C1 → S2. The smallest average difference (score: 14.2) was rated for the case in which source and receiver resided in the same room (C2 → C1). One participant rated the overall perceived differences with a score of 0 for this configuration, for both speech and music.

With only few exceptions, the perceived differences in tone color, metallic tone color, distance, and loudness are quite small. For the rated differences in reverberation time and level, the good match in the EDCs for configurations S1 → C1, S2 → C1, and C2 → C1 is represented, whereas some (significant) mismatch is observed for C1 → S2. Metallic tone color and naturalness appear to be correlated in such a way that a pronounced metallic tone color has one of the strongest impacts on (un)naturalness. Differences for both attributes were perceived to be mostly small and, again, largest for C1 → S2. Inter-individual standard deviations were mostly small, indicating that ratings were consistent across participants. The largest standard deviations occurred for source width and envelopment.

A paired-samples t test with Bonferroni correction for multiple comparisons was performed to assess whether the average scores were significantly different ($p < 0.05$) from zero (indicated by the asterisk symbols in Fig. 10). Although the overall perceived differences were always significant, most conditions for the perceptual attributes were not significantly different from zero.

3 DISCUSSION

A computationally efficient method to render late reverberation in rooms with inhomogeneously distributed absorption coefficients and coupled rooms was proposed and implemented in the perceptually plausible room acoustics simulator (RAZR, [26]). It was demonstrated that late reverberation can be represented by simplified simulation methods and limited spatial resolution with overall good agreement of technical and perceptual measures. Thus, the computational cost for simulating late reverberation and demands for spatially rendering late reverberation can be kept low, which is particularly important in the context of real-time VAEs.

Although a particular room acoustics simulation method featuring an FDN was used, the current results are applicable more generally to different simulation approaches for late reverberation, provided a number of incoherent late reverberation signals is generated.

3.1 Spatial Subsampling

A direction-based warping method was proposed to spatially subsample the different absorption coefficients at room boundaries by deriving mapping weights for the available VRS using only vector operations together with the well-established VBAP method. Moreover, for axis-aligned shoebox rooms (orthogonal walls), VBAP simplifies to just using the absolute values of the corresponding elements of the warped VRS vectors (per octant), dramatically reducing computational complexity. For many applications in which such a shoebox geometry is sufficient, this method thus offers very low computational complexity.

However, for low numbers of VRS and when the direction of the VRS coincide with the surface normals that form the vector base for VBAP in each octant, the method becomes invariant to the solid angle under which a boundary occurs. This is obvious in the current study for six VRS, in which, e.g., the IC is independent of position (see Fig. 3, positions A and D). The method could be further improved, mainly for low numbers of VRS, by integrating an estimate of the FOV of each boundary and the average FOV covered by each VRS. Improvements would be observed mainly in position D and for the lowest number of six VRS, which cannot be recommended for application in 6-DOF VAEs, anyhow.

As an extension for arbitrary (triangle-based) geometry, the suggested (more general) method based on ray tracing with a moderate number of deterministic directions around the listener can be used, if a "proxy" shoebox room approximation of the underlying room geometry is not sufficient. Using this method, different absorption coefficients can be applied to different triangles, enabling inhomogeneous distribution of absorption coefficients on the boundary planes. In contrast, the simple shoebox geometry used so far only represents averaged absorption coefficients in six orthogonal directions. Future developments for arbitrary geometry can be based on extension to stochastic ray tracing including temporal averaging methods. Alternatively, rasterization and image down-sampling methods after projection and rendering of the environment to a cube or spherical map, as frequently used in computer graphics/vision might be applied. Such more-advanced methods offer the possibility to define the spatial distribution of absorption coefficients using textures, independent of triangle geometry.

3.2 Coupled Rooms

The suggested coupled rooms method strongly simplifies the underlying problem by assuming independent decay processes in the two rooms. This simplification enables the use of two separate FDNs (or generally late reverberation models) and linear superposition of the outputs, considering the solid angle covered by the door opening and the surrounding room for the spatial mapping. The key assumption is that the opening is small in comparison to the surface areas of the rooms, although it is large enough so that no (or at least no frequency-dependent) diffuse field sound transmission loss of a "large aperture" can be assumed (e.g., [62]). Thus, without further evaluation, the current method is suited for typical indoor scenarios with rooms connected by doors, reflecting a class of highly relevant acoustic conditions (e.g., [29]). Future research might be directed to a more accurate simulation of the underlying acoustic coupling by, e.g., using coupled or grouped FDNs [63]. Moreover, the current spatial mapping of two (or more) discrete reverberation processes, which so far accounts for the direction- and frequency-dependent reverberation level, might be extended to anisotropic reverberation times (e.g., [16]) also suited to account for effects of coupled sub-volumes.

An aspect for further computational optimization is to avoid the current spatial warping of the VRS. In the current concept, the number of VRS would increase with the number of coupled rooms. Alternatively, a fixed number of VRS might be used, and a spatial subsampling method for the solid angle covered by the door could be employed, similar to the proposed rendering for inhomogeneous boundary conditions.

3.3 Physically-Based FDN Parameters

In line with a perceptual evaluation in [33], the current analysis of IC and ILDs indicates that a relatively low number of 12 to 24 VRS is sufficient to render late reverberation for inhomogeneous and homogeneous boundary conditions. For high computational efficiency, an according low number of FDN channels is desirable. However, the lower the number of FDN channels, the more pronounced recurrent temporal features occur in the FDN output. To avoid artifacts for low numbers of FDN channels, a high number of FDN channels (e.g., 96), independent of the number of VRS and downmixing, can be used in offline simulations.

For real-time applications, a perceptually optimal choice of FDN parameters for a low number of FDN channels has to be found. Although several studies optimized FDNs using, e.g., time-varying parameters or considering mode density (e.g., [64, 44]), RAZR uses a physically-based design of the FDN, in which the delays are derived from the dimensions of the room. This is also motivated by interpreting the FDN as a rough approximation of radiance transfer [65]. The current default setting as suggested and evaluated in [26] uses 12 VRS for lowest computational complexity. According to the current finding, a representation of random sound directions travelling through the room has to be found for 24 VRS. This higher number also allows for a better sampling of the distribution of sound travelling times between boundaries as, e.g., calculated from radiance transfer, and can be assumed to generally improve the perceptual quality of late reverberation.

4 CONCLUSION

Highly efficient methods for rendering anisotropic late reverberation in coupled rooms and rooms with inhomogeneous distribution of absorption coefficients were proposed. The methods use a limited number of VRS highly suited for application in (interactive) VAEs. For a "proxy" shoebox representation of arbitrary room geometry, the (av-

erage) absorption coefficients assigned to each boundary are smoothly mapped to a desired number spatially equally distributed VRS, depending on the receiver position in the room. For an axis-aligned proxy shoebox, the required operations can be strongly reduced, and computational efficiency further improved. For arbitrary geometry, an extension of the concept using a simplified deterministic ray tracing approach and spatial subsampling was suggested. Technical evaluation in terms of IC and ILD suggests that across multiple positions in a room with strongly inhomogeneous absorption properties, a number of 24 VRS can suffice for accurate reproduction in 6-DOF applications.

To account for anisotropic late reverberation in coupled rooms, VRS were spatially distributed to represent the solid angle covered by a door connection and the remaining room surface. In the assessed two room scenario, the VRS were driven by two independent FDNs representing the decay process of each room, mapped to 12 VRS each. The method is suited to account for a smooth onset of the impulse response for dual-slope decays and conditions with an obstructed source in the neighbor room. A technical and perceptual evaluation showed a good agreement with measured BRIRs.

The room acoustics simulator RAZR [26] including the current extensions is freely available at www.razrengine.com.

5 ACKNOWLEDGMENT

This work was funded by the Deutsche Forschungsgemeinschaft, DFG – Project-ID 352015383 – SFB 1330 C5 and DFG SPP Audictive – Project-ID 444827755.

6 REFERENCES

[1] J. H. Rindel, "The Use of Computer Modeling in Room Acoustics," *J. Vibroeng.*, vol. 3, no. 4, pp. 219–224 (2001 Feb.).

[2] N. Raghuvanshi and J. Snyder, "Parametric Directional Coding for Precomputed Sound Propagation," *ACM Trans. Graph.*, vol. 37, no. 4, paper 108 (2018 Aug.). https://doi.org/10.1145/3197517.3201339.

[3] H. Hu, L. Zhou, and and Z. Wu H. Ma, "HRTF Personalization Based on Artificial Neural Network in Individual Virtual Auditory Space," *Appl. Acoust.*, vol. 69, no. 2, pp. 163–172 (2008 Feb.). https://doi.org/10.1016/j.apacoust.2007.05.007.

[4] M. Cord, D. Baskent, S. Kalluri and B. C. Moore, "Disparity Between Clinical Assessment and Real-World Performance of Hearing Aids," *Hear. Rev.*, vol. 14, pp. 22–26 (2007 Jun.).

[5] J. Jerger, "Ecologically Valid Measures of Hearing Aid Performance," https://order.starkeypro.com/pdfs/sas/Starkey_Audiology_Series_v1i1.pdf (2009).

[6] K. M. Miles, G. Keidser, K. Freeston et al., "Development of the Everyday Conversational Sentences in Noise Test," *J. Acoust. Soc. Am.*, vol. 147, no. 3, pp. 1562–1576 (2020 Mar.). https://doi.org/10.1121/10.0000780.

[7] B. U. Seeber, S. Kerber, and E. R. Hafter, "A System to Simulate and Reproduce Audio–Visual Environments for Spatial Hearing Research," *Hear. Res.*, vol. 260, no. 1–2, pp. 1–10 (2010 Feb.). https://doi.org/10.1016/j.heares.2009.11.004.

[8] F. Pausch, L. Aspöck, M. Vorländer and J. Fels, "An Extended Binaural Real-Time Auralization System With an Interface to Research Hearing Aids for Experiments on Subjects With Hearing Loss," *Trends Hear.*, vol. 22, 18 pages (2018 Oct.). https://doi.org/10.1177/2331216518800871.

[9] G. Grimm, J. Luberadzka, and V. Hohmann, "A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology," *Acta Acust. united Acust.*, vol. 105, no. 3, pp. 566–578 (2019 May/Jun.). https://doi.org/10.3813/aaa.919337.

[10] M. Schutte, S. D. Ewert, and L. Wiegrebe, "The Percept of Reverberation Is Not Affected by Visual Room Impression in Virtual Environments," *J. Acoust. Soc. Am.*, vol. 145, no. 3, pp. EL229–EL235 (2019 Mar.). https://doi.org/10.1121/1.5093642.

[11] P. Luizard, B. F. G. Katz, and C. Guastavino, "Perceptual Thresholds for Realistic Double-Slope Decay Reverberation in Large Coupled Spaces," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. 75–84 (2015 Jan.). https://doi.org/10.1121/1.4904515.

[12] W. Lachenmayr, A. Haapaniemi, and T. Lokki, "Direction of Late Reverberation and Envelopment in Two Reproduced Berlin Concert Halls," presented at the *140th Convention of the Audio Engineering Society* (2016 May), paper 9503.

[13] D. Romblom, C. Guastavino, and P. Depalle, "Perceptual Thresholds for Non-Ideal Diffuse Field Reverberation," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3908–3916 (2016 Nov.). https://doi.org/10.1121/1.4967523.

[14] B. Alary, P. Massé, V. Välimäki, and M. Noisternig, "Assessing the Anisotropic Features of Spatial Impulse Responses," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*, pp. 43–48 (Paris, France) (2019 Sep.). https://doi.org/10.25836/sasp.2019.32.

[15] S. D. Ewert, O. Buttler, and H. Hu, "Computationally Efficient Parametric Filter Approximations for Sound-Source Directivity and Head-Related Impulse Responses." in *Proceedings of the International Conference on Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–6 (2021 Sep.). https://doi.org/10.1109/I3DA48870.2021.9610923.

[16] B. Alary, A. Politis, S. Schlecht and V. Välimäki, "Directional Feedback Delay Network," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 752–762 (2019 Oct.). https://doi.org/10.17743/jaes.2019.0026.

[17] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950 (1979 Apr.). https://doi.org/10.1121/1.382599.

[18] J. Borish, "Extension of the Image Model to Arbitrary Polyhedra," *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836 (1984 Jun.). https://doi.org/10.1121/1.390983.

[19] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique," *J. Sound Vibr.*, vol. 8, no. 1, pp. 118–125 (1968 Jul.). https://doi.org/10.1016/0022-460X(68)90198-3.

[20] M. Vorländer, "Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing/Image-Source Algorithm," *J. Acoust. Soc. Am.*, vol. 86, no. 1, pp. 172–178 (1989 Jul.). https://doi.org/10.1121/1.398336.

[21] M. Hodgson and E.-M. Nosal, "Experimental Evaluation of Radiosity for Room Sound-Field Prediction," *J. Acoust. Soc. Am.*, vol. 120, no. 2, pp. 808–819 (2006 Aug.). https://doi.org/10.1121/1.2216559.

[22] S. Siltanen, T. Lokki, S. Kiminki and L. Savioja, "The Room Acoustic Rendering Equation," *J. Acoust. Soc. Am.*, vol. 122, no. 3, pp. 1624–1635 (2007 Sep.). https://doi.org/10.1121/1.2766781.

[23] J.-M. Jot and A. Chaigne, "Digital Delay Networks for Designing Artificial Reverberators," presented at the *90th Convention of the Audio Engineering Society* (1991 Feb.), paper 3030.

[24] S. J. Schlecht, and E. A. P. Habets, "Scattering in Feedback Delay Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 1915–1924 (2020 Jun.). https://doi.org/10.1109/taslp.2020.3001395.

[25] D. Schröder, "Physically Based Real-Time Auralization of Interactive Virtual Environments," Ph.D. dissertation, Institute of Technical Acoustics, RWTH Aachen University, Berlin, Germany (2011 Feb.).

[26] T. Wendt, S. van de Par, and S. D. Ewert, "A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation," *J. Audio Eng. Soc.*, vol. 62, no. 11, pp. 748–766 (2014 Nov.). https://doi.org/10.17743/jaes.2014.0042.

[27] F. Brinkmann, L. Aspöck, D. Ackermann et al., "A Round Robin on Room Acoustical Simulation and Auralization," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2746–2760 (2019 Apr.). https://doi.org/10.1121/1.5096178.

[28] L. Aspöck, and M. Vorländer, "Simulation of a Coupled Room Scenario Based on Geometrical Acoustics Simulation Models," *Proc. Mtgs. Acoust.*, vol. 36, no. 1, pp. 015002 (2019 May). https://doi.org/10.1121/2.0001041.

[29] S. van de Par, S. D. Ewert, L. Hladek, et al., "Auditory-Visual Scenes for Hearing Research," *Acta Acust.*, vol. 6, no. 55, 14 pages (2022 Nov.). https://doi.org/10.1051/aacus/2022032.

[30] V. Välimäki, and K. Prawda, "Late-Reverberation Synthesis Using Interleaved Velvet-Noise Sequences," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1149–1160 (2021 Feb.). https://doi.org/10.1109/taslp.2021.3060165.

[31] F. Jacobsen, and T. Roisin, "The Coherence of Reverberant Sound Fields," *J. Acoust. Soc. Am.*, vol. 108, no. 1, pp. 204–210 (2000 Jun.). https://doi.org/10.1121/1.429457.

[32] K. Hiyama, S. Komiyama, and K. Hamasaki, "The Minimum Number of Loudspeakers and Its Arrangement for Reproducing the Spatial Impression of Diffuse Sound Field," presented at the *113th Convention of the Audio Engineering Society* (2002 Oct.), paper 5674.

[33] C. Kirsch, J. Poppitz, T. Wendt, S. van de Par, and S. D. Ewert, "Spatial Resolution of Late Reverberation in Virtual Acoustic Environments," *Trends Hear.*, vol. 25, 17 pages (2021 Oct.). https://doi.org/10.1177/23312165211054924.

[34] C. Kirsch, J. Poppitz, T. Wendt S. van de Par, and S. D. Ewert, "Computationally Efficient Spatial Rendering of Late Reverberation in Virtual Acoustic Environments," *in Proceedings of the International Conference on Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–8 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/i3da48870.2021.9610896.

[35] J. M. Airey, J. H. Rohlf, and F. P. Brooks, "Towards Image Realism With Interactive Update Rates in Complex Virtual Building Environments," *SIGGRAPH Comput. Graph.*, vol. 24, no. 2, pp. 41–50 (1990 Mar.). https://doi.org/10.1145/91394.91416.

[36] T. Akenine-Möller, E. Haines, and N. Hoffman, *Real-Time Rendering* (CRC Press. New York, NY, 2019), 3rd ed. https://doi.org/10.1201/9781315365459.

[37] S. J. Teller, and C. H. Séquin, "Visibility Preprocessing for Interactive Walkthroughs," *SIGGRAPH Comput. Graph.*, vol. 25, no. 4, pp. 61–70 (1991 Jul.). https://doi.org/10.1145/127719.122725.

[38] D. Luebke, and C. Georges, "Portals and Mirrors: Simple, Fast Evaluation of Potentially Visible Sets," in *Proceedings of the Symposium on Interactive 3D Graphics*, pp. 105–106. (Monterey, CA) (1995 Apr.). https://doi.org/10.1145/199404.199422.

[39] C. Faller and J. Merimaa, "Source Localization in Complex Listening Situations: Selection of Binaural Cues Based on Interaural Coherence," *J. Acoust. Soc. Am.*, vol. 116, no. 5, pp. 3075–3089 (2004 Nov.). https://doi.org/10.1121/1.1791872.

[40] J. Grosse, F. Hungar, S. Klockgether S. van de Par, "Wahrgenommene Quellbreite Einer Lautsprecheranordnung in Abhängigkeit Der Physikalischen Quellbreite," in *Proceedings of the Fortschritte der Akustik*, pp. 898–901 (Nürnberg, Germany) (2015 March).

[41] A. Walther, and C. Faller, "Assessing Diffuse Sound Field Reproduction Capabilities of Multichannel Playback Systems," presented at the *130th Convention of the Audio Engineering Society* (2011 May), paper 8428.

[42] C. Kirsch, and S. D. Ewert, "Low-Order Filter Approximation of Diffraction for Virtual Acoustics," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 341–345 (New Paltz, NY) (2021 Oct.). https://doi.org/10.1109/WASPAA52581.2021.9632674.

[43] L. Cremer and H. A. Müller, *Die Wissenschaftlichen Grundlagen Der Raumakustik* (S. Hirzel Verlag, Stuttgart, Germany, 1978), 2nd ed.

[44] S. J. Schlecht, and E. A. P. Habets, "Time-Varying Feedback Matrices in Feedback Delay Networks and Their Application in Artificial Reverberation," *J. Acoust. Soc. Am.*, vol. 138, no. 3, pp. 1389–1398 (2015 Sep.). https://doi.org/10.1121/1.4928394.

[45] S. J. Schlecht and E. A. P. Habets, "Feedback Delay Networks: Echo Density and Mixing

Time," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 2, pp. 374–383 (2017 Sep.). https://doi.org/10.1109/taslp.2016.2635027.

[46] H. Wadell, "Volume, Shape, and Roundness of Quartz Particles," *J. Geol.*, vol. 43, no. 3, pp. 250–280 (1935 Apr.). https://doi.org/10.1086/624298.

[47] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466 (1997 Jun.).

[48] F. Brinkmann, A. Lindau, S. Weinzierl, et al. "The Fabian Head-Related Transfer Function Data Base," Technische Universität Berlin (2017 Feb.). https://doi.org/10.14279/depositonce-5718.

[49] R. G. Klumpp and H. R. Eady, "Some Measurements of Interaural Time Difference Thresholds," *J. Acoust. Soc. Am.*, vol. 28, no. 5, pp. 859–860 (1956 Sep.). https://doi.org/10.1121/1.1908493.

[50] J. Zwislocki, and R. S. Feldman, "Just Noticeable Differences in Dichotic Phase," *J. Acoust. Soc. Am.*, vol. 28, no. 5, pp. 860–864 (1956 Sep.). https://doi.org/10.1121/1.1908495.

[51] A. Brughera, L. Dunai, and W. M. Hartmann, "Human Interaural Time Difference Thresholds for Sine Tones: The High-Frequency Limit," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2839–2855 (2013 May). https://doi.org/10.1121/1.4795778.

[52] D. E. Robinson and L. A. Jeffress, "Effect of Noise Correlation on Binaural-Signal Detection," *J. Acoust. Soc. Am.*, vol. 34, no. 5, pp. 727–728 (1962 May). https://doi.org/10.1121/1.1937229.

[53] M. A. Biot, and I. Tolstoy, "Formulation of Wave Propagation in Infinite Media by Normal Coordinates With an Application to Diffraction," *J. Acoust. Soc. Am.*, vol. 29, no. 3, pp. 381–391 (1957 Mar.). https://doi.org/10.1121/1.1908899.

[54] U. P. Svensson, R. I. Fred, and J. Vanderkooy, "An Analytic Secondary Source Model of Edge Diffraction Impulse Responses," *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2331–2344 (1999 Oct.). https://doi.org/10.1121/1.428071.

[55] M. R. Schroeder, "Natural Sounding Artificial Reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223 (1962 Jul.).

[56] R. Kruse, A. Häußler, and S. van de Par, "An Omnidirectional Loudspeaker Based on a Ring-Radiator," *Appl. Acoust.*, vol. 74, no. 12, pp. 1374–1377 (2013 Dec.). https://doi.org/10.1016/j.apacoust.2013.04.018.

[57] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), paper 5093.

[58] S. Müller and P. Massarani, "Transfer-Function Measurement With Sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471 (2001 Jun.).

[59] A. Lindau, *Spatial Audio Quality Inventory (SAQI)* (Technische Universität Berlin, Berlin, Germany, 2014). http://dx.doi.org/10.14279/depositonce-1.2.

[60] A. Lindau, V. Erbes, S. Lepa et al., "A Spatial Audio Quality Inventory (SAQI)," *Acta Acust. united Acust.*, vol. 100, no. 5, pp. 984–994 (2014 Sep.). https://doi.org/10.3813/AAA.918778.

[61] K. Wagener, T. Brand, and B. Kollmeier, "Entwicklung Und Evaluation Eines Satztests Für Die Deutsche Sprache III: Evaluation Des Oldenburger Satztests," *Zeitschrift für Audiologie*, vol. 38, no. 3, pp. 86–95 (1999).

[62] F. Sgard, H. Nelisse, and N. Atalla, "On the Modeling of the Diffuse Field Sound Transmission Loss of Finite Thickness Apertures," *J. Acoust. Soc. Am.*, vol. 122, no. 1, pp. 302–313 (2007 Jul.). https://doi.org/10.1121/1.2735109.

[63] O. Das and J. S. Abel, "Grouped Feedback Delay Networks for Modeling of Coupled Spaces," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 486–496 (2021 July). https://doi.org/10.17743/jaes.2021.0026.

[64] D. Rocchesso, and J. O. Smith, "Circulant and Elliptic Feedback Delay Networks for Artificial Reverberation," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 1, pp. 51–63 (1997 Jan.). https://doi.org/10.1109/89.554269.

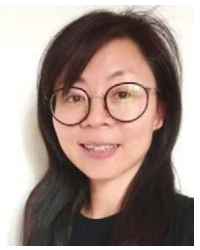
[65] H. Bai, G. Richard, and L. Daudet, "Late Reverberation Synthesis: From Radiance Transfer to Feedback Delay Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2260–2271 (2015 Dec.). https://doi.org/10.1109/taslp.2015.2478116.

THE AUTHORS


Christoph Kirsch


Torben Wendt


Steven van de Par


Hongmei Hu


Stephan D. Ewert

Christoph Kirsch studied Media and Acoustical Engineering in Mittweida, Germany, and received his Master's degree in Signal Processing and Acoustics in 2019 from Aalborg University, Denmark. Currently, he is working as a Ph.D. student at Carl von Ossietzky Universität, Oldenburg, Germany. His work is focused on efficient techniques for perceptually plausible acoustic rendering. His contribution regarding rendering and requirements on the spatial resolution of late reverberation won the best student paper award at the International Conference on Immersive and 3D Audio 2021.

•

Torben Wendt was born in Germany in 1986. He studied physics at the Christian-Albrechts-Universität zu Kiel, Germany, and at the Universität Oldenburg, Germany, and received a Ph.D. degree in 2018 from the Universität Oldenburg. For his Bachelor's thesis, he did research in basic binaural psychoacoustics. In his Master's thesis, he started working in the field of room acoustical simulations, which he continued as a Ph.D. student. Currently, he is working as a software developer.

•

Steven van de Par studied physics at the Eindhoven University of Technology, Eindhoven, The Netherlands, and there he received a Ph.D. degree in 1998 on a topic related to binaural hearing. As a Postdoctoral Researcher at the same university, he studied auditory-visual interaction and was a Guest Researcher at the University of Connecticut Health Center. In 2000, he joined Philips Research, Eindhoven, to do applied research on low–bit-rate audio coding, and music information retrieval. Since April 2010, he is professor in acoustics at the Carl-von-Ossietzky Universität Oldenburg, Germany, with a research focus on auditory perception and its application to virtual acoustics, vehicle acoustics and computational auditory-scene analysis. He has published various papers on binaural auditory perception, auditory-visual synchrony perception, audio coding, virtual acoustics, and computational auditory scene analysis.

•

Hongmei Hu received her Ph.D. in Information and Signal Processing at Southeast University (China) in 2008. She was a lecturer at Jiangsu University (China) afterward until she moved to the United Kingdom in 2012 and worked as a Marie Curie senior researcher at Southampton University from 2010 to 2012. She was promoted to a tenured associate professor in Jiangsu University in 2012. From 2013 to now, she has been a senior researcher in Carl von Ossietzky Universität Oldenburg (Germany). Her current research interests are mainly focusing on the speech perception and binaural hearing of bilateral cochlear implant (CI) users. Topics include CI fitting strategies based on subjective perception (psychoacoustic) and objective (Electroencephalography) measurements, advanced signal processing and electric modeling for CIs, and speech in noise perception of tonal and non-tonal languages. More recently, virtual acoustic rending techniques and application in (spatial) hearing research have become of interest.

•

Stephan D. Ewert studied physics and received his Ph.D. degree in 2002 from the Carl von Ossietzky Universität Oldenburg, Germany. During his Ph.D. project, he spent a 3-month stay as visiting scientist at the Research Lab of Electronics at the Massachusetts Institute of Technology (MIT), Cambridge, MA. From 2003 to 2005, he was Assistant Professor at the Centre of Applied Hearing Research at the Technical University of Denmark (DTU), Lyngby, Denmark. He re-joined Medizinische Physik at the Universität Oldenburg in 2005, and there he has been the head of the Psychoacoustic and Auditory Modeling Group since 2008. His field of expertise is psychoacoustics and acoustics with a strong emphasis on perceptual models of hearing and virtual acoustics. Dr. Ewert has published various papers on spectro-temporal processing, binaural hearing, and speech intelligibility. He also focused on perceptual consequences of hearing loss, hearing-aid algorithms, instrumental audio quality prediction, and room acoustics simulation.