



Computationally-efficient rendering of diffuse reflections for geometrical acoustics based room simulation

Stephan D. Ewert^{1,*} , Nico Gößling¹, Oliver Buttler¹, Steven van de Par², and Hongmei Hu^{1,2} 

¹ Medizinische Physik and Cluster of Excellence “Hearing4all”, Carl von Ossietzky Universität Oldenburg, Carl-von-Ossietzky-Straße 9-11, 26129 Oldenburg, Germany

² Akustik and Cluster of Excellence “Hearing4all”, Carl von Ossietzky Universität Oldenburg, Carl-von-Ossietzky-Straße 9-11, 26129 Oldenburg, Germany

Received 6 February 2024, Accepted 16 September 2024

Abstract – Geometrical acoustics is well suited for real-time room acoustics simulation and is often implemented using the image source model (ISM). One drawback of the ISM is its limitation to specular reflections, while sound scattering plays an important role in real environments. Here, computationally-efficient, digital-filter approximations are proposed to account for effects of non-specular scattered reflections in the ISM. For scattering at large surfaces such as room boundaries, each reflection is energetically split into a specular and a scattered part, based on the scattering coefficient. The scattered sound is coupled into a diffuse reverberation model. Temporal effects of the underlying surface scattering for an infinite ideal diffuse (Lambertian) reflector are derived and the resulting monotonic decay is simulated using cascaded all-pass filters. Effects of scattering and multiple (inter-) reflections caused by larger geometric structures at walls, and by objects in the room are accounted for in a highly simplified manner. A single parameter is used to quantify deviations from an empty shoebox room. The cumulated temporal effect of scattering along a reflection path is mimicked using cascaded all-pass filters adjusted to obtain a gamma-distribution-shaped envelope. The proposed method was perceptually evaluated with both music and pulse stimuli against dummy head recordings of real rooms. The results show a better agreement between the recording and the simulation for transient stimuli. In a technical evaluation, the temporal evolution of echo density showed a comparable profile for the suggested method and real rooms.

Keywords: Scattering, Virtual acoustics, Geometrical acoustics, Digital filters, Diffraction

1 Introduction

Room acoustics simulation has applications in various fields, ranging from room-acoustical planning and education, hearing research, rehabilitation and training (e.g., [1–3]), to virtual reality and artificial reverberation in video games (e.g., [4, 5]). Accordingly, the research field has gained increasing popularity [6–8] with a growing demand for real-time applicability and interactive virtual acoustic environments (e.g., [9–15]). These developments require computationally efficient simulation techniques. Geometrical acoustics (GA) is frequently used in this context and well suited to model direct sound and early reflections, assuming ray-like sound propagation and specular reflections.

In real rooms and environments, however, sound scattering plays an important role, which is not accounted for by straightforward application of GA, for example, using the image source model (ISM; [16, 17]) or ray tracing [18, 19]. One important scattering phenomenon is edge

diffraction which has been integrated into GA considering diffracted rays with “bended” propagation paths (e.g., [20–22]). While the underlying diffraction solutions involve computationally expensive integral expressions (e.g., [20–23]), computationally highly-efficient, digital filter-based diffraction solutions and approximations for GA were proposed in [24–26]. These filter-based techniques offer a physically-based simulation of diffraction from arbitrary infinite and finite wedges (and objects composed thereof). For complex geometries with many edges and objects, however, diffraction path tracing (e.g., [27, 28]) and detailed diffraction modelling become computationally expensive and modelling individual edges may not be perceptually required. Thus, more simplified or macroscopic approximations for combined scattering effects of several edges might be helpful. Approaches for objects using parametric filters estimated by machine learning have been studied in [29, 30].

Besides the above described effects of scattering at distinct large structures and objects, scattering also occurs at boundaries due to small-scale geometric or spatial impedance variations, which can be considered as surface

*Correspondence: stephan.ewert@uni-oldenburg.de

“roughness” and may be effectively described by the frequency-dependent (random-incidence) scattering coefficient [31]. Unlike GA specular reflections at boundaries, which are accounted for by the ISM, sound reflections in real rooms typically involve scattering at such rough surfaces, resulting in diffuse reflections. Diffusively reflected sound energy typically dominates after the third reflection order in real rooms [32]. The effect of diffuse reflections has been shown to be audible [33], and is obvious in comparisons of measured and simulated impulse responses (e.g., [34]), potentially related to the temporal evolution of echo density [35], necessitating consideration in virtual acoustic environments.

For example, approaches like extending the ISM by incorporating incoherent reflections from rough room surfaces [36], or by combining the ISM with a feedback delay network (FDN; [37]) as diffuse late reverberation model have been explored. Other approaches for diffuse late reverberation use decaying noise or specific noise sequences (e.g., [38, 39]). Spatio-temporal and spatial modulations in diffuse reverberation have been realized using FDNs (e.g., [40, 41]), with temporal and directional fluctuations affecting listener envelopment [42]. In [41, 43], the ISM was used for low-order specular reflections, while diffuse (later and non-specular) reflections were modelled in a simplified manner using a FDN with room-dependent parameters and spatial rendering. However, with this and similar processing schemes, the transition from purely specular to purely diffuse reflections occurs at a distinct, pre-defined reflection order. Conversely, a gradual transition is observed in real rooms, where scattering contributes to the diffuse sound field for all, including early (non-specular) reflections. Diffuse reflections can generally be handled in the context of GA by models using the bidirectional reflection distribution function (BRDF), known from computer graphics, to include diffuse parts, or for solely diffuse acoustic radiance transfer [44], generalizing acoustic radiosity methods (e.g., [45, 46]). For real-time rendering, however, either impulse responses can be generated from BRDF time-energy responses or frequency-domain implementations can be used [47]. Some authors [48, 49] combined acoustic radiance transfer with path tracing to account for diffuse reflections (and edge diffraction) in interactive acoustic environments, both generating impulse responses and using convolution and interpolation. In [49], late reverberation is separately implemented using a Schroeder-type artificial reverberator [50]. Sound scattering at surfaces is considered in scattering delay networks (SDNs; [51, 52]), which combine principles of GA and digital waveguide networks (to efficiently account for late reverberation, similar to FDNs). In ray tracing approaches, diffuse reflections can be modelled by spatially distributing the directions of reflected rays depending on the scattering coefficient in a frequency band [53, 54], and by collecting diffuse sound contributions from each point of reflection at the receiver [14, 55] (see also [56] for radiosity). Like acoustic radiance and radiosity, ray tracing provides time-energy responses and impulse responses can be obtained by modulating noise sequences, e.g., derived from a Poisson process [14, 55, 57], with the energy envelope.

For real-time room simulations, time-domain implementations (see [58]), making use of digital filters are overall desirable to limit latency and to improve computational efficiency. Ewert et al. [59] suggested a simplified digital-filter approximation to account for effects of scattering in the framework of GA, including a coupling of scattered sound reflections from the ISM into the diffuse reverberation model.

In artificial reverberators, temporal diffusion as an effect of scattered reflections (see, e.g., [60, 61]) is often an integral part, using computationally efficient implementations. Popular FDN-based artificial reverberators as, e.g., outlined in [6], include temporal spread via Schroeder-reverberators in the feedback loop. FIR filters were used in [62]. One basic idea of these filters is a faster increase of echo density, often referred to as “diffusion” in the context of reverberators, and is required particularly for highly efficient FDNs with low number of channels. Other examples in reverberators are “cluster filters” to process the early reflections [63] or nested all-pass filters [64]. In contrast to room acoustics simulation, where highly abstracted techniques such as an FDN are suited as a late reverberation model, in artificial reverberators such abstracted techniques might also account for early reflections. For real-time room acoustics simulation and interactive virtual acoustics, solutions that account for temporal diffusion, similar to those in artificial reverberators, might be applicable if they have a physically-based motivation to include the effects of sound scattering.

Besides the underlying methods to account for scattering, a critical question is under which conditions the inclusion of scattering effects in room acoustics is perceptually relevant. For example, regarding echo thresholds, no differences were found for reflections with and without temporal diffusion for music in [61], while for speech, offering more transients, diffuse reflections were less easily detected as separate echo, although effects were overall small. Perceptual effects of including diffuse reflections in room acoustics simulation should thus be evaluated for transient and less transient stimuli.

Taken together, computationally-efficient approximations for scattering at surfaces, and for sound paths involving partly obstructing larger scale geometric structures and objects in a room are desirable for real-time virtual acoustics. Parametric digital filters appear particularly suited to achieve the required low computational complexity. So far, filter solutions for scattering phenomena, such as discrete diffracted sound paths at edges and objects (e.g., [26]), have been thoroughly derived.

Here, we extend the study described in [59] by deriving and further evaluating a strongly simplified and computationally-highly efficient digital-filter approach to approximate the global effects of sound scattering and resulting diffuse reflections for room acoustic simulation. The goal is to i) provide a physically-based extension of the ISM approach for scattered reflections and realistic temporal evolution of echo density, and to ii) provide computationally-efficient simplifications aiming at perceptual plausibility.

Firstly, considering the frequency-dependent scattering coefficient [31] of a surface, we assume ideal diffuse (Lambertian) reflections (e.g., [44, 65]) in addition to the GA specular reflection in the ISM. For modelling diffuse reflections from (larger) surfaces such as the room boundaries, we derive a simplified temporal response of an infinite ideal diffuse reflector and propose splitting the energy of each image source (IS) into a specular and non-specular (diffuse) part using IS decomposition filters. The non-specular reflected part of each IS is then fed into an FDN serving as diffuse reverberation model ([41, 43]). As a consequence, the here suggested approach realistically generates (partly) diffuse reflections for all ISs orders. In the time domain, ideal diffuse reflections from a surface lead to a temporally distributed summed reflection, given that each surface point contributes with different travel time and attenuation. To approximate this effect, we propose the use of highly-efficient cascaded all-pass filters to temporally “smear” the non-specular parts of the ISs. The accompanying spatial spread is accounted for by mapping the diffuse reflections to the according directions of virtual reverberation sources used in the FDN spatialization [66, 67].

Secondly, complementing the existing physically-based filter solution for detailed modelling of discrete diffracted sound paths at edges and objects of [26], we suggest to account for the global effect of scattering, including multiple (inter-) reflections at larger geometric structures and by interior objects in the room in a highly simplified manner: We use a single parameter to quantify geometric deviations from an ideal empty shoebox room. Using this parameter, each IS reflection is temporally distributed using cascaded all-pass filters.

Both proposed methods to account for surface and object scattering were implemented in a state-of-the-art room acoustics simulation and were perceptually evaluated against dummy head recordings of real rooms using a subset of the spatial audio quality inventory [68]. As a technical evaluation, the temporal evolution of echo density in simulated impulse responses with the here suggested methods was compared to real rooms and an alternative state-of-the-art ray tracing implementation of surface scattering.

2 Diffuse reflections

2.1 Surface scattering

2.1.1 Non-specular reflection from an infinite wall

In order to characterize the effects of surface scattering, for simplicity, a single rigid infinite wall (no absorption; reflection coefficient $\rho = 1$) is assumed with the source and receiver located in front of the wall at the same point P , at a distance R from the wall (see Fig. 1). Following Fermat’s principle, the shortest reflection path corresponds to the specular reflection path, which is in this case perpendicular to the wall with total distance of $2R$. Assuming an

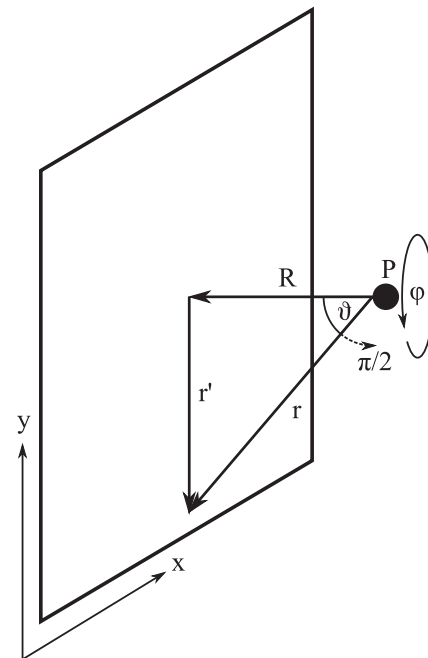


Figure 1. Contribution of non-specular reflections from an infinite wall. The source and receiver are located in the same point P at a distance R from the wall. They either radiate to or receive sound from all points at the wall surface with distance r and angles ϑ , φ .

omnidirectional (point) sound source and the speed of sound c , the direct sound arrives without any delay and distance attenuation at the receiver (collocated with the source), whereas the specular reflection arrives with a travel time of $t = 2R/c$ and with a distance attenuation of $1/(2R)^2$. With the source power P_s , the intensity of the specular reflection for this geometrical arrangement is

$$I_\sigma = \frac{P_s}{16\pi R^2}. \quad (1)$$

Further assuming ideal (Lambertian) diffuse reflections (e.g., [44, 65]), “non-specular” reflections from each point on the wall surface at a distance r also arrive at the receiver in addition to the specular reflection. The contributing intensities of the resulting diffuse sound sources for each reflection point on the surface can be calculated by Lambert’s cosine law. We base our derivation on the differential (outgoing) intensity, dI , of the ideal diffuse reflection given by [65]:

$$dI(r, \vartheta) = \rho B_0 \frac{\cos \vartheta}{\pi r^2} dS, \quad (2)$$

where B_0 is the irradiance caused by the source, dS is a differential surface segment, and $\rho = 1$ (assuming no absorption). The assumptions of a co-located source and receiver result in an equal angle of incidence and diffuse reflection ϑ , so that the irradiance on a surface segment caused by the source becomes

$$B_0 = \frac{P_s \cos \vartheta}{4\pi r^2}. \quad (3)$$

With the differential surface area in equation (2) expressed in spherical coordinates, $dS = r'dr'd\varphi$, using $r' = \sqrt{r^2 - R^2}$, the sound intensity at the receiver contributed by each point on the surface can now be written as

$$dI(r, \vartheta) = \frac{P_s \cos^2 \vartheta}{4\pi^2 r^4} r' dr' d\varphi. \quad (4)$$

Hence, integrating over $d\varphi$, the total intensity of the diffuse reflection I_δ of the wall surface is given by

$$I_\delta = \int_0^\infty \frac{P_s \cos^2 \vartheta}{2\pi r^4} r' dr'. \quad (5)$$

With $\cos \vartheta = R/r$ and replacing $r' = \sqrt{r^2 - R^2}$, this results in

$$I_\delta = \frac{P_s R^2}{2\pi} \int_R^\infty \frac{1}{r^5} dr = \frac{P_s}{8\pi R^2}. \quad (6)$$

Using equations (1) and (6), the ratio between the intensity I_δ of the diffuse reflection and the intensity of the specular reflection I_σ is

$$\frac{I_\delta}{I_\sigma} = 2 \quad (7)$$

or $10 \log(I_\delta/I_\sigma) \approx 3$ dB. It should be noted that this relation only holds for the here assumed highly simplified case and of an infinite wall. For finite walls, the ratio in equation (7) is smaller than 2. As an example, for a finite square wall with an edge length $2R$, with a geometrical arrangement as given in Figure 1, a ratio of $I_\delta/I_\sigma \approx 1.6$ would be observed (and $I_\delta/I_\sigma \approx 0.84$ for a smaller square wall with edge length R). Additionally, diffuse reflections are typically frequency-dependent and their intensity in relation to the specular reflection can be expressed using the scattering coefficient [31], depending on the size of the wall irregularities and roughness in relation to the wavelength of sound. Typically, scattering increases towards high frequencies, resulting in different spectra for the diffusively and specularly reflected sound.

The contribution from each point on the wall to the diffusively reflected sound arrives with a delay relative to the specular reflection at the receiver, caused by the longer path length $2r \geq 2R$. The contribution of each point on the wall to the reflected sound thus represents a *spatial spread*, while at the point of the receiver this consequently leads to a *temporal spread*, due to the different path lengths.

By assuming a step excitation from the sound source at $t = 0$, $P_{s,\text{step}}(t) = \mathcal{H}(t)P_s$ in equation (6), with \mathcal{H} denoting the Heaviside step function, and further using the time-of-arrival at the receiver $\tau = 2r/c$ from each point on the surface, the integration variable dr can be substituted by

$$dr = \frac{c}{2} d\tau, \quad (8)$$

and the diffuse intensity $I_{\delta,\text{step}}(t)$ becomes

$$\begin{aligned} I_{\delta,\text{step}}(t) &= \mathcal{H}\left(t - \frac{2R}{c}\right) \frac{8P_s R^2}{\pi c^4} \int_{\frac{2R}{c}}^t \frac{1}{\tau^5} d\tau \\ &= \mathcal{H}\left(t - \frac{2R}{c}\right) \frac{2P_s R^2}{\pi c^4} \left(\frac{c^4}{16R^4} - \frac{1}{t^4}\right). \end{aligned} \quad (9)$$

Consequently, the time dependence of the diffuse intensity can be expressed by defining the diffuse intensity impulse response $IR_\delta(t)$

$$IR_\delta(t) = \mathcal{H}\left(t - \frac{2R}{c}\right) \frac{8P_s R^2}{\pi c^4 t^5} \propto \frac{1}{t^5}, \quad (10)$$

$$I_{\delta,\text{step}}(t) = \mathcal{H}(t)P_s * IR_\delta(t), \quad (11)$$

where the asterisk denotes convolution. Figure 2a shows the (peak) normalized intensity impulse response against time for distances R of the source and receiver of 1, 5, and 10 m. As can be observed, the temporal spread increases with distance. With increasing R the lower integration boundary in equation (9) increases and $IR_\delta(t) \propto 1/t^5$ starts at a less steep point. This leads to a slower relative decrease for larger distances R . Figure 2b shows the same responses on a logarithmic intensity scale, normalized to the intensity of the specular reflection for each distance, respectively. It is obvious that the peak of the decay curves drops with distance R (in relation to the longer decay time) and is considerably lower than that of the specular reflection. It should be noted that for more directed, and not ideally diffuse scattering (not shown here), contributions from farther locations on the reflecting surface will be weaker and accordingly the diffuse reflection will generally decay faster (until the decay disappears for a specular reflection).

2.1.2 Effective implementation of temporal spread

To quantify the resulting temporal spread of the diffuse reflection, the monotonic temporal decay derived above, expressed in dB as shown in Figure 2b, is similar to an energy decay curve (EDC; e.g. [69]) as used for room reverberation. The temporal effect of surface scattering can thus conceptually be interpreted as a local energy decay, following each specular reflection. For simplicity, assuming an exponential energy decay (a linear decay on the logarithmic intensity scale of Fig. 2b), a local decay time T_s can be defined to approximate the initial decay of the above derived power function. Fitting the initial 10-dB decrease (intensity reduction to 1/10) of the EDC by an exponential decay, the resulting decay time T_s for a decrease of 60 dB can be derived from equations (9) and (10) as $T_s = 6(\sqrt[5]{10} - 1)2R/c \cong 7R/c$. The assumption of an exponential decay appears reasonable for the initial 20 dB of the decay curve (see left panel of Fig. 2), however, underestimates the temporal spread overall.

To approximate a pressure impulse response of the diffuse reflection for auralization, additional assumptions have to be made. The above derivation is intensity based and is valid for walls where each point acts as an ideal Lambertian

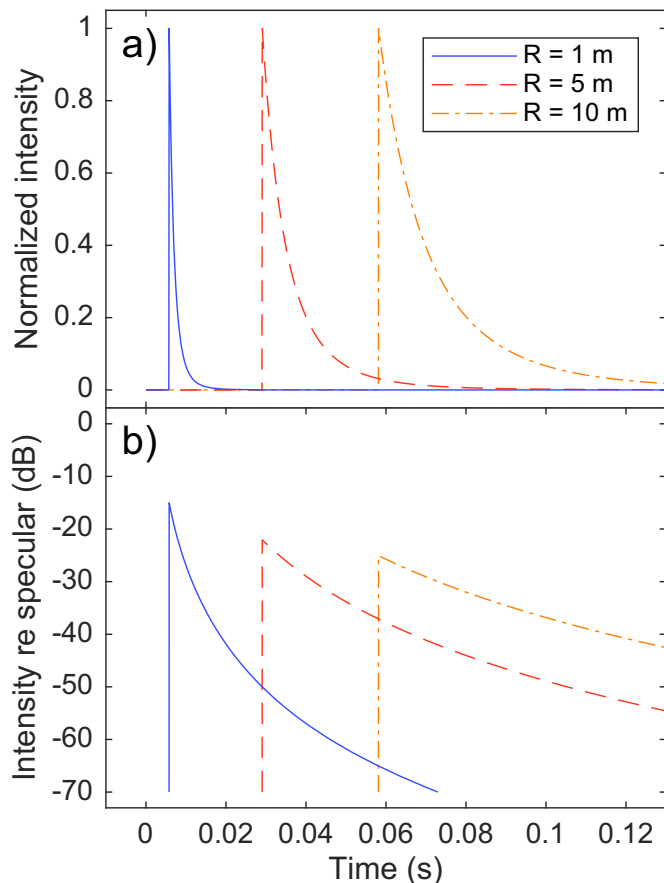


Figure 2. a) shows the (peak) normalized intensity of the diffuse reflections over time for distances 1, 5, and 10 m. b) shows the same decay curves on a logarithmic intensity scale in dB, with each curve normalized to the specular reflection intensity for the respective distance R .

reflector. The approximated exponential intensity decay is converted to a corresponding pressure decay (see, e.g., [36, 57]), representing the envelope function of the equivalent impulse response. Regarding the fine structure of the impulse response, white noise could be assumed to represent the effects of surface roughness, bumpiness, or impedance variations along the real wall’s surface on the pressure waveform of the diffuse reflection. We thus assume that the non-specular energy is temporally diffused according to the derived decay with random phase perturbations, while the long-term magnitude spectrum is not affected. We propose to efficiently approximate these main properties of the decaying diffuse reflection response using all-pass filter structures [50, 70, 71], originally suggested for artificial room reverberation. To achieve a pulse density that is sufficiently high for the human listener, Schroeder [50] connected these all-pass reverberators in series to form an all-pass cascade (APC). Based on [50], the here suggested decay filter (with “colorless” long-term spectrum) uses a series of four all-pass filters of the form

$$H_i(z) = \frac{g_i + z^{-\tau_i}}{1 + g_i z^{-\tau_i}}, \quad i = 1 \dots 4, \quad (12)$$

with

$$\tau_i = \frac{\tau_s f_s}{\eta^{i-1}}, \quad \tau_s = \frac{T_s \log |g^{-1}|}{3}, \quad (13)$$

where all gains $g_i = g = (\sqrt{2})^{-1}$ are fixed and identical, f_s is the sampling frequency, z is the discrete variable of the z -transform, and the delay ratio of the different all-pass filters depends on the parameter $\eta = \pi$, to achieve a (nearly) non-commensurate ratio [71]. The delays τ_i are rounded to the closest integer sample value. The local decay time T_s can be derived based on the distances to the reflecting wall as outlined above. Alternatively, for a given room, the mean distance to all walls, or the mean free path length may be used, neglecting the (time-variant) source and receiver positions, and resulting in a time-invariant APC. The number of four all-pass filters was chosen to keep computational costs low, while achieving a sufficiently high pulse density for typical room dimensions (see [50]). Figure 3a shows the resulting filter output (black) for $R = 10$ m ($T_s = 0.2$ s) in addition to the corresponding exponential decay (red).

In addition to the temporal spread at the receiver which can be implemented by the above filter approximation, the spatial spread caused by the scattered diffuse reflections with contributions impinging from the solid angle covered by the reflecting surface, might be considered for auralization. In contrast to the specular reflection, which impinges from a single direction, the diffuse reflection perceptually results in a widened apparent source width. Depending on the sound rendering system, this can be implemented in different ways. In the context of the diffuse reverberation model and rendering in [41, 43], the spatial spread can be represented in a simplified manner by mapping the output of the all-pass filter to those spatially distributed virtual reverberation sources of the FDN spatialization which represent the reflecting surface (for details see Sect. 3). For simplicity, the dot product between the normal vector of the considered wall surface and the direction of the virtual reverberation sources can be used, equivalent to the angular dependence in equation (2).

2.1.3 Decomposition filters

Specularly reflected and scattered sound are implemented for each image source in an ISM using parametric decomposition filters based on the frequency-dependent scattering coefficient [31]. The temporal spread of the diffuse reflections is introduced by applying the above described APC, before feeding the diffuse part into a diffuse reverberation model. Here, this scheme was implemented in [41, 43] using a spatially-mapped FDN as diffuse reverberation model (for more details see Sect. 3). The mapping of the FDN output to spatially evenly-distributed virtual reverberation sources [41, 66] around the listener is utilized and diffuse reflections are assigned to virtual reverberation sources representing the direction of the reflecting surface. It should be noted that the local decay time only depends on the geometrical arrangement and the spatial distribution of scattering (here assumed to be ideal diffuse), and does

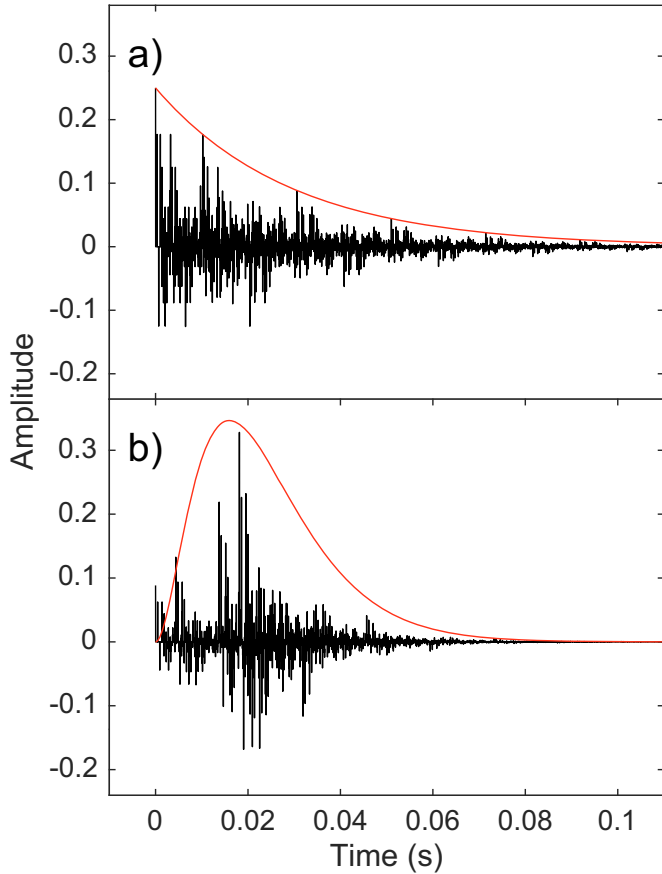


Figure 3. a) Response of the suggested cascaded all-pass filter (black) for surface scattering. The distance was $R = 10$ m ($T_s = 0.2$ s). The red trace shows the respective exponential decay. b) Modified all-pass cascade filter for object scattering (black) with a group delay of 20 ms. The red trace shows the roughly approximated Gamma distribution function with shape parameter $k = 3$ and scale parameter $\theta = 2\gamma/5$.

not depend on the scattering coefficient. Accordingly, the frequency-dependent scattering coefficient determines the attenuation of the scattered part.

Two power-complementary IS decomposition filters (second-order, low-shelving filter for the specular part, see Fig. 4) were implemented assuming that the scattering coefficient δ is generally low for low frequencies and high for high frequencies (see, e.g., [65]). The transition between the low-frequency specular part H^σ and the high-frequency diffuse part H^δ , $|H^\sigma|^2 + |H^\delta|^2 = 1$, is specified by the cross-over frequency f_c . With this simplified filter design only two parameters, δ and f_c , are used to approximate measured scattering coefficients for each wall.

2.2 Object scattering

For scattering from a flat, yet rough surface, the parameters of an APC were chosen to account for a monotonic temporal decay, representing diffusely reflected sound from increasing distances at the surface. To effectively simulate sound scattering caused by interior objects such as furniture

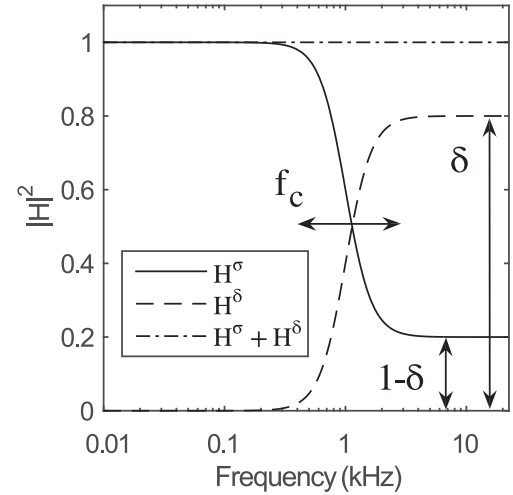


Figure 4. Parameterized image source decomposition using two power-complementary filters H^σ and H^δ for the specular and diffuse reflection, respectively.

or larger irregularities at room boundaries, a further APC-based approach is suggested. Here the assumption is that the simplified effect of such geometric deviations from an (ideal) empty shoebox is that no single specular reflection or diffuse reflection from each wall with only one associated single delay (and surface scattering tail) is observed. Instead, multiple scattered (inter-) reflections between surfaces of interior objects are assumed to occur along the path of each reflection at the walls of the surrounding shoebox room, arriving at the receiver with different delays. For the temporal distribution of the summed reflections from multiple objects at the receiver, an envelope shape similar to a gamma distribution function can be assumed [36].

A single-valued geometric deviation parameter is introduced, referred to as ζ , expressing the deviation from an (ideal) empty shoebox room. For an empty shoebox, the geometric deviation is defined as $\zeta = 0$. For other rooms with interior objects or geometric deviations from the shoebox shape, larger values, $0 < \zeta < 1$, are heuristically chosen. A typical office room with some shelves at the walls and desks would be characterized by an averaged geometric deviation of about 0.05–0.1, depending on the size of the shelves relative to the room dimensions. A storage room filled with shelves and goods would have a geometric deviation of 0.4–0.6. It should be noted that, so far, there is no exact expression to determine ζ for a given room with interior objects, instead it reflects a rather rough approximation.

To efficiently approximate the desired temporal distribution of summed reflections, again a series of four all-pass filters as in equation (12) is used. However, here the gain factors g_i and delays τ_i were heuristically altered to approximate the desired gamma-distribution-shaped envelope:

$$\mathbf{g} = \left(\sqrt{2}^{-1}, \sqrt{2}^{-1}, \sqrt{2}^{-2}, \sqrt{2}^{-3} \right), \quad (14)$$

$$\tau = \left(\frac{\tau_0 f_s}{\eta^3}, \frac{\tau_0 f_s}{\eta^2}, \frac{\tau_0 f_s}{\eta}, \tau_0 f_s \right), \tau_0 = \frac{\gamma}{\sum_{k=0}^3 \eta^{-k}}. \quad (15)$$

The delays τ_i are rounded to integer sample values. γ is the desired group delay of the APC and is calculated from the (specular) reflected overall path length d (based on the surrounding shoebox room boundaries), the speed of sound c , and the geometric deviation parameter ζ :

$$\gamma = \zeta \frac{d}{c}. \quad (16)$$

For $\zeta > 0$, the reflections in the ISM are filtered with a specific APC with parameters as defined in equations (14)–(16). For each image source, a single APC is applied, independent of the reflection order. The amount of temporal diffusion thus only depends on the overall path length d resulting from all reflections. Figure 3b shows the filter output (black) for a group delay of $\gamma = 20$ ms. The red trace shows the Gamma distribution with a shape parameter $k = 3$ and a scale parameter $\theta = 2\gamma/5$, which is approximately matched with the suggested APC. Without such a filter, particularly for large rooms, where individual early reflections from the room boundaries are temporally sparse, pure specular/diffuse reflections in the ISM (including a surface scattering tail as described in Sect. 2.1) might result in an unrealistic “crackling” impression for transient sounds. While such an impression might occur for large, completely empty shoebox rooms, real rooms almost always deviate from this geometry. In this case, the simplified effect of object scattering is assumed to be audible, even for a small geometric deviation parameter of $\zeta = 0.05$ (5%), as could be assumed for a nearly empty room with few shelves at walls, support beams, wall projections, or pillars.

3 Perceptual evaluation

The perceptual evaluation was performed in two steps: In Experiment 1, the focus was on the effect of object scattering and the parameter ζ , given that the suggested approximation is highly abstracted. Based on the results of Experiment 1, the parameter space was adapted in Experiment 2. Both the suggested (physically-based) implementation of surface scattering, and object scattering were evaluated separately as well as in combination. All tests were performed by comparing measured and simulated binaural room impulse responses (BRIRs) for several example rooms. For the simulations, the scattering approximations were implemented in the room acoustics simulator, RAZR, see [41, 43].

3.1 Methods

3.1.1 Listeners

In Experiment 1, eight normal-hearing listeners (all male) aged 29–46 years participated. A second group of fifteen normal-hearing listeners (4 female, 11 male) aged 18–57 years participated in Experiment 2. None of the

participants reported any hearing problems. All listeners were working in the field of (virtual) acoustics or hearing research and/or were experienced in performing psychoacoustic experiments. They were therefore considered expert listeners.

3.1.2 Rooms and room simulation

Measured BRIRs for four rooms taken from existing databases in the literature were used as reference. For additional details on these rooms, the reader is referred to the respective publications and databases.

1. Room A (Experiments 1 and 2): a large aula (Aula Carolina, Aachen, Germany; Aachen impulse response (AIR) database, [72]), roughly shoebox-shaped, but with some pillars inside and arches at the ceiling. The walls are mostly rigid (bricks, windows), except for some draperies. Compared to the room dimensions (approx. 19×30 m², 10 m height, 5700 m³ volume), the distance between source and receiver was quite small (at approx. 3 m), which results in a relatively high direct to reverberant energy ratio in the sound. The broadband reverberation time (T30) is about 4.7 s.
2. Room C (Experiment 1): a medium-sized corridor [41] with uniform and even finish coat (plaster) at all side walls and tiles on the floor. 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. The outer dimensions (used for the shoebox approximation) are: $2.65 \times 7.60 \times 2.49$ m³ (= 50.15 m³). The measured broadband reverberation time (T30) is about 0.9 s.
3. Room S (Experiment 1; AIR database, [72]): a shoebox-shaped seminar room (10.8×10.9 m², 3.15 m height, 370.8 m³ volume) with concrete walls, windows at three sides and parquet floor. Interior objects are tables and chairs. The broadband reverberation time (T30) is about 0.8 s.
4. Room U (Experiment 2; [73, 74]): the underground (U-Bahn) station Theresienstraße in Munich, Germany. It is a large, elongated space containing the platform and two tracks of about $120 \times 15.7 \times 4.16$ m³, with a total volume of about 11000 m³ and a measured broadband reverberation time (T30) of about 1.6 s. The platform floor and the tunnel walls are covered with tiles. The receiver was standing near the middle of the platform and the source was located at 6.37 m distance in front of the receiver.

The two suggested scattering methods were implemented in the framework of the room acoustics simulator RAZR (see [41, 43] for a detailed description) which uses an ISM for early specular reflections and a spatially rendered FDN as diffuse reverberation model. Figure 5 shows a schematic diagram of the signal processing with the added components for object (bold red) and surface scattering (bold blue). The upper left part of the schematic depicts the ISM (light red background), the right-hand part depicts

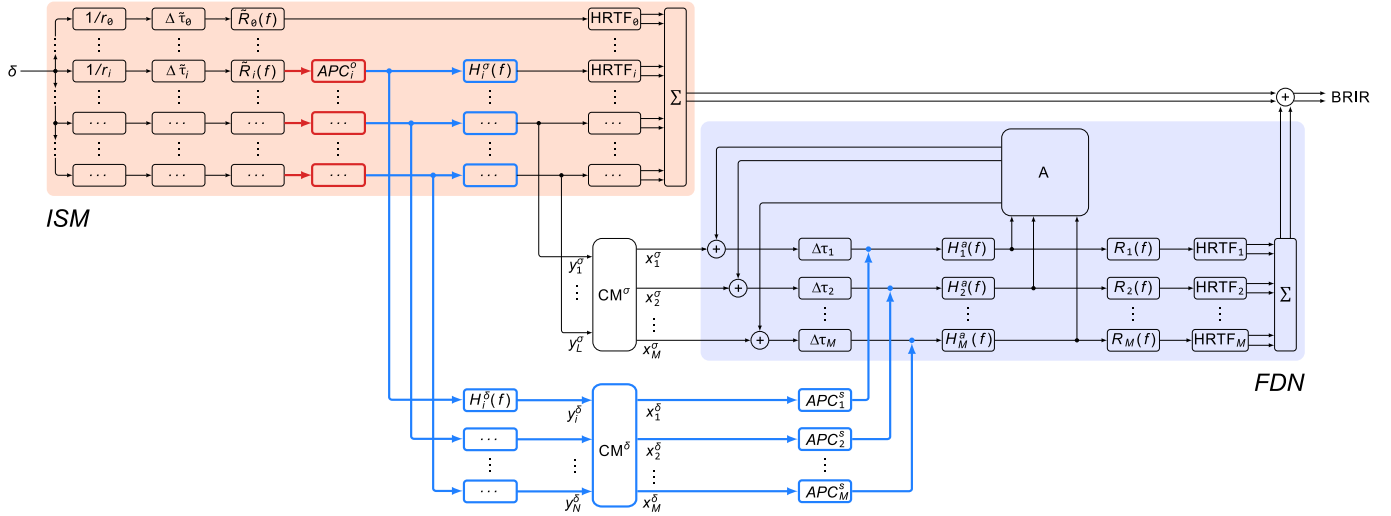


Figure 5. Block diagram of the implementation of the object and surface scattering module (thick red and blue components, respectively) in the framework of the room acoustics simulator RAZR (thin lines) consisting of the ISM (light red block) and FDN (light blue block) module. For object scattering, the IS signals are subjected to cascaded all-pass filters APC^O (red). Consecutively, the outputs are separated into specular and non-specular reflected parts by the decomposition filters $H^\sigma(f)$ and $H^\delta(f)$ (blue). Additionally, the non-specular part is temporally diffused by Schroeder-reverberators with cascaded all-pass filters APC^S (blue).

the FDN (light blue background). The transition between the ISM and FDN is achieved by coupling the diffuse part of each reflection (bold blue) into the FDN and by coupling both the diffuse and specular part of the last ISM order into the FDN. Given that individual reflections of the ISM are temporally distributed, early FDN outputs and late ISM inputs temporally overlap, ensuring a smooth transition. The channel mapping matrices (CM) either map each specular ISM output to those FDN channels representing the spatially opposite hemisphere (including the FDN delays, thus mimicking sound having travelled once through the room and being reflected on the opposite side), or map each diffuse ISM output to those FDN channels on the same hemisphere (initially skipping the FDN delays, thus representing spatially distributed diffuse reflections directly following the respective specular reflection). The ISM used a shoebox proxy room, disregarding the exact geometry of the real rooms, however, using the same volume. The effect of interior objects, such as furniture, on the reverberation time was indirectly represented by the wall absorption coefficients in the ISM, which were estimated from the measured frequency-dependent reverberation times using the inverse form of Eyring’s formula (e.g., [65]). The maximum image source order was set to a default value of 3, providing a good trade-off between accuracy and computational cost [43]. The late reverberation tail was generated by a 12-channel FDN. The FDN parameters were derived to account for the (measured) frequency-dependent reverberation time.

In Experiment 1, the temporal spread caused by object scattering was implemented for each reflection path in the ISM using the geometric deviation parameter ζ . Figure 4 shows the implementation of the object scattering module in RAZR as bold red processing blocks, termed APC^O . For rooms A, C, and S, BRIRs for relative geometric

deviations of 0 % (i.e., no object scattering), 5 %, and 20 % were created. Head-related impulse responses of the MK2 artificial head by Cortex were used in the simulation (as used for the recordings in room C).

In Experiment 2, surface scattering was additionally implemented with decomposition filters, and temporal smearing was applied for all reflection orders in the ISM (see bold blue processing blocks in Fig. 5). For rooms A and U, BRIRs without any object and surface scattering, with isolated object scattering, ζ of 5 %, and 20 %, with isolated surface scattering ($f_c = 1000$ Hz, $\delta = 0.25$ and 0.5), and with combined object (ζ of 5 %) and surface scattering ($f_c = 1000$ Hz, $\delta = 0.25$) were created. Head-related impulse responses of the Head acoustics HMS were used in the simulation (as used for the recordings in room A and U). To reduce coloration differences between the recordings and simulations, an octave-smoothed equalization filter was applied in the simulation.

3.1.3 Apparatus and procedure

The participants were seated in a sound-attenuating listening booth and listened using Sennheiser HD 650 headphones (equalized transfer function), driven by an RME ADI-2 (Exp. 1) and RME UCX (Exp. 2) soundcard at a sampling rate of 44.1 kHz. The participant’s task was to rate perceived differences between measured (reference) and simulated rooms with respect to certain perceptual attributes. For this, the measured and synthesized BRIRs were auralized by convolution with dry test stimuli.

In Experiment 1, perceptual ratings of a subset of seven attributes from the spatial audio quality inventory (SAQI; [68]) were obtained. In SAQI, each attribute consists of a name (“perceptual quality”), a circumscription, and scale end labels. The chosen attributes (including scale end

labels) were: *Tone color bright – dark* (darker – brighter), *Metallic tone color* (less pronounced – more pronounced), *Width* (less wide – wider), *Reverberation level* (less – more), *Envelopment* (by reverberation) (less pronounced – more pronounced), *Distortion* (less intense – more intense), *Naturalness* (lower – higher). In order to better account for the audible effects specifically associated with scattering, two attributes were specifically introduced in comparison to earlier studies [41, 43], *Distortion* and *Flutter echoes*, based on informal listening tests. The attribute *Distortion* was here applied although non-linear distortions cannot occur. However, the above described sparse early reflection can lead to a “crackling” sound impression, which is best described as distortion. The attribute *Flutter echoes* is not part of the SAQI, albeit is contained in the room acoustical quality inventory (RAQI, [75]). It was defined in this study as follows: “Perception of roughness and flutter in fast repeating reflections or echoes for transient sounds. For transients in music and speech, fluttering reflections particularly occur for high-frequency parts of the sound. Example: Speech or hand clapping between two parallel flat concrete walls.” The definitions of the original SAQI and RAQI attributes can be found in the respective manuals [68, 75]. While the attribute “width” referred to the sound source, all other attributes referred to the overall acoustic impression.

Attribute ratings were unipolar or bipolar. Ratings were performed by adjusting sliders on a graphical user interface. Values ranged from 0 to 100 for unipolar scales, and from –50 to 50 for bipolar scales. The acoustic scenes were presented in a randomized order. In contrast to the original SAQI-associated test procedure [68], three parallel sliders representing the three values of parameter ζ were used for each attribute. The order of the three sliders was randomly chosen for each rating. According to the original test procedure, for each room the overall perceived difference between measured and synthesized scenes (unipolar scale: “difference: none – very large”) was rated. If there was a perceived difference (i.e., rating > 0) for any of the three conditions, all other attributes were rated. During the introduction of the test, all perceptual attributes including their original definitions were presented to the participants and potential misunderstandings were clarified. In addition, the attribute definitions were displayed on the screen before each rating as a reminder.

Based on the results of the first experiment, the procedure in Experiment 2 was adapted and only the four perceptual attributes with the strongest effects, namely *Tone color*, *Naturalness*, *Distortion* and *Flutter echoes* were considered. Moreover, the reference signal itself was included for comparison (similar to the “hidden reference” in a MUSHRA procedure, [76]), adding one additional stimulus.

3.1.4 Stimuli

Two source stimuli were used in Experiment 1: a music excerpt and a pulse. Both exhibit clear transient parts. The music signal was a slapped bass guitar and is referred to as

music in the following. By the choice of this instrument and this playing technique, the rooms were excited in a wide frequency range and with transient parts. The here used pulse signal was a filtered version of a (digital) delta pulse to obtain a pink-spectrum-shaped frequency response. In Experiment 2, a castanets excerpt was additionally presented. This sound offers a 1-s fast sequence of rhythmic castanets claps with many transient parts. In contrast to the single transient event in the pulse signal, the sequence of transients overlaps with the pattern of early reflections. All (reverberated) stimuli were about 4 s (music) and 2 s (pulse, castanets) in duration. The mean sound pressure level for the presentation was 65 dB SPL (music), 54 dB SPL (initial 0.2 s of pulse), 56 dB SPL (initial 1 s of castanets).

3.2 Results of experiment 1

The rated perceived differences for object scattering between the measured (reference) and simulated conditions, averaged over all listeners are shown in Figure 6a for music and in Figure 6b for pulse. Ratings for different object scattering (geometric deviation) parameters (0%, 5%, 20%) are indicated by the differently colored symbols (see Fig. 6 legend). The error bars indicate inter-individual standard errors. For both test signals, the *Overall difference* to the measured scenes (top left sub panels) is in the medium range, and standard errors are largest compared to all other attributes. For music (a), no notable differences between the three levels of object scattering were perceived: A two-way repeated-measured analysis of variance (rmANOVA) [room (3 levels) \times object scattering (3 levels)] showed no significant main effect for room [$F(2, 14) = 2.17, p = 0.15$], object scattering [$F(2, 14) = 2.40, p = 0.12$], nor any significant interaction [$F(4, 28) = 1.16, p = 0.35$]. For the pulse (b), the overall differences to the measured reference decrease with increasing scattering parameter and are generally larger than for music. Here, a two-way rmANOVA showed a highly significant main effect of object scattering [$F(2, 14) = 29.45, p < 0.001$], while no significant main effect of room [$F(2, 14) = 1.0, p = 0.39$], and no significant interaction [$F(4, 14) = 1.48, p = 0.23$] were observed. Accordingly, the pulse is the more critical test signal compared to music, and the suggested scattering approximation leads to a higher perceived similarity between the recording and the simulation with increasing scattering parameter. Based on the statistical significant effect of object scattering for pulse in the overall difference, we further assessed the individual sound attributes for pulse (other panels in Fig. 6b) by analyzing the main effect of object scattering (two-way rmANOVA). For most attributes, perceived differences were generally small and no statistically significant effect of scattering was found. A statistically significant main effect of the scattering simulation was observed for the attributes *Distortion* [$F(2, 14) = 25.53, p < 0.001$], *Naturalness* [$F(2, 14) = 24.67, p < 0.001$], and *Flutter echoes* [$F(2, 14) = 4.17, p = 0.04$]. Differences to the measured conditions decreased with increasing object scattering parameter.

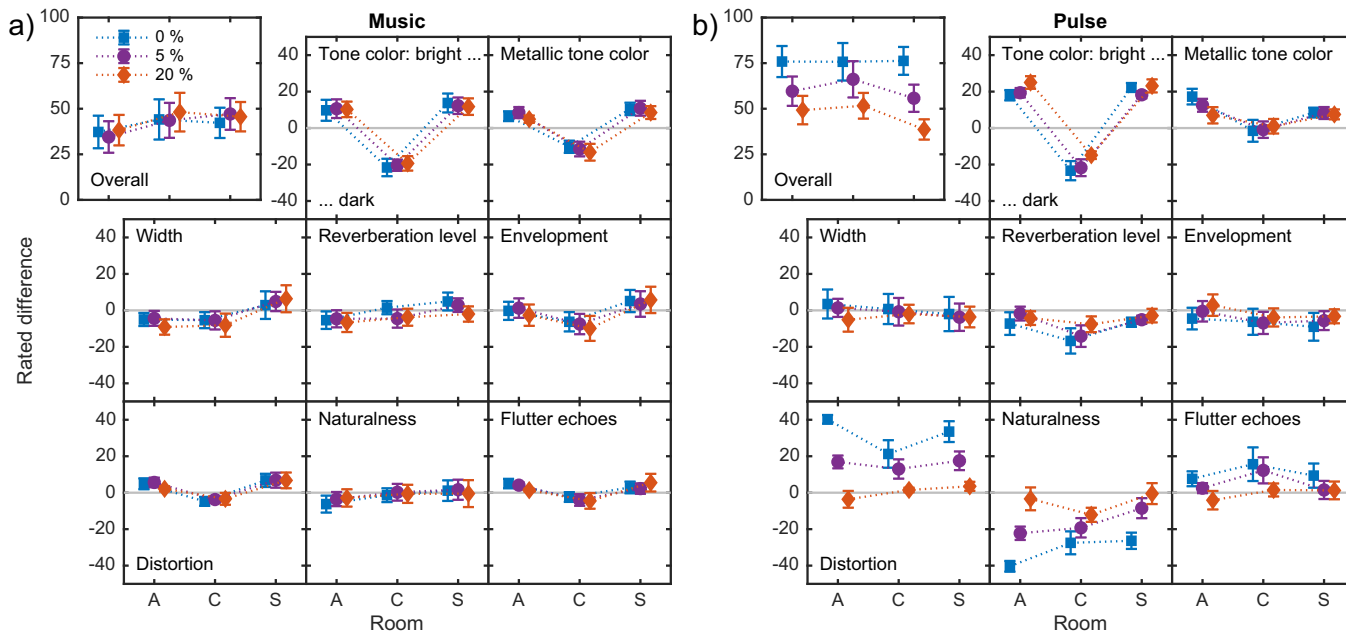


Figure 6. Average rated differences for object scattering between simulated and measured (reference) conditions for music (a) and pulse (b). The different colors and symbols indicate the object scattering parameter (see legend). The different sound attributes are indicated in the panels, the three different rooms are indicated on the abscissa (A: aula, C: corridor, S: seminar room). Error bars indicate inter-individual standard errors. Depending on the attribute, ordinate scales ranged from, e.g., “less pronounced” to “more pronounced” or semantically fitting descriptors.

Regarding the effect of room, for *Tone color*, ratings were very similar for music and pulse. Deviations in both directions (brighter, darker) indicate overall spectral differences between the recording and the simulation. A similar, however, less pronounced pattern was observed for *Metallic tone color*.

In summary, consistent perceived differences between measurement and simulation exist for *Tone color* and *Metallic tone color* for music and pulse. For the pulse, the different choices of geometric deviations for object scattering were clearly audible. It was found that a geometric deviation parameter of 20% led to the best match between simulated and measured room BRIRs as indicated by *Distortion*, *Naturalness*, and *Flutter echoes*. Based on the current results, the further evaluation in Experiment 2 focused on the four attributes which showed large effects: *Tone color*, *Distortion*, *Naturalness*, and *Flutter echoes*. To further assess the effect of scattering in natural sounds, castanets were added as additional stimulus.

3.3 Results of experiment 2

The results for the rated perceived differences between the simulated and measured (reference) conditions, averaged over all listeners are shown in Figure 7 for four quality attributes as indicated in the panels. The pulse (blue squares) was presented in the underground and aula (open and closed symbols, respectively). Castanets (red circles) and music (yellow triangles) were presented in the aula. The different scattering parameters are indicated on the

abscissa: No scattering was applied for the left-most data point, followed by object scattering parameters 5% and 20%. In the next two data points, surface scattering was applied with a scattering coefficient of 0.25 and 0.5. On the right-hand side, the label “Both” indicates combined object and surface scattering (5% and 0.25). The right-most data point shows the (hidden) reference, labeled “Ref”.

A distinct pattern is observed for *Distortion* and *Naturalness* for the pulse (underground and aula) as well as for castanets with clear deviations from the reference for the condition without any scattering (None) and for conditions without surface scattering. Similarly, the largest deviations for *Flutter echoes* are observed for pulse and castanets in the aula. For *Distortion*, a two-way rmANOVA [factors room (2 levels) \times scattering condition (7 levels)] for the pulse showed a significant main effect of scattering [$F(6, 84) = 30.57, p < 0.001$], no effect of room [$F(1, 14) = 4.39, p = 0.055$], and a significant interaction [$F(6, 84) = 11.52, p < 0.001$]. Post-hoc pairwise comparisons (using Bonferonni correction) revealed significantly different marginal means (averaged across room) to the reference for None, Obj 5%, Obj 20% (all $p < 0.001$). For the aula, the same significant differences to the reference were found for None, Obj 5%, Obj 20% (all $p < 0.001$), while for the underground a significant difference was found only for None ($p < 0.001$), in line with the significant interaction term. Thus, with surface scattering in the aula, and with object and surface scattering in the underground, as well as with their combination (Both), *Distortion* for pulse was rated the same as for the reference. For castanets,

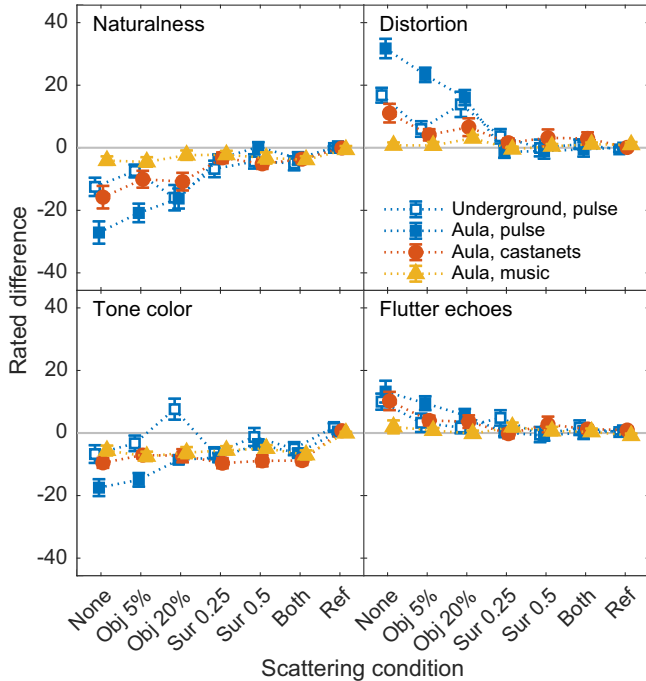


Figure 7. Average rated differences between measured and simulated auralizations for four different sound attributes (see panels), plotted against the scattering conditions. No scattering is on the left, the hidden reference rating is on the right. Different symbols and colors (see legend) indicate the combination of rooms (open: underground, closed: aula) and stimuli (blue: pulse, red: castanets, yellow: music). Error bars indicate inter-individual standard errors.

a significant main effect of scattering was also found for *Distortion* [one-way rmANOVA; $F(6, 84) = 6.13, p < 0.001$], and post-hoc tests revealed a significant difference for None ($p < 0.05$) to the reference.

For *Naturalness* a similar, however, inverted pattern of results as for *Distortion* was observed. Here, the two-way rmANOVA for pulse also showed a significant main effect of scattering, [$F(6, 84) = 18.32, p < 0.001$], no effect of room [$F(1, 14) = 1.94, p = 0.18$], and a significant interaction [$F(6, 84) = 8.09, p < 0.001$]. Post-hoc pairwise comparisons (with Bonferonni correction) revealed significantly different marginal means (averaged across rooms) to the reference for None and Obj 5% (both $p < 0.001$), Obj 20% ($p < 0.01$), and Sur 0.25 ($p = 0.039$). In the aula, significant differences to the reference were found for None and Obj 5% (both $p < 0.001$), Obj 20% ($p < 0.01$), and for None and Obj 20% (both $p < 0.05$) in the underground. Likewise, castanets showed a significant main effect of scattering for *Naturalness* [one-way rmANOVA; $F(6, 84) = 10.25, p < 0.001$], and a significant post-hoc test for None ($p < 0.01$) compared to the reference.

For *Flutter echoes*, overall smaller effects were observed in Figure 7d, with a similar, yet much less pronounced pattern of results as for *Distortion*. Here, the two-way rmANOVA for pulse showed a significant main effect of scattering, [$F(6, 84) = 8.74, p < 0.001$], no effect of room

[$F(1, 14) = 0.25, p = 0.63$], and a significant interaction [$F(6, 84) = 2.44, p < 0.05$]. Post-hoc pairwise comparisons (with Bonferonni correction) only revealed significantly different marginal means (averaged across rooms) to the reference for None ($p < 0.05$). For castanets, a significant main effect of scattering was found for *Flutter echoes* [one-way rmANOVA; $F(6, 84) = 4.06, p < 0.01$], however, no significant post-hoc pairwise comparisons were found.

For music, no consistent changes were observed for *Distortion*, *Naturalness*, and *Flutter echoes* in Figure 7, supported by no significant main effects of scattering condition. For *Tone color*, all stimuli and all conditions resulted in a “darker” rating compared to the reference in Figure 7, except for Obj 20% in the underground. Here a significant main effect of scattering condition was found for all stimuli.

Taken together, Experiment 2 showed clear perceptual differences to the reference, particularly in the absence of any scattering simulation (“None”) for the pulse stimulus, in line with Experiment 1. A similar, however, less pronounced pattern of results was observed for the castanets. For the music excerpt, no effect of scattering was observed, also in line with Experiment 1. The inclusion of the (hidden) reference allowed for investigating perceptual differences to the ratings for the reference in the post-hoc pairwise comparisons. Increasing object scattering generally resulted in decreasing, yet still persisting perceptual differences. In contrast, surface scattering or a combination of object and surface scattering (“Both”, with 5 % object scattering and a surface scattering coefficient of 0.25) resulted in no significant differences to the reference. Overall the results suggest “Both” as the best general solution for transient stimuli.

4 Technical evaluation

To objectively assess the effect of the suggested scattering approximations, the reverberation echo density was compared to that of the measured (reference) BRIRs in the aula and underground. In addition, the current room acoustics simulations were compared to a purely specular image source model of the (empty) proxy shoebox rooms and to alternative GA simulations using ray tracing. The reverberation echo density was estimated as suggested in [35], by comparing the number of sample values that fall outside the interval defined by the IR’s standard deviation within a sliding temporal window. The proportion of “outlier” samples is normalized to the expected value for a Gaussian distribution. Consequently, the (normalized) echo density measure approximates 1 when the echo density becomes sufficiently high after a certain (room-dependent) time, and the statistics of the IR become indistinguishable from Gaussian. A window duration of 25 ms and a raised-cosine window were used here for both rooms (for further details see [35]). Additionally, the resulting echo density estimate was further smoothed by convolution with the same raised-cosine window.

The solid black traces in Figures 8a and 8b show the estimated normalized echo density for the measured BRIRs (averaged across both ears) in the aula and underground,

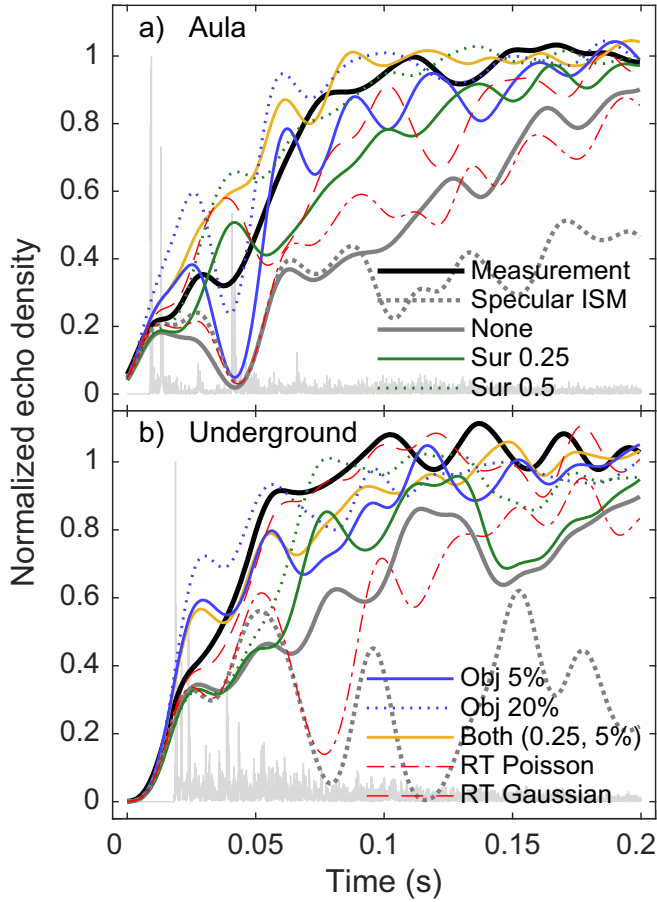


Figure 8. Comparison of the normalized echo density measure for different room acoustics simulations (see legend) and the recorded BRIR (solid black trace) in the aula and underground in panels a) and b), respectively. Echo density estimates were averaged across both ears. The faint grey trace in the background depicts the absolute value of the recorded impulse response.

respectively. The underlying (rectified) BRIRs are indicated in faint grey in the background. It is evident that the echo density increases during the initial part of the IR and reaches 1 after about 100 ms in the Aula and after about 80–90 ms in the underground, in conjunction with an initially steeper increase.

To estimate the overall effect of surface scattering and object scattering (including reflections on interior structures), the echo density of the measurement can be compared to that of a (purely specular) image source model of the equivalent empty room (dotted grey trace; Specular ISM). For both rooms, the echo density of the ISM is considerably lower than that of the measurement. While a gradual increase in echo density can be observed in the (smaller) Aula, the specular reflections in the much larger underground are still so sparse, that the echo density measure fluctuates without showing a clear increase within the depicted initial 200 ms of the impulse response. The solid grey trace (None) represents the above described underlying hybrid model without scattering extension (see [41, 43]). It is obvious that the use of the FDN-based diffuse

reverberation model after the 3rd ISM order leads to a faster increase of echo density compared to the ISM (dotted grey). However, without considering scattering for the initial three ISM reflection orders, echo density still increases considerably slower than for the measurements in both rooms. The solid and dotted green traces show the effect of including surface scattering alone (Sur 0.25, Sur 0.5), respectively. For the Aula, Sur 0.25 shows a considerably increased echo density compared to None, however, underestimating the measured echo density in the temporal region around 100 ms. Sur 0.5 (dotted green) fits the measurement well. In the underground, Sur 0.25 also shows an overall too small echo density with larger deviations from the measurement compared to the aula. Sur 0.5 shows a good agreement with the measurement after about 80 ms. The solid and dotted blue traces are for object scattering alone (Obj 5%, Obj 20%). In the aula, Obj 5% and Obj 20%, respectively, slightly under- and overestimate the measured echo density. In the underground, both are comparable after about 100 ms, and otherwise show an initially too steep increase of echo density. The solid yellow trace depicts combined surface and object scattering (Both; 0.25, 5%) as also perceptually evaluated in the previous section. In the aula, the measured echo density is well explained, except for some overestimation between 20 and 80 ms. In the underground the fit is also good, with some underestimation between 50 and 150 ms.

The red dash-dotted and dashed traces are for a combined ISM and ray tracing (RT) implementation as described in [14, 55], representing an alternative GA solution. Diffuse reflections were simulated in nine octave frequency bands centered at 62.5 Hz to 16 kHz, using stochastic vector-based (see, e.g., [54]) reflection directions based on the ideal diffuse (Lambert) spatial distribution and the specular reflection direction added according to the scattering coefficient per band. Here, the same spectral profile for the scattering coefficient as for the here suggested surface scattering approach was used (see Fig. 4), with $\delta = 0.25$, along with the same absorption coefficients and room dimensions. For each ray intersection with a wall (reflection point), the contribution of scattered sound at the receiver was estimated assuming ideal diffuse scattering. The energy histogram was collected in 1-ms time bins using 128,000 rays in each band (highly sufficient for the here analyzed first 200 ms of the IR). For a single large perfectly reflecting and ideal diffuse scattering wall ($\rho = 1$, $\delta = 1$), the ray tracing implementation was verified to numerically show the same energy decay as derived in Section 2.1.1. The final impulse response was obtained by combining specular reflections from the ISM, weighted with $(1 - \delta)$, with the ray-traced diffuse sound.

Given that the diffuse part of the impulse response has to be constructed from the ray-traced time-energy representation, the echo density depends on the underlying assumptions made for this conversion step: For the red dash-dotted trace (RT Poisson), an inhomogeneous Poisson process was used to generate a stochastic echo sequence which was modulated with the time-energy response in the nine octave bands (see [14, 55]). Hereby, the rate of the Poisson process

is based on the time evolution of the echo density estimated for image sources in a rectangular room ([77]; as described in [14]) with the maximum rate limited to 20 kHz. As such, the initial echo density of the so simulated diffuse reflections is too low compared to the measurement, and the echo density is overall broadly comparable to that obtained for the current approach without simulating scattering (None, solid grey line). Conversely, the red dashed trace (RT Gaussian) uses a Gaussian noise instead of the Poisson noise (comparable results would be obtained by using a constant rate of 20 kHz for the Poisson noise). Here the estimated echo density of the RT simulation fits the measurements better than RT Poisson. For the aula, the results are similar to the suggested surface scattering solution with the same scattering coefficient (Sur 0.25, solid green), showing an overall too low echo density between 50 and 150 ms of the IR. For the underground, RT Gaussian fits the measurement very well, better than the suggested surface scattering solution (Sur 0.25, solid green) with the same parameters.

5 Discussion

5.1 Relation to other approaches for surface scattering

The current derivation of surface scattering has some differences to an earlier solution [36] for modeling incoherent reflections from rough room surfaces. The solution in [36] is based on the far-field approximation for the source and wave scattering by a surface covered with circular bosses, based on Biot's rough surface modeling theory [78–80]. Depending on the size of the bosses, their scattering model is only valid for low frequencies, in their provided example below 500 Hz. The current solution is based on sound intensity and uses ideal diffuse Lambert scattering (as can also be used as kernel in, e.g., the BRDF in the acoustic rendering equation [44]). While in [36] the non-specular reflection results in an exponentially decaying tail, the current solution follows a power-law that was approximated by an exponential decay. One important difference is that for the far-field source, low-frequency approximation of [36], the decay time strongly depends on the incident angle and their solution results in no temporal spread for perpendicular incidence angles of the sound (see Fig. 5 in [36]). In contrast, the current solution was derived for exactly that perpendicular case and shows a clear temporal spread. This result appears reasonable for scattering from real surfaces, at least at high frequencies. Further numerical simulations (not shown) indicate that for the current ideal diffuse scattering model, a temporal decay is generally observed for all incidence angles. Given that the scattering coefficient of real surfaces is typically large at high-frequencies, the applicability of the current solution for high frequencies offers advantages for application in GA-based room acoustics simulation and virtual acoustics.

The comparison of echo density for the current solution and for ray tracing with Gaussian noise “carrier” showed a comparable behaviour in the aula and some differences in the underground when using the exact same parameters

(surface scattering only, $\delta = 0.25$; Sur 0.25; solid green and dashed red traces in Fig. 8). In the aula, with a moderate ratio of the room dimensions ($19 \times 30 \times 10 \text{ m}^3$), the use of the identical temporal decay for all reflections in the current solution, agrees well with surface scattering modelled by ray tracing (dashed red). Both models, however, show a systematic underestimation of echo density in comparison to the measurement. This can be attributed to geometric deviations (e.g., church aisles and columns) of the real aula to the here used empty proxy shoebox room. In contrast, in the underground, the ray tracing solution approximates the measurement very well using the empty proxy shoebox, while the current solution (Sur 0.25; solid green) considerably underestimates echo density. It appears that the use of identical temporal decays for all reflections in the current solution is less suited for elongated rooms with largely different room dimensions ($120 \times 15.7 \times 4.16 \text{ m}^3$), although the real geometry is well approximated by an empty shoebox in this case, as demonstrated with the ray tracing model. Combining moderate amounts of surface and object scattering in the current solution (Both; solid yellow traces in Fig. 8) agreed reasonably well with the measured echo density in both rooms. In the aula, object scattering can be assumed to meaningfully represent the acoustic effect of the geometric deviations from the empty shoebox.

It is evident that the temporal evolution of echo density (and the agreement with the measurements) for ray tracing critically depends on the assumptions for converting ray-traced time-energy histograms to impulse responses. To achieve an estimated echo density around 1 as observed in the measurements after about 100 ms, a Gaussian noise or Poisson noise “carrier” with a rate of 20 kHz was required. Originally [57] suggested a rate of at least 5 kHz, while [14, 55] use a maximum rate of 10 kHz.

Taken together, the current approach shows a clear improvement of GA room simulation regarding echo density in comparison to a purely specular ISM and a hybrid ISM/FDN model without scattering. In comparison to ray tracing, the current surface scattering with fixed temporal spread for all reflections shows limitations in rooms with large dimensional ratios. However, the current solution is much simpler and can be directly implemented in the time domain, without the necessity to translate time-energy histograms to impulse responses.

Even though the current results suggest a limited perceptual relevance of including temporal diffusion as a consequence of surface scattering in room acoustics simulation for natural stimuli such as music, the perceptual results for pulse and castanets show a clear improvement of the simulation, in line with the improved temporal evolution of echo density in comparison to the real rooms.

5.2 Differences between surface and object scattering

The two suggested approximations for surface and object scattering bear some similarities in their digital filter implementation and underlying concepts, however, also show some differences. The approximation of surface

scattering was directly derived from the physics of an ideal diffuse reflection from an infinite flat surface. The resulting power-law temporal decay for the idealized co-located source and receiver was then translated into an exponential decay, matching the initial 10-dB decrease of the diffuse reflection. The advantage of the exponential decay is that it can be efficiently simulated using an all-pass cascade [50] which also models phase perturbations expected from diffuse reflections. The simulated exponential decay declines faster compared to the power-law for later stages of the decay process. Perceptually, it is expected that this initial decay is more important, however, no formal perceptual evaluation was performed for this specific design choice in the current study.

The current derivation of the time spread is only strictly valid for co-located source and receiver, and other combinations, particularly with the source or the receiver much closer to the surface would result in a different time spread. However, accounting for such differences would require the use of time-variant APCs for each room boundary, while for simplicity a fixed, solely room-dependent APC is more reasonable to be used for all boundaries, with a decay based on the mean free path length.

While the simplifications made for surface scattering may not be physically exact, the level of abstraction is considerably higher for the suggested object scattering approximation: Here, the main underlying physical concept is that multiple consecutive surface scattering processes with monotonic decay will result in a gamma-distribution-shaped temporal envelope of the scattered response [36]. The main difference to the monotonic decay for (single) surface scattering is the sloped attack of the envelope, comparable to sound propagation from room-to-room through openings or room-in-room reproduction [41, 81]. In the current object scattering approximation, only this basic physical principle is considered, while the digital filter implementation and the parameter selection are heuristic. In future work, machine learning might be applicable to estimate the geometric deviation parameter from room models or camera images.

Another difference between the two approaches lies in the treatment of the scattered energetic component. In the surface scattering, the scattered energetic part is fed into the diffuse reverberation model and thus contributes to the build-up of the simulated diffuse reverberation from the first reflection order onward. However, this was not considered for object scattering so far. One reason is the lack of an underlying description of the scattered energy, as is provided by the frequency-dependent scattering coefficient for individual surfaces. As a first approximation for object scattering, similar low- and high-shelving filters as those used for surface scattering could be employed. The choice of the crossover frequency would depend on the average size of the scattering objects. For surface scattering, the orientation of the reflecting surface is used to determine the direction from which the diffusively reflected sound originates and how it is integrated in the diffuse field model. Given that object scattering is based on a spatially unresolved geometrical deviation from the empty shoebox, the

scattered sound energy should be assumed to spatially evenly radiate in all directions.

While future work could aim at further harmonization and integration of both concepts, it should be noted that each concept needs to be independently implemented and requires independent adjustment to simulate different rooms. For example, in large empty halls, surface scattering can be assumed to dominate, while in small cluttered rooms object scattering dominates.

5.3 Connection to edge diffraction and diffusion

The current techniques can be combined with a more detailed simulation of edge diffraction (e.g., [26, 27]) to account for specific geometric details of the environment. Conceptually, surface scattering and edge diffraction represent more detailed simulations of macroscopically different phenomena, while the here suggested object scattering describes both processes in combination with multiple reflections for a sound propagation path with a certain coverage of geometrically unspecified objects. For computationally-efficient room simulation, it is thus recommended to explicitly simulate edge diffraction for large structures, such as room or building corners, to simulate surface scattering for larger surfaces, and to account for the remaining smaller scale geometric details through object scattering. While earlier approaches (e.g., [48, 49]) combined radiosity and path tracing to estimate impulse responses for convolution, the here suggested approach is suited for a straightforward time-domain implementation using digital filters. It can be seamlessly integrated with filter-based edge diffraction [26, 27], without processing in sub-bands (e.g., [28, 49]).

Both suggested methods effectively increase the echo density in the room simulation, for both early reflections and late reverberation. This can be conceptually compared to other approaches aimed at increasing the echo density, e.g., in FDNs [62, 82–84] by considering scattering and perceptual effects of echo density. However, in the current approach, scattering effects are implemented in an ISM, which is then followed by an FDN which is solely used as a late reverberation model. An increase in initial echo density is often referred to as “diffusion” in the context of artificial reverberation (for an overview, see [6]) as used in music effects processors. Hereby diffusion, as a design parameter, controls the extent to which echo density increases over time. The current contribution provides a physical basis for understanding this design parameter and its associated effects, at least for surface scattering. For object scattering, the geometrical deviation parameter ζ currently remains a rather heuristically adjusted parameter. In future research, objective methods to estimate ζ from the room geometry including interior objects should be investigated.

5.4 Perceptual impact

The current work was initially motivated by informal listening tests and the observation that simulating

particularly large volumes using a (shoebbox) ISM, without accounting for sound scattering, fails to produce perceptually satisfactory results, at least for transient sounds. Hereby, the main challenge for large volumes and shoebbox specular reflections are unrealistically sparse patterns of early reflections. Moreover, the assumption of purely specularly reflected high-frequency sounds appears unrealistic in most real environments. Technically, the sparseness or “granularity” of early reflections can be quantified by the echo density which clearly showed a much faster increase for the here tested real rooms than for the specular ISM. While some earlier studies noted the relevance of scattering for room acoustics simulation (e.g., [33]), others [61] found only small perceptual effects of temporal diffusion (mimicking a scattered reflection) for speech and no effect for music. They speculated that their speech source material offered more pronounced transients than their music sample, a piano phrase. In [85], the authors found detrimental effects of temporal diffusion by “envelope distorting” reflections, resulting in more muddy sound because of a partial break down of the precedence effect [86, 87].

In line with [61], the current results showed little to no perceptual effect for the current music excerpt, a slapped bass guitar. For the transient pulse and for castanets, however, the current listening tests demonstrated a clear perceptual impact. For the pulse (similar to the sound produced by a bursting balloon), the attribute *Distortion* best described the crackling effect of sparse early reflections, and showed a clear reduction of the deviation to the measured room with increasing geometrical deviation parameter ζ for object scattering (particularly in Exp. 1) and surface scattering (Exp. 2). Similarly, naturalness generally improved with increased object or surface scattering. Although the slapped-bass-guitar music excerpt, was deliberately designed to also include transient attacks, it can be suspected that these transients had not enough high-frequency content to evoke a perceptual effect. For the castanets (Exp. 2) no inclusion of scattering in the simulation still had a detrimental effect, however, less pronounced than for the pulse. Although increasing the amount of scattering in the here tested range generally resulted in better perceptual agreement with the measured rooms, there were specific cases in which increased scattering resulted in less agreement with the measured rooms: For *Naturalness* and *Distortion* in the underground in Experiment 2 (Fig. 7, open symbols), increased object scattering (Obj 20%) was rated considerably worse than less scattering (Obj 5%). In fact, 20% was rated quite different to the measurement, similar to the simulation without scattering (None). A similar trend was also observed for the castanets in the aula. Thus, it cannot be concluded that more scattering is generally better, particularly given that the here tested range of scattering parameters was already limited to a reasonable range: Considering the room geometry, including the presence of interior objects, higher amounts of object scattering, e.g., 40%, appeared an unreasonably high value for testing. Likewise, the surface properties in the real rooms did not justify scattering coefficients beyond $\delta = 0.5$. In connection with the estimated echo density, the increased perceptual

difference for 20% object scattering in comparison to the measurement in the real underground station might be related to an overly fast increase of echo density in the simulation compared to the measurement (see Fig. 8b, dotted blue vs. solid black).

A considerable part of the overall difference in Experiment 1, for both the music and pulse stimulus, was likely caused by *Tone color* differences. The tone color ratings show a distinct pattern with deviations in the direction bright for room A and S, and in direction dark for room C. These deviations were most likely caused by the different HRTFs used in room A and S from the AIR database, while room C was recorded with a different dummy head at the University of Oldenburg. It remains unclear whether the differences in tone color affected the rating of the effect of object scattering in Experiment 1. In Experiment 2, a different set of HRTFs was used (in agreement with the aula and underground recordings) and remaining coloration differences were compensated for based on octave-smoothed spectral differences of the BRIRs. Still significant coloration changes remained, with the simulation mainly perceived as darker than the reference. It appears unlikely that these remaining *Tone color* differences consistently affected the other ratings.

The current evaluation of scattering focused on temporal diffusion for several example rooms. Hereby, potential scattering effects for conditions with a strongly dominating nearby reflection and specifically effects of spatial diffusion for, e.g., lateral reflections were not covered. To also cover spatial effects, the attribute *Width* (of the source) and *Envelopment* were part of Experiment 1, however, showed no effect of object scattering, neither for music nor pulse. Although the current study found no perceptually relevant effects of including scattering in the room acoustics simulation for the tested music sample, it cannot be generally excluded that scattering has no audible effect except for those demonstrated here for pulse and castanets. While it can be expected that temporal diffusion caused by scattering will not be audible for other music stimuli or stimuli with less transient character, such as speech, there might be specific conditions in other rooms with stronger perceptual impact. Potential effects on speech for temporally separated echoes interfering with the syllable rate were not covered here.

Future research should specifically focus on speech and conditions with a dominating reflection, which might introduce spatial effects. Moreover, beyond the here selected attributes of the SAQI and RAQI, other attributes could be tested. We suspect, however, that for temporal diffusion effects, the choice of other potentially appropriate attributes, e.g., from RAQI *Irregularity in sound decay* or *Attack*, would have led to similar results. In both current experiments, *Naturalness* and *Distortion* showed broadly the same pattern (with inverted scale), indicating that listeners mapped the perceived difference to both attributes. It is unclear whether other descriptive attributes would have helped to better separate different perceptual effects.

Based on the current results, it is evident that the suggested scattering approximations are generally suited to improve room acoustics simulations, as supported by the

statistically significant effects for pulse and castanets, and by the technical evaluation of echo density. However, the perceptual effect of temporal diffusion caused by scattering was also demonstrated to strongly depend on the stimulus *per se*. Thus, for room acoustics simulation, particularly in the interest of computational efficiency in real-time applications, the relevance of including scattering effects has to be considered. Depending on the expected sound material, scattering could be disabled or could be enabled for certain scenarios only.

Finally, it should be noted that although the here suggested scattering approaches are physically grounded and the evaluation of echo density showed a clear improvement over simpler room simulation approaches disregarding scattering for early reflections, it cannot be excluded that the use of other diffusion or temporal smearing filters such as used in artificial reverberators would lead to perceptually similar results.

6 Summary and conclusion

Two simplified digital-filter approximations were suggested to account for the effects associated with scattered reflections at surfaces and objects for GA-based room acoustics simulation. Computationally highly-efficient all-pass cascades are used to mimic effects of scattering for each specular reflection in an image source model. The parameters of the all-pass cascades are chosen to either account for a monotonically decaying temporal spread caused by surface scattering or for a more gamma-distribution-shaped envelope of the spread caused by multiple scattered (inter-) reflections from objects in the room.

The suggested surface scattering approach transfers scattered sound energy of each reflection in the ISM to a diffuse reverberation model using decomposition filters based on the frequency-dependent scattering coefficient. The spatial spread is accounted for by mapping the scattered reflections to spatially distributed virtual reverberation sources used in the diffuse reverberation model, according to the solid angle covered by the considered surface.

The suggested object scattering, using a single parameter to specify the geometric deviation from an empty shoebox room, avoids unrealistic sparse specular reflections, particularly for large rooms, without the necessity to model a large number of individual reflecting surfaces.

Perceptual evaluation in comparison to real room recordings showed the relevance of scattering for highly transient signals such as a pulse and castanets, and room acoustics simulations including scattering effects were perceptually rated more similar to the real rooms. For a music excerpt, even including transients, no perceptual effect of scattering was observed.

A technical evaluation showed a considerably faster increase and comparable temporal evolution of echo density as observed in the real room recordings for simulations including the suggested scattering approaches. The results also showed a good agreement with the more complex ray

tracing method as an alternative GA approach, for which results critically depend on the inherent assumptions for echo density in the necessary conversion step from time-energy histograms to impulse responses.

Both suggested approaches are freely available in the framework of the room acoustics simulator (RAZR; www.razengine.com) and are highly suited for real-time applications.

Acknowledgments

The authors thank Torben Wendt for support in the data collection and figure compilation for the listening tests in the first experiment.

Funding

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

Conflicts of interest

The authors declare no conflict of interest.

Data availability statement

The data are available from the corresponding author on request.

Supplementary material

Aula_Reference_Castanets.wav: Castanets with measured BRIR in Aula Carolina.

Aula_Reference_Music.wav: Slapped bass guitar with measured BRIR in Aula Carolina.

Aula_Reference_Pulse.wav: Pulse with measured BRIR in Aula Carolina.

Aula_Surf0_Obj0_Castanets.wav: Castanets with simulated BRIR in Aula Carolina, no surface scattering, no object scattering.

Aula_Surf0_Obj0_Music.wav: Slapped bass guitar with simulated BRIR in Aula Carolina, no surface scattering, no object scattering.

Aula_Surf0_Obj0_Pulse.wav: Pulse with simulated BRIR in Aula Carolina, no surface scattering, no object scattering.

Aula_Surf0_Obj20_Castanets.wav: Castanets with simulated BRIR in Aula Carolina, no surface scattering, object scattering parameter 20%.

Aula_Surf0_Obj20_Music.wav: Slapped bass guitar with simulated BRIR in Aula Carolina, no surface scattering, object scattering parameter 20%.

Aula_Surf0_Obj20_Pulse.wav: Pulse with simulated BRIR in Aula Carolina, no surface scattering, object scattering parameter 20%.

Aula_Surf0_Obj5_Castanets.wav: Castanets with simulated BRIR in Aula Carolina, no surface scattering, object scattering parameter 5%.

Aula_Surf0_Obj5_Music.wav: Slapped bass guitar with simulated BRIR in Aula Carolina, no surface scattering, object scattering parameter 5%.

Aula_Surf0_Obj5_Pulse.wav: Pulse with simulated BRIR in Aula Carolina, no surface scattering, object scattering parameter 5%.

Aula_Surf025_Obj0_Castanets.wav: Castanets with simulated BRIR in Aula Carolina, surface scattering coefficient 0.25, no object scattering.

Aula_Surf025_Obj0_Music.wav: Slapped bass guitar with simulated BRIR in Aula Carolina, surface scattering coefficient 0.25, no object scattering.

Aula_Surf025_Obj0_Pulse.wav: Pulse with simulated BRIR in Aula Carolina, surface scattering coefficient 0.25, no object scattering.

Aula_Surf025_Obj5_Castanets.wav: Castanets with simulated BRIR in Aula Carolina, surface scattering coefficient 0.25, object scattering parameter 5%.

Aula_Surf025_Obj5_Music.wav: Slapped bass guitar with simulated BRIR in Aula Carolina, surface scattering coefficient 0.25, object scattering parameter 5%.

Aula_Surf025_Obj5_Pulse.wav: Pulse with simulated BRIR in Aula Carolina, surface scattering coefficient 0.25, object scattering parameter 5%.

Aula_Surf050_Obj0_Castanets.wav: Castanets with simulated BRIR in Aula Carolina, surface scattering coefficient 0.50, no object scattering.

Aula_Surf050_Obj0_Music.wav: Slapped bass guitar with simulated BRIR in Aula Carolina, surface scattering coefficient 0.50, no object scattering.

Aula_Surf050_Obj0_Pulse.wav: Pulse with simulated BRIR in Aula Carolina, surface scattering coefficient 0.50, no object scattering.

Underground_Reference_Pulse.wav: Pulse with measured BRIR in underground station.

Underground_Surf0_Obj20_Pulse.wav: Pulse with simulated BRIR in underground station, no surface scattering, object scattering parameter 20%.

Underground_Surf0_Obj5_Pulse.wav: Pulse with simulated BRIR in underground station, no surface scattering, object scattering parameter 5%.

Underground_Surf025_Obj0_Pulse.wav: Pulse with simulated BRIR in underground station, surface scattering coefficient 0.25, no object scattering.

Underground_Surf025_Obj5_Pulse.wav: Pulse with simulated BRIR in underground station, surface scattering coefficient 0.25, object scattering parameter 5%.

Underground_Surf050_Obj0_Pulse.wav: Pulse with simulated BRIR in underground station, surface scattering coefficient 0.50, no object scattering.

The audio files from Experiment 2 are available as supplementary material at <https://acta-acustica.edpsciences.org/10.1051/aacus/2024062/olm>.

References

1. L. Aspöck, S. Pelzer, F. Wefers, M. Vorländer: A real-time auralization plugin for architectural design and education, in: Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 3–5 April, 2014.
2. S. Pelzer, L. Aspöck, D. Schröder, M. Vorländer: Integrating real-time room acoustics simulation into a cad modeling software to enhance the architectural design process, *Buildings* 4 (2014) 113–138. <https://doi.org/10.3390/buildings4020113>.
3. J.H. Rindel: The use of computer modeling in room acoustics, *Journal of Vibroengineering* 3 (2000) 219–224.
4. N. Raghuvanshi, J. Snyder: Parametric directional coding for precomputed sound propagation, *ACM Transactions on Graphics* 37 (2018) 1–14. <https://doi.org/10.1145/3197517.3201339>.
5. S.V. Amengual Garí, C. Schissler, R. Mehra, S. Featherly, P. Robinson: Evaluation of real-time sound propagation engines in a virtual reality framework, in: Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio, York, UK, 27–29 March, 2019.
6. V. Valimaki, J.D. Parker, L. Savioja, J.O. Smith, J.S. Abel: Fifty years of artificial reverberation, *IEEE Transactions on Audio, Speech, and Language Processing* 20 (2012) 1421–1448. <https://doi.org/10.1109/TASL.2012.2189567>.
7. F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, S. Weinzierl: A round robin on room acoustical simulation and auralization, *Journal of the Acoustical Society of America* 145 (2019) 2746–2760. <https://doi.org/10.1121/1.5096178>.
8. L. Savioja, U.P. Svensson: Overview of geometrical room acoustic modeling techniques, *Journal of the Acoustical Society of America* 138 (2015) 708–730. <https://doi.org/10.1121/1.4926438>.
9. B.U. Seeber, S.W. Clapp: Interactive simulation and free-field auralization of acoustic space with the rTSOFE, *Journal of the Acoustical Society of America* 141 (2017) 3974–3974. <https://doi.org/10.1121/1.4989063>.
10. A. Ahrens, K.D. Lund, M. Marschall, T. Dau: Sound source localization with varying amount of visual information in virtual reality, *PLoS One* 14 (2019) e0214603. <https://doi.org/10.1371/journal.pone.0214603>.
11. F. Pausch, J. Fels: Localization performance in a binaural real-time auralization system extended to research hearing aids, *Trends in Hearing* 24 (2020) 2331216520908704. <https://doi.org/10.1177/2331216520908704>.
12. H. Hu, L. Zhou, H. Ma, Z. Wu: HRTF personalization based on artificial neural network in individual virtual auditory space, *Applied Acoustics* 69 (2008) 163–172. <https://doi.org/10.1016/j.apacoust.2007.05.007>.
13. F. Pausch, L. Aspöck, M. Vorländer, J. Fels: An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss, *Trends in Hearing* 22 (2018) 2331216518800871. <https://doi.org/10.1177/2331216518800871>.
14. D. Schröder: Physically based real-time auralization of interactive virtual environments, Rheinisch-Westfälischen Technischen Hochschule Aachen, Aachen, Germany, 2011.
15. H. Hu, L. Zhou, H. Ma, F. Yang, Z. Wu: Externalization of headphone based virtual sound system, *Journal of Southeast University (Natural Science Edition)* 38 (2008) 1–5.
16. J.B. Allen, D.A.: Image method for efficiently simulating small-room acoustics, *Journal of the Acoustical Society of America* 65 (1979) 943–950. <https://doi.org/10.1121/1.382599>.
17. J. Borish: Extension of the image model to arbitrary polyhedra, *Journal of the Acoustical Society of America* 75 (1984) 1827–1836. <https://doi.org/10.1121/1.390983>.
18. A. Krokstad, S. Strom, S. Sørsdal: Calculating the acoustical room response by the use of a ray tracing technique, *Journal of Sound and Vibration* 8 (1968) 118–125. [https://doi.org/10.1016/0022-460x\(68\)90198-3](https://doi.org/10.1016/0022-460x(68)90198-3).
19. M. Vorländer: Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm, *Journal of the Acoustical Society of America* 86 (1989) 172–178. <https://doi.org/10.1121/1.398336>.
20. A.D. Pierce: Diffraction of sound around corners and over wide barriers, *Journal of the Acoustical Society of America* 55 (1974) 941–955. <https://doi.org/10.1121/1.1914668>.

21. R.G. Kouyoumjian, P.H. Pathak: A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface, *Proceedings of the IEEE* 62 (1974) 1448–1461. <https://doi.org/10.1109/PROC.1974.9651>.
22. U.P. Svensson, R.I. Fred, J. Vanderkooy: An analytic secondary source model of edge diffraction impulse responses, *Journal of the Acoustical Society of America* 106 (1999) 2331–2344. <https://doi.org/10.1121/1.428071>.
23. M.A. Biot, I. Tolstoy: Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction, *Journal of the Acoustical Society of America* 29 (1957) 381–391. <https://doi.org/10.1121/1.1908899>.
24. C. Kirsch, S.D. Ewert: Low-order filter approximation of diffraction for virtual acoustics, in: 2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 17–20 October, IEEE, 2021, pp. 341–345.
25. S.D. Ewert: A filter representation of diffraction at infinite and finite wedges, *JASA Express Letters* 2 (2022) 092401. <https://doi.org/10.1121/10.0013686>.
26. C. Kirsch, S.D. Ewert: A universal filter approximation of edge diffraction for geometrical acoustics, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (2023) 1636–1651. <https://doi.org/10.1109/taslp.2023.3264737>.
27. A. Erraji, J. Stienen, M. Vorländer: The image edge model, *Acta Acustica* 5 (2021) 17. <https://doi.org/10.1051/aacus/2021010>.
28. C. Schissler, G. Mückl, P. Calamia: Fast diffraction pathfinding for dynamic sound propagation, *ACM Transactions on Graphics (TOG)* 40 (2021) 1–13. <https://doi.org/10.1145/3450626.3459751>.
29. V. Pulkki, U.P. Svensson: Machine-learning-based estimation and rendering of scattering in virtual reality, *Journal of the Acoustical Society of America* 145 (2019) 2664–2676. <https://doi.org/10.1121/1.5095875>.
30. J. Mannall, L. Savioja, P. Calamia, R. Mason, E. De Sena: Efficient diffraction modeling using neural networks and infinite impulse response filters, *Journal of the Audio Engineering Society* 71 (2023) 566–576. <https://doi.org/10.17743/jaes.2022.0107>.
31. M. Vorländer, E. Mommertz: Definition and measurement of random-incidence scattering coefficients, *Applied Acoustics* 60 (2000) 187–199. [https://doi.org/10.1016/S0003-682X\(99\)00056-0](https://doi.org/10.1016/S0003-682X(99)00056-0).
32. H. Kuttruff: A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries, *Journal of the Acoustical Society of America* 98 (1995) 288–293. <https://doi.org/10.1121/1.413727>.
33. R.R. Torres, M. Kleiner, B.-I. Dalenbäck: Audibility of “diffusion” in room acoustics auralization an initial investigation, *Acta Acustica united with Acustica* 86 (2000) 919–927.
34. L. Aspöck, M. Vorländer: Differences between measured and simulated room impulse responses, in: INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Glasgow, Scotland, 21–24 August, 2022, Institute of Noise Control Engineering, 2023, pp. 3209–3217.
35. J. S. Abel, P. Huang: A simple, robust measure of reverberation echo density, *Audio Engineering Society Convention* 121, San Francisco, CA, USA, 5–8 October, 2006.
36. S. Siltanen, T. Lokki, S. Tervo, L. Savioja: Modeling incoherent reflections from rough room surfaces with image sources, *Journal of the Acoustical Society of America* 131 (2012) 4606–4614. <https://doi.org/10.1121/1.4711013>.
37. J.-M. Jot, A. Chaigne: Digital delay networks for designing artificial reverberators, in: *Audio Engineering Society Convention* 90, Paris, France, 19–22 February, 1991.
38. E.A. Lehmann, A.M. Johansson: Diffuse reverberation model for efficient image-source simulation of room impulse responses, *IEEE Transactions on Audio, Speech, and Language Processing* 18 (2010) 1429–1439. <https://doi.org/10.1109/tasl.2009.2035038>.
39. V. Välimäki, K. Prawda: Late-reverberation synthesis using interleaved velvet-noise sequences, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021) 1149–1160. <https://doi.org/10.1109/taslp.2021.3060165>.
40. B. Alary, A. Politis, S.J. Schlecht, V. Välimäki: Directional feedback delay network, *Journal of the Audio Engineering Society* 67 (2019) 752–762. <https://doi.org/10.17743/jaes.2019.0026>.
41. C. Kirsch, T. Wendt, S. Van De Par, H. Hu, S.D. Ewert: Computationally-efficient simulation of late reverberation for inhomogeneous boundary conditions and coupled rooms, *Journal of the Audio Engineering Society* 71 (2023) 186–201. <https://doi.org/10.17743/jaes.2022.0053>.
42. S. Riedel, M., Smith F. Zotter: The effect of temporal and directional density on listener envelopment, *Journal of the Audio Engineering Society* 71 (2023) 455–467. <https://doi.org/10.17743/jaes.2022.0088>.
43. T. Wendt, S. Van de Par, S.D. Ewert: A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation, *Journal of the Audio Engineering Society* 62 (2014) 748–766. <https://doi.org/10.17743/jaes.2014.0042>.
44. S. Siltanen, T. Lokki, S. Kiminki, L. Savioja: The room acoustic rendering equation, *Journal of the Acoustical Society of America* 122 (2007) 1624–1635. <https://doi.org/10.1121/1.2766781>.
45. H. Kuttruff: Simulated reverberation curves in rectangular rooms with diffuse sound fields, *Acta Acustica united with Acustica* 25 (1971) 333–342.
46. E.-M. Nosal, M. Hodgson, I. Ashdown: Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms, *Journal of the Acoustical Society of America* 116 (2024) 970–980.
47. S. Siltanen, T. Lokki, L. Savioja: Frequency domain acoustic radiance transfer for real-time auralization, *Acta Acustica united with Acustica* 95 (2009) 106–117. <https://doi.org/10.3813/AAA.918132>.
48. L. Antani, A. Chandak, L. Savioja, D. Manocha: Interactive sound propagation using compact acoustic transfer operators, *ACM Transactions on Graphics* 31 (2012) 1–12. <https://doi.org/10.1145/2077341.2077348>.
49. C. Schissler, R. Mehra, D. Manocha: High-order diffraction and diffuse reflections for interactive sound propagation in large environments, *ACM Transactions on Graphics* 33 (2014) 1–12. <https://doi.org/10.1145/2601097.2601216>.
50. M.R. Schroeder: Natural sounding artificial reverberation, *Journal of the Audio Engineering Society* 10 (1962) 219–223.
51. E.D. Sena, H. Hacıhabiboğlu, Z. Cvetković, J.O. Smith: Efficient synthesis of room acoustics via scattering delay networks, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23 (2015) 1478–1492. <https://doi.org/10.1109/taslp.2015.2438547>.
52. T.B. Atalay, Z.S. Gul, E. De Sena, Z. Cvetkovic, H. Hachabiboglu: Scattering delay network simulator of coupled volume acoustics, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022) 582–593. <https://doi.org/10.1109/taslp.2022.3143697>.

53. U. Stephenson: Eine Schallteilchen-Computersimulation zur Berechnung der für die Hörsamkeit in Konzertsälen massgebenden Parameter, *Acta Acustica united with Acustica* 59 (1985) 1–20.
54. J.H. Rindel: Computer simulation techniques for acoustical design of rooms, *Acoustics Australia* 23 (1995) 81–86.
55. D. Schröder, M. Vorländer; Raven: A real-time framework for the auralization of interactive virtual environments, in: *Forum Acusticum, European Acoustics Association*, 2011, pp. 1541–1546.
56. T. Lewers: A combined beam tracing and radiatn exchange computer model of room acoustics, *Applied Acoustics* 38 (1993) 161–178.
57. K.H. Kuttruff: Auralization of impulse responses modeled on the basis of ray-tracing results, *Journal of the Audio Engineering Society* 41 (1993) 876–880.
58. L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen: Creating interactive virtual acoustic environments, *Journal of the Audio Engineering Society* 47 (1999) 675–705.
59. S.D. Ewert, N. Gößling, O. Buttler, S. van de Par, H. Hu: Computationally-efficient and perceptually-motivated rendering of diffuse reflections in room acoustics simulation, in: *10th Convention of the European Acoustics Association (Forum Acusticum 2023)*, Turin, Italy, 11–15 September, 2023.
60. T.J. Cox, P. D’Antonio: *Acoustic absorbers and diffusers: theory, design and application*, 2nd edn., Taylor & Francis, London, New York, 2009.
61. P.W. Robinson, A. Walther, C. Faller, J. Braasch: Echo thresholds for reflections from acoustically diffusive architectural surfaces, *Journal of the Acoustical Society of America* 134 (2013) 2755–2764. <https://doi.org/10.1121/1.4820890>.
62. S.J. Schlecht, E.A.P. Habets: Scattering in feedback delay networks, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020) 1915–1924. <https://doi.org/10.1109/taslp.2020.3001395>.
63. T. Carpentier, M. Noisternig, O. Warusfel: Hybrid reverberation processor with perceptual control, in: *17th International Conference on Digital Audio Effects-DAFx-14*, Erlangen, Germany, 1–5 September, 2014, pp. 93–100.
64. J.-M. Jot, R. Audfray, M. Hertensteiner, B. Schmidt: Rendering spatial sound for interoperable experiences in the audio metaverse, in: *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Bologna, Italy, 8–10 September, IEEE, 2021, pp. 1–15.
65. H. Kuttruff: *Room acoustics*, 6th edn., CRC Press, Boca Raton, FL, 2017, p. 97.
66. C. Kirsch, J. Poppitz, T. Wendt, S. van de Par, S.D. Ewert: Spatial resolution of late reverberation in virtual acoustic environments, *Trends in Hearing* 25 (2021) 1–17. <https://doi.org/10.1177/233121652110054>.
67. C. Kirsch, J. Poppitz, T. Wendt, S.V.D. Par, S.D. Ewert: Computationally efficient spatial rendering of late reverberation in virtual acoustic environments, in: *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Bologna, Italy, 8–10 September, IEEE, 2021, pp. 1–8.
68. A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, S. Weinzierl: A spatial audio quality inventory (SAQI), *Acta Acustica united with Acustica* 100 (2014) 984–994. <https://doi.org/10.3813/AAA.918778>.
69. V.L. Jordan: Acoustical criteria for auditoriums and their relation to model techniques, *Journal of the Acoustical Society of America* 47 (1970) 408–412. <https://doi.org/10.1121/1.1911535>.
70. M.R. Schroeder: Digital simulation of sound transmission in reverberant spaces, *Journal of the Acoustical Society of America* 47 (1970) 424–431. <https://doi.org/10.1121/1.1911541>.
71. M.R. Schroeder, B.F. Logan: Colorless artificial reverberation, *IRE Transactions on Audio AU-9* (1961) 209–214. <https://doi.org/10.1109/TAU.1961.1166351>.
72. M. Jeub, M. Schafer, P. Vary: A binaural room impulse response database for the evaluation of dereverberation algorithms, in: *16th International Conference on Digital Signal Processing*, Santorini, Greece, 5–7 July, IEEE, 2009, pp. 1–5.
73. L. Hladek, B. Seeber: Underground station environment (1.1) [dataset], Zenodo, 2022. Available at <https://doi.org/10.5281/zenodo.6025631>.
74. S. van de Par, S.D. Ewert, L. Hladek, C. Kirsch, J. Schütze, J. Llorca-Bofi, G. Grimm, M.M. Hendrikse, B. Kollmeier, B.U. Seeber: Auditory-visual scenes for hearing research, *Acta Acustica* 6 (2022) 55. <https://doi.org/10.1051/aacus/2022032>.
75. S. Weinzierl, S. Lepa, D. Ackermann: A measuring instrument for the auditory perception of rooms: the room acoustical quality inventory (RAQI), *Journal of the Acoustical Society of America* 144 (2018) 1245–1257. <https://doi.org/10.1121/1.5051453>.
76. ITU-R: Recommendation ITU-R BS.1534–3 method for the subjective assessment of intermediate quality level of audio systems, *International Telecommunication Union Radiocommunication Assembly*, 2015.
77. L. Cremer: *Die wissenschaftlichen Grundlagen der Raumakustik. Band I: Geometrische Raumakustik*, Hirzel, Stuttgart, 1948.
78. M.A. Biot: Generalized boundary condition for multiple scatter in acoustic reflection, *Journal of the Acoustical Society of America* 44 (1968) 1616–1622.
79. M.A. Biot: On the reflection of acoustic waves on a rough surface, *Journal of the Acoustical Society of America* 30 (1958) 479–480.
80. M.A. Biot: Reflection on a rough surface from an acoustic point source, *Journal of the Acoustical Society of America* 29 (1957) 1193–1200.
81. A. Haeussler, S. van de Par: Crispness, speech intelligibility, and coloration of reverberant recordings played back in another reverberant room (room-in-room), *The Journal of the Acoustical Society of America* 145 (2019) 931–944. <https://doi.org/10.1121/1.5090103>.
82. J. Fagerström, B. Alary, S.J. Schlecht, V. Välimäki: Velvet-noise feedback delay network, in: *23rd International Conference on Digital Audio Effects (DAFx)*, Vienna, Austria, 8–12 September, 2020, pp. 219–226.
83. S.J. Schlecht, E.A.P. Habets: Feedback delay networks: Echo density and mixing time, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25 (2017) 374–383. <https://doi.org/10.1109/taslp.2016.2635027>.
84. S.J. Schlecht, E.A.P. Habets: Dense reverberation with delay feedback matrices, in: *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 20–23 October, IEEE, 2019, pp. 150–154.
85. T. Lokki, J. Pätynen, S. Tervo, S. Siltanen, L. Savioja: Engaging concert hall acoustics is made up of temporal envelope preserving reflections, *Journal of the Acoustical Society of America* 129 (2011) EL223–EL228. <https://doi.org/10.1121/1.3579145>.
86. H. Haas: The influence of a single echo on the audibility of speech, *Acustica* 1 (1951) 49–58.
87. H. Wallach, E.B. Newman, M.R. Rosenzweig: The precedence effect in sound localization, *American Journal of Psychology* 62 (1949) 315–336. <https://doi.org/10.2307/1418275>.

Cite this article as: Ewert SD, Gößling N, Buttler O, van de Par S, & Hu H. 2025. Computationally-efficient rendering of diffuse reflections for geometrical acoustics based room simulation. Acta Acustica, **9**, 9. <https://doi.org/10.1051/aacus/2024062>.