Diffuse reverberation model for efficient image-source simulation of room impulse responses

Eric A. Lehmann*, *Member, IEEE,* and Anders M. Johansson

***Abstract*— In many research fields of engineering and acoustics, the image-source model represents one of the most popular tools for the simulation of sound fields in virtual reverberant environments. This can be seen as a result from the relative simplicity and flexibility of this particular method. However, the associated computational costs constitute a well known drawback of image-source implementations, as the required simulation times grow exponentially with the considered reflection order. This paper proposes a method that allows for a fast synthesis of room impulse responses according to the image-source technique. This is achieved by modeling the diffuse reverberation tail as decaying random noise, where the decay envelope for the considered acoustic environment is determined according to a recently proposed method for the prediction of energy decay curves in image-source simulations. The diffuse reverberation model presented in this paper thus produces impulse responses that are representative of the specific virtual environment under consideration (within the general assumptions of geometrical room acoustics), in contrast to other artificial reverberation techniques developed on the basis of perceptual measures or assuming a purely exponential energy decay. Furthermore, since image-source simulations are only used for the computation of the early reflections, the proposed approach achieves a reduction of the computational requirements by up to two orders of magnitude for the simulation of full-length room impulse responses, compared to a standard image-source implementation.**

***Index Terms*— Room acoustics, image-source method, room impulse response, energy decay, diffuse reverberation model. EDICS category: AUD-ROOM.**

I. Introduction

The development and refinement of techniques for the simulation of sound fields in room acoustics has attracted a significant research effort to date, with the scientific literature containing a vast array of such techniques. Among others, these include many methods based on the well known principles of ray tracing [1], beam tracing [2, 3], boundary and finite element methods [4, 5], digital waveguide meshes [6, 7], and image-source methods [8–10]. Alternative approaches are also regularly proposed on the basis of innovative models of the sound propagation in acoustic environments [11–14].

Despite this abundance of sophisticated room-acoustics simulation methods, the relatively basic image-source model (ISM) of Allen & Berkley [8] remains as popular as ever as a technique for the simulation of sound propagation in reverberant settings. This popularity is attested by the ever increasing number of scientific publications relying on ISM simulations for a variety of purposes such as generating stimuli for perceptual and psychoacoustic tests [15, 16], validating algorithms or systems designed to operate in reverberant conditions [17, 18], sound field analysis and synthesis [19–21], sound rendering and auralization in virtual auditory systems [22, 23], and design of acoustic spaces [24, 25]. The prominence of the ISM technique can be attributed to a number of important benefits compared to other approaches based on simulation, scale modeling or physical measurements, including mainly:

i) simplicity of algorithmic implementation,
ii) high degree of flexibility, with many simulation parameters (such as room dimensions, acoustic absorption coefficients, source and microphone positions, reverberation time, etc.) adjustable in software,
iii) the ability to generate realistic room impulse responses that are very similar to those obtained from real-room measurements [10, Fig. 1],
iv) the ability to investigate the effects of reverberation in isolation, separately from other sources of disturbance such as additive noise,
v) the guarantee to find all valid specular reflections in a given environment, which is critical when modeling the early part of a RIR where individual reflections (or the lack thereof) can have a large effect on the perceived acoustic characteristics.

On the other hand, the most significant and well known drawback of an image-source implementation is its intrinsic computational cost. The number of image sources necessary for the computation of a room impulse response (RIR) by means of the ISM is proportional to the cube of the RIR length, and grows exponentially with the reflection order [26]. This large computational requirement is typically prohibitive in real-time applications such as immersive or interactive auditory environments [2, 12, 22]. Similarly, the ISM can become overly burdensome when a large number of RIRs need to be computed at once [27, 28]. For instance, in the case of a moving sound source in microphone array applications [29], the source trajectory is typically discretized into a series of finely-spaced locations, from which the RIRs to each considered sensor are computed. The audio data at the microphones is then obtained by convolving successive frames of source signal with the computed RIRs, using an overlap-add scheme. In such applications, sampling the source trajectory too coarsely usually leads to adjacent RIRs (to a given sensor) differing significantly from each other, which can generate clicking noise and other audible artifacts in the resulting audio signals. The computational costs of the ISM technique are further exacerbated by the need to simulate "full-length" RIRs, i.e., to

*Corresponding author. E. A. Lehmann is with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Division of Mathematical & Information Sciences, Private Bag 5, Wembley WA 6913, Australia (e-mail: Eric.Lehmann@csiro.au, tel.: +61 (0)8 9333 6123, fax: +61 (0)8 9333 6121).

A. M. Johansson is with the Swedish Defence Research Agency, Electronic Warfare Division, Linköping, Sweden (e-mail: ajh@foi.se).

ensure that most of the RIR energy has been captured without significant cropping. This is typically necessary so as to avoid unnatural-sounding reverberation or unpredictable effects on the audio processing system under scrutiny. A number of different techniques have been presented in the literature in an attempt to mitigate the computational costs of the ISM method [27, 28, 30–33].

The decomposition of a RIR into two parts, namely early reflections and late (diffuse) reverberation, is a widely used model in the acoustics and engineering literature [18, 23, 31, 34]. The early part of the RIR typically contains the direct path together with a series of discrete early reflections. The diffuse part is characterized by a reflection density that is too high for individual reflections to be discerned in the received signal. Due to its noise-like nature, the late reverberation part is commonly modeled as exponentially decaying Gaussian noise [18, 34], which forms the basis of many artificial reverberation techniques, for instance [35, 36]. While providing acceptable perceptual results, the assumption of a purely exponential RIR decay however fails to account for the particular characteristics of the considered acoustic environment. A nonuniform definition of the wall absorption coefficients, for instance, will typically produce a nonexponential decay of the sound energy [32, 34]. As an alternative, existing acoustic environments can also be explicitly measured in order to model the diffuse decay of a given real room by means of a parametric algorithm for artificial reverberation [23, 36]. However, this represents a solution usually involving substantial hardware setup costs.

This article presents a new method to simulate RIRs in virtual acoustic environments, based on the following diffuse reverberation modeling (DRM) approach. It uses the above mentioned decomposition of the RIR into early reflections and late reverberation, where the early part is simulated according to the standard ISM technique. For the late part of the RIR, the proposed DRM approach draws on the work presented by the authors in [10], which describes a method for the prediction of the energy decay curve (EDC) in ISM-simulated RIRs. It is shown here how this EDC prediction technique, which is briefly reviewed in Section II, can be used to compute an accurate envelope for the synthesis of the diffuse RIR tail as a decaying random noise process. The resulting decay in the simulated RIR is therefore representative of the considered environment, rather than based on perceptual or other theoretical measures. Section III presents the details of the proposed DRM approach, based on the results obtained in [10] as well as a statistical analysis of the reverberant tail of ISM-simulated RIRs. This is followed by a discussion on the limitations of the adopted model in Section IV. By restricting the use of image-source simulations to the early reflections only, the DRM method presented in this paper allows for a significant reduction of the computational costs without compromising the statistical accuracy and energy content of the diffuse RIR decays. Experimental results demonstrating the accuracy of the proposed technique are provided in Section V, together with a quantitative analysis of the resulting savings in computational requirements.

It is important to note that this paper only attempts to *reproduce* the results from ISM simulations. The aim is not to improve or modify the acoustic model assumed by the ISM technique. Due to several simplifying assumptions (empty rectangular room, lack of diffraction and diffusion effects, etc.), the shortcomings of Allen & Berkley's ISM algorithm in modeling physical acoustics phenomena are well known, but this particular issue does not represent the focus of this work.

II. Background review

This section provides a brief review of the two concepts central to the work presented later in Section III, namely the image-source simulation technique and the method proposed by the authors for the prediction of energy decay curves.

A. Image-source method

The image-source method was originally presented for rectangular enclosures in [8] and can be briefly described as follows. A model of a shoebox enclosure with dimensions $L_x \times L_y \times L_z$ contains a sound source and an acoustic receiver, located at $\boldsymbol{p}_s = [x_s\ y_s\ z_s]^{\mathrm{T}}$ and $\boldsymbol{p}_r = [x_r\ y_r\ z_r]^{\mathrm{T}}$, respectively. The acoustic properties of this virtual room are characterized by means of a reflection coefficient β for each of the six enclosure surfaces: $\boldsymbol{\beta} = [\beta_{x,1}\ \beta_{x,2}\ \beta_{y,1}\ \beta_{y,2}\ \beta_{z,1}\ \beta_{z,2}]^{\mathrm{T}}$.

In order to determine the RIR between the source and the receiver, the ISM technique assumes specular reflections of the sound on the enclosure boundaries (assumptions of geometrical room acoustics). The RIR $h(t)$, $t \geqslant 0$, is then obtained by considering a grid of mirrored image sources extending in all dimensions, and by summing the contribution of each image source at the receiver:

$$h(t) = \sum_{\boldsymbol{u}=0}^{1} \sum_{\boldsymbol{v}=-\infty}^{\infty} A(\boldsymbol{u}, \boldsymbol{v}) \cdot \delta\big(t - \tau(\boldsymbol{u}, \boldsymbol{v})\big)\,, \tag{1}$$

where $\boldsymbol{u} = (u_x, u_y, u_z)$ and $\boldsymbol{v} = (v_x, v_y, v_z)$ are triplet parameters controlling the indexing of the image sources in all dimensions, $A(\cdot)$ is the amplitude factor and $\tau(\cdot)$ the time delay of the considered image source, respectively. In (1), the sum over $\boldsymbol{u}$ (respectively $\boldsymbol{v}$) is used to represent a triple sum over each of the triplet's internal indices (see [8] for more details).

Once the RIR $h(\cdot)$ is computed, the energy decay curve $E(\cdot)$ can be obtained using the normalized Schroeder integration method:

$$E(t) = 10 \cdot \log_{10}\left(\frac{\int_t^\infty h^2(\xi)\,\mathrm{d}\xi}{\int_0^\infty h^2(\xi)\,\mathrm{d}\xi}\right). \tag{2}$$

Many ISM variants and extensions based on this standard technique have been proposed in the literature [9, 32, 37, 38]. This paper makes use of the modified ISM algorithm described in [10], and readers are referred to this previous work for more detailed information.

B. EDC prediction method

The work presented in [10] provides a mathematical formulation of the energy decay in ISM-simulated RIRs. This EDC prediction method is based on the ISM principles reviewed in Section II-A and is thus also limited to the case of

$$\hat{h}_P(t) = \frac{1}{8\rho\,\bar{r}} \cdot \begin{cases} \frac{B_z}{\log\left(\frac{B_y}{B_x}\right)} \cdot \Big(\mathrm{Ei}\big(\log\big(\frac{B_z}{B_x}\big)\big) + \log\big(\log\big(\frac{B_z}{B_x}\big)\big) - \mathrm{Ei}\big(\log\big(\frac{B_z}{B_y}\big)\big) - \log\big(\log\big(\frac{B_z}{B_y}\big)\big) \Big) & \text{if } B_x \neq B_y \neq B_z\,, \\ \frac{B_z}{\log\left(\frac{B_z}{B}\right)} \cdot \Big(\mathrm{Ei}\big(\log\big(\frac{B_z}{B}\big)\big) + \log\big(\log\big(\frac{B_z}{B}\big)\big) + \gamma \Big) & \text{if } B_z = B_y \neq B_x \triangleq B \text{ or } B_z = B_x \neq B_y \triangleq B\,, \\ \frac{B - B_z}{\log\left(\frac{B}{B_z}\right)} & \text{if } B_z \neq B_x = B_y \triangleq B\,, \\ B & \text{if } B_x = B_y = B_z \triangleq B\,. \end{cases} \tag{3}$$

small-room acoustics in rectangular enclosures with specular reflections. The approach used for EDC prediction relies on the following observation: the acoustic power $h_P(t)$ received at the microphone at a given time t corresponds to the addition of the contributions from all the image sources located on a sphere of radius $\rho = c \cdot t$ around the receiver (c denotes the propagation speed of acoustic waves). Given the known position of each image source on the considered grid, the discrete summation is then considered as a Riemann sum which can be turned into the integral of a continuous function over the sphere. Analytically solving the integration finally yields a closed-form expression for the approximated power impulse response $\hat{h}_P(\cdot)$. This expression is reproduced here in (3) from [10] for convenience, with $\gamma = 0.5772157\ldots$ the Euler–Mascheroni constant, $\mathrm{Ei}(\cdot)$ denoting the first-order exponential integral, and with the following definitions:

$$B_x = (\beta_{x,1}\beta_{x,2})^{\rho/L_x}, \tag{4}$$

$$B_y = (\beta_{y,1}\beta_{y,2})^{\rho/L_y}, \tag{5}$$

$$B_z = (\beta_{z,1}\beta_{z,2})^{\rho/L_z}, \tag{6}$$

$$\bar{r} = \frac{L_x + L_y + L_z}{3}\,. \tag{7}$$

On the basis of (2) and (3), an estimate $\widehat{E}(\cdot)$ of the EDC can be obtained as

$$\widehat{E}(t) = 10 \cdot \log_{10}\left(\frac{\int_t^\infty \hat{h}_P(\xi)\,\mathrm{d}\xi}{\int_0^\infty \hat{h}_P(\xi)\,\mathrm{d}\xi}\right). \tag{8}$$

As with the standard ISM algorithm reviewed in Section II-A, this EDC prediction method can also be extended to include additional parameters such as frequency and angle-dependent reflection coefficients, directional sound source and/or microphone, etc. Such implementations are however not carried out in this work for the sake of conciseness and presentation clarity. Readers are further referred to [10] for more detail on the EDC approximation method reviewed in this subsection.

III. Proposed RIR synthesis method

A. Simulation concept

The basis of the proposed RIR simulation technique relies on the observation that the tail of a typical RIR usually presents a distinct noise-like character, due to the isotropic addition of a multitude of late acoustic reflections (diffuse field). This is illustrated in Figure 1, which displays an example of ISM-simulated RIR obtained for a $6.6 \times 6.1 \times 2.9\,\mathrm{m}^3$ room with $\boldsymbol{p}_r = [2.6\ 3.2\ 0.8]^\mathrm{T}$ and $\boldsymbol{p}_s = [1.2\ 1.9\ 1.1]^\mathrm{T}$ (in m), and using nonuniform reflection coefficients leading to a reverberation time of $T_{60} \approx 0.3\,\mathrm{s}$.

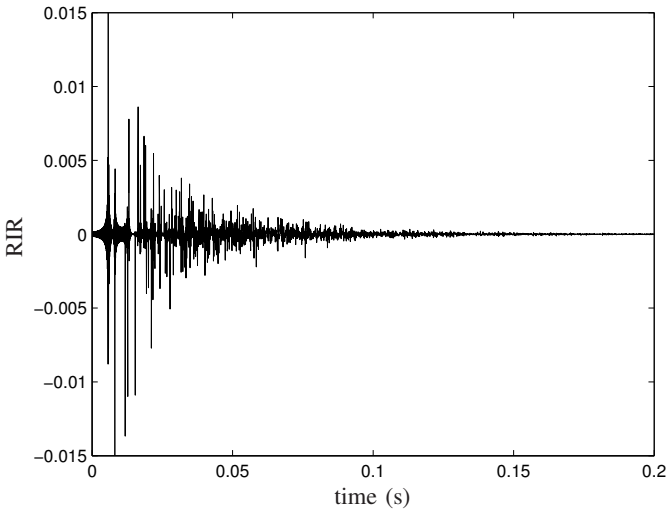

Fig. 1. Typical example of ISM-simulated RIR, computed for a $6.6 \times 6.1 \times 2.9\,\mathrm{m}^3$ room with nonuniform reflection coefficients and $T_{60} \approx 0.3\,\mathrm{s}$ (the ordinate axis has been cropped for display purposes).

In practice, a large computational effort is spent by the ISM algorithm on the computation of the higher-order reflections that are ultimately combined to produce this diffuse part of the RIR. Significant computation costs can therefore be avoided by modeling the tail of the RIR as a decaying random signal. Existing methods based on this principle typically assume a purely exponential decay of the RIR energy [18, 34, 39]. This however represents an approximation that is not necessarily valid in practice when attempting to model a given acoustic environment. To achieve a more accurate modeling, the formula for the RIR power in (3) can be used here in order to generate a decay that is true to the specific environment under consideration.

While the decaying random noise assumption provides a good model of the late reverberation, it is clear that this approach cannot be used for the earlier part of the RIR. The strong discrete reflections usually present during that phase cannot be well replicated using a pure noise signal. Also, the various assumptions underlying the derivation of the EDC prediction technique in [10], and leading to the expression in (3), are known to be inaccurate for $t \to 0$ (early part of the RIR). It is therefore not possible to guarantee that the expression in (3) will provide a good approximation of the RIR's very early decay. As a result, standard ISM-based simulation must still be used to generate the early reflections in the RIR.

B. Transition from early reflections to diffuse field

The RIR model assumed so far requires the definition of a specific transition time between the early reflections and the late reverberation part in the RIR. This cut-off point, denoted as t_c in this work, effectively determines how much of the total RIR length is simulated with the ISM, and hence strongly influences the computational load of the resulting algorithm.

A typical definition of the transition time commonly found in the literature is $t_c = 80\,\text{ms}$ [34, 39]. This value can however be seen as being somewhat arbitrary, and several authors use the definition of a more generic transition period ranging from as low as $50\,\text{ms}$ and up to $150\,\text{ms}$ [36, 40]. The acoustics literature contains various other definitions of the t_c parameter, including some resulting from psychoacoustic studies (see, e.g., [41] and the references therein).

For the purpose of the proposed RIR synthesis method, t_c should be chosen large enough so as to ensure that the characteristic features of the early reflections are left unaltered. In other words, the t_c value should be such that the decaying-noise model is only applied to the diffuse part of the RIR. In this work, we define the transition point t_c as the time for which the overall acoustic energy in the RIR has decreased by a certain amount Δ_c (in dB):

$$t_c \triangleq E^{-1}(-\Delta_c)\,, \tag{9}$$

where $E^{-1}(\xi)$ corresponds to the time lag t for which $E(t) = \xi$. This effectively defines a room-dependent cut-off parameter that can be tuned to capture most of the early reflections in the specific environment under consideration.

C. Modeling the diffuse reverberation decay

Once the early reflections, i.e., $h(t)$ for $t \leqslant t_c$, have been computed using the ISM technique, the diffuse part of the RIR can be simulated by generating a noise signal whose energy decay is determined by the EDC for the considered acoustic environment. To this purpose, the expression in (3) is used in order to obtain an accurate prediction of this specific energy decay pattern.

The power impulse response $\hat{h}_P(\cdot)$ described in Section II-B effectively corresponds to the instantaneous acoustic power in the RIR $h(\cdot)$ (power envelope). This relation can be mathematically expressed as

$$\hat{h}_P(t) \approx \frac{\lambda}{T} \cdot \int_{t-T/2}^{t+T/2} h^2(\xi)\,\mathrm{d}\xi\,, \tag{10}$$

for a small window length T. The parameter λ is a multiplicative constant introduced for normalization purpose, and can be explained as follows.

As mentioned earlier, the model used as the basis for the EDC approximation technique in [10] is known to become inaccurate as $t \to 0$, i.e., for small values of the sphere radius ρ (see Section II-B). This in turn leads to issues when attempting to properly normalize the expression for $\hat{h}_P(\cdot)$ given in (3). Hence, this expression effectively delivers a "relative" measure rather than the absolute value of the power impulse response.[1] In the present work however, a properly normalized expression for $\hat{h}_P(\cdot)$ is needed in order to achieve a proper correspondence, as given by (10), between the squared ISM-simulated RIR $h^2(\cdot)$ and the approximated power values $\hat{h}_P(\cdot)$ (which are used for the simulation of the late reverberations). This normalization is here effectively achieved with the introduction of the constant factor λ in (10).

In practice, an estimate for λ can be determined by adjusting $\hat{h}_P(\cdot)$ to the already computed ISM results (early reflections), which can be achieved as follows. The ISM-simulated part of the RIR $h(t)$, previously computed for the early reflections up to time t_c, is subdivided into K frames of length T. Thus, $K = \lfloor t_c/T \rfloor$ and the frame centers are defined as $t_k = (k - 1/2) \cdot T$ for $k = 1, 2, \ldots, K$. For each frame, an estimate of the RIR power can be computed by integrating $h^2(\cdot)$ over the considered time interval. The value of the predicted power function $\hat{h}_P(\cdot)$ can also be determined according to (3) for the same sequence of frame center points t_k. Using the index-based notation

$$\hat{h}_P(k) \triangleq \hat{h}_P(t)\big|_{t=(k-1/2)\cdot T}\,, \tag{11}$$

an estimate of the factor λ for each frame k then follows directly from (10) as

$$\lambda_k = \frac{T \cdot \hat{h}_P(k)}{\int_{kT-T}^{kT} h^2(\xi)\,\mathrm{d}\xi}\,. \tag{12}$$

A global estimate $\hat{\lambda}$ of the normalization factor can be obtained based on the values computed for each separate window. The first few frames in the sequence are typically discarded during this process. This accounts for potential discrepancies between the ISM-simulated RIR $h(\cdot)$ and the predicted power function $\hat{h}_P(\cdot)$ resulting from the presence of strong and discrete early reflections.[2] In this work, an average of the λ_k values was found to achieve good simulation results:

$$\hat{\lambda} = \frac{1}{K_\lambda} \cdot \sum_{k=k_s}^{K} \lambda_k\,, \tag{13}$$

with the start index $k_s = K - K_\lambda + 1$, and with K_λ denoting the total number of frame values used to compute the estimate of the λ parameter ($0 < K_\lambda < K$).

D. Statistical distribution of the diffuse reflections

Together with (13), the predicted power function $\hat{h}_P(\cdot)$ is used to generate the diffuse RIR decay. This procedure involves the choice of a suitable probability density function (PDF) from which to generate the noise signal. Figure 2 displays a histogram of the normalized RIR values $h(\cdot)$ during the late reverberation phase in ISM-simulated RIRs, collected from a total of 100 different RIRs. Each RIR was computed with a randomly-generated source–receiver configuration in a

[1] Note that this particular issue does not directly affect the derivation of the EDC approximation method presented in [10] since the EDC is based on a normalized version of the Schroeder integration method, as shown in (8).

[2] In practice, this also means that it is desirable to define a value of t_c large enough to include the start of the diffuse phase in the ISM simulation of the early RIR part, so as to obtain an accurate estimate of the λ parameter.

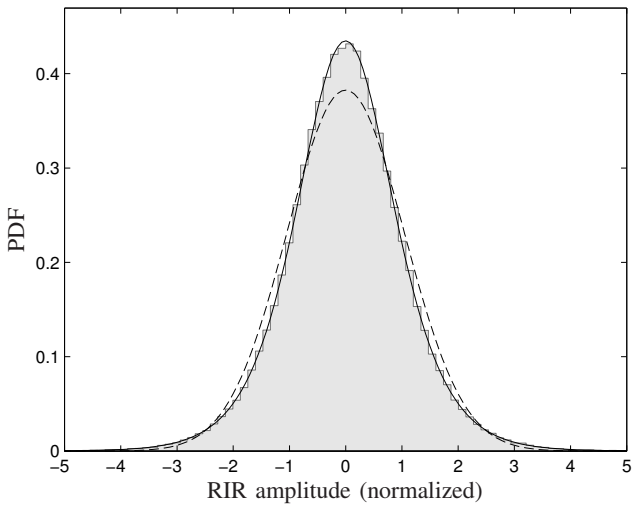

Fig. 2. Amplitude distribution of the late RIR reverberations (normalized) for environments with $T_{60} \approx 0.3\,\text{s}$. This plot shows: *i)* histogram of normalized diffuse RIR values (i.e., $h(t)$ for $t > t_c$) from 100 different RIR simulations, *ii)* Gaussian PDF fit of the histogram data (dashed line), and *iii)* logistic PDF fit (solid line).

virtual enclosure with random volume selected in the range between 20 and $250\,\text{m}^3$. In each environment, nonuniform values of reflection coefficients were also randomly set in such a way as to achieve a reverberation time of $T_{60} \approx 0.3\,\text{s}$.

The histogram data was obtained by first simulating each full-length RIR using the standard ISM algorithm [10]. The early reflections were subsequently cropped out for $t \leqslant t_c$, with $\Delta_c = 20\,\text{dB}$ chosen here so as to ensure that the remaining signal would only contain the diffuse RIR tail. This signal was then normalized in order to compensate for the average decay pattern, leading to a stationary signal with unit variance. This signal data was finally collected by concatenation for each of the 100 RIRs, leading to a total of approximately $5 \cdot 10^5$ data samples. The resulting histogram of amplitude values is plotted as a shaded area in Figure 2.

The Gaussian distribution might appear as a potential candidate to model the diffuse RIR reflections, which can be seen as an isotropic addition of a multitude of wave fronts with large modal overlap impinging on the receiver from many directions [18, 34, 42]. Figure 2 however demonstrates that this PDF does not represent the best fit for typical ISM-simulated late reverberation data.[3] The dashed line in this plot shows a Gaussian PDF with distribution parameters (mean and variance) computed as maximum likelihood estimates from the histogram data. While the Gaussian distribution might provide acceptable perceptual results in practice, a better fit is provided by the logistic PDF $\mathcal{L}(\cdot)$, which exhibits a slightly longer tail and a higher kurtosis compared to the Gaussian PDF. The expression for the logistic PDF (evaluated at ξ) is defined as

$$\mathcal{L}(\xi; \mu, s) = \frac{\exp(-(\xi-\mu)/s)}{s \cdot \big(1 + \exp(-(\xi-\mu)/s)\big)^2}\,, \qquad (14)$$

[3]This particular result is likely to originate from the simplifying assumptions of geometrical room acoustics (including the lack of an acoustic air absorption variable) defined in the frame of the image-source model.

with mean μ, scale parameter s, and standard deviation resulting in $\sigma = \pi \cdot s/\sqrt{3}$. The solid line in Figure 2 shows a logistic PDF fit of the simulation data, here again plotted using maximum likelihood estimates of the parameters μ and s (p-value from the one-sample Kolmogorov-Smirnov test resulting in $p = 0.565$). In addition, a whiteness test on the same input data also shows that the signal samples are statistically uncorrelated in time.[4]

On the basis of these observations, and using the results from Section III-C, the diffuse part of the RIR is thus modeled as decaying random noise as follows:

$$h(t) = n(t) \cdot \sqrt{\frac{\hat{h}_P(t)}{\hat{\lambda}}}\,, \quad t > t_c\,, \qquad (15)$$

where $n(\cdot)$ is a zero-mean random process with unit variance and distributed according to the logistic PDF:

$$n(\cdot) \sim \mathcal{L}(0, \sqrt{3}/\pi)\,. \qquad (16)$$

In practice, random samples from the logistic PDF can be easily generated by using an inversion method based on the known logistic cumulative distribution function (CDF).

E. Summary and further implementation details

Based on the developments presented in this section, the proposed diffuse reverberation modeling (DRM) technique for ISM-based RIR synthesis can be summarized with the following steps.

1) *Early reflections:* compute the early part of the RIR $h(t)$, i.e., for the time interval $t \in [0, t_c]$, using the ISM simulation method described in [10].
2) *Predicted RIR power envelope:*
   a) using the results from Step 1), compute the average of the RIR power $h^2(\cdot)$, i.e., the denominator of (12), for K_λ windows of length T with frame centers t_k, $k = k_s, \ldots, K$,
   b) compute the predicted RIR power envelope $\hat{h}_P(t)$ as given by (3) for $t \geqslant 0$ and up to the desired total RIR length t_{RIR},
   c) based on the results from Steps 2a) and 2b), use the values of $\hat{h}_P(\cdot)$ obtained at the time points t_k to compute an estimate $\hat{\lambda}$ of the normalization factor according to (12) and (13).
3) *Late reverberation:* simulate the diffuse tail of the RIR $h(t)$, for $t \in (t_c, t_{\text{RIR}}]$, according to (15) and (16).
4) *Reconstructed RIR:* concatenate the two RIR sections generated in Steps 1) and 3).

In the rest of this document, this algorithm will be denoted as IS-DRM, emphasizing the fact that the proposed approach represents a DRM approximation of the image-source (IS) simulation technique.

[4]It must be noted that the results presented in this subsection (Figure 2) are similar when using different values of T_{60}, or when the RIRs are randomly selected in one single environment (average not considered over different rooms).

In this algorithm, the transition between the early part and the diffuse part of the RIR is defined as being abrupt, i.e., the final impulse response is reconstructed by simply concatenating the two different sections. This has shown to produce good results in the current work, provided the ISM-simulated part of the RIR includes the beginning of the diffuse reverberation phase. If necessary, an alternative approach based on cross-fading between the two RIR sections [41] may be used in order to produce perceptually better results. This may be required, e.g., if smaller values of the cut-off parameter t_c are to be used or to achieve a smooth transition over a larger early-to-late transition period.

Step 1) above involves carrying out ISM simulations up to the time lag t_c for which the overall RIR energy has decreased by Δ_c dB. In practice, this time lag can be determined "on the fly" as the RIR is being simulated. Alternatively, the EDC prediction method in [10] can also be used here to estimate the value of t_c, thus avoiding the need for a constant monitoring of the EDC during the computations. This latter approach is used in this work since it allows for a faster ISM synthesis of the early reflections.

Finally, evaluating the function $\hat{h}_P(\cdot)$ in Step 2b) can also be made computationally more efficient by taking advantage of the smoothness of this particular function (see [10] and the example further in Figure 4). Instead of computing $\hat{h}_P(t)$ for every time index $t \geqslant 0$, virtually identical results can be obtained by calculating the $\hat{h}_P(\cdot)$ values at larger time intervals, and subsequently interpolating between them. Significant savings in simulation times for Step 2b) can be achieved with this approach.

IV. Discussion

A. Model limitations

Due to the simplifying assumptions made in the development of both the ISM technique and the EDC approximation method, the IS-DRM approach has some limitations in its applicability. In particular, its validity is limited to the domain of small-room acoustics in rectangular enclosures with specular reflections. The proposed method is thus likely to be unsuitable for other types of spaces such as long and narrow ducts, open street canyons, etc.

Furthermore, the considered late-reverberation model explicitly relies on the envelope (average decay) of the RIR power. Satisfactory results can thus only be expected in environments where the RIRs are "well behaved", i.e., display a rather smooth and regular decay of acoustic energy. Some particular RIR decays are thus impossible to model properly with the IS-DRM technique. This is the case, e.g., with environments exhibiting a very uneven distribution of the acoustic absorption among the walls. Figure 3 presents an example of such a problematic RIR, obtained in a $2.8 \times 5.9 \times 2.7\,\mathrm{m}^3$ room with $\boldsymbol{\beta} = [0.72\ 0.58\ 0.78\ 0.84\ 0.69\ 0.55]^{\mathrm{T}}$, $\boldsymbol{p}_s = [0.5\ 0.7\ 0.9]^{\mathrm{T}}$ (m), and $\boldsymbol{p}_r = [1.2\ 3.7\ 1.4]^{\mathrm{T}}$ (m). Empirically, it was found that such irregular decay patterns would typically occur in environments where the relative variability η of the average acoustic absorption of the enclosure's six boundary surfaces

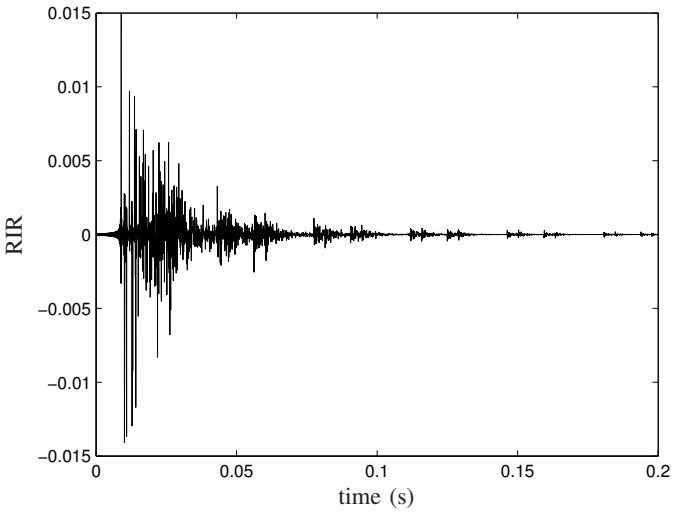

Fig. 3. Example of RIR with irregular decay pattern, computed for a $2.8 \times 5.9 \times 2.7\,\mathrm{m}^3$ room with nonuniform reflection coefficients (the ordinate axis has been cropped for display purposes).

is above the following threshold:

$$\eta = \frac{\sqrt{\mathrm{var}(\boldsymbol{\alpha} \otimes \boldsymbol{S})}}{S} \gtrsim 0.04\,. \tag{17}$$

In this expression, S is the total surface area of the considered enclosure, $\mathrm{var}(\cdot)$ represents the variance, $\otimes$ denotes the element-wise multiplication (Hadamard product), $\boldsymbol{\alpha}$ is the vector of average acoustic absorption coefficients (with $\alpha = 1 - \beta^2$):

$$\boldsymbol{\alpha} = [\alpha_{x,1}\ \alpha_{x,2}\ \alpha_{y,1}\ \cdots\ \alpha_{z,2}]^{\mathrm{T}}, \tag{18}$$

and $\boldsymbol{S}$ is the vector of surface area for each wall:

$$\boldsymbol{S} = [S_{x,1}\ S_{x,2}\ S_{y,1}\ \cdots\ S_{z,2}]^{\mathrm{T}}. \tag{19}$$

The threshold parameter η in (17) effectively corresponds to an Eyring-like estimate of the area-weighted absorption coefficient variance in the considered setting. Environments for which (17) holds true may exhibit a significantly uneven distribution of the sound absorption across the room, and (17) thus provides a basic rule of thumb to assess whether the proposed RIR synthesis technique is applicable to a given environmental setup. The threshold in (17) was further found to be relatively invariant with respect to the value of reverberation time T_{60}.

In environments deemed problematic according to this criterion, the IS-DRM simulation technique will produce a correct *average* decay of the RIR energy. The particular characteristics (decay pattern) of the RIRs in these cases can however only be generated accurately by means of a standard ISM simulation. It must also be noted here that in order to obtain relevant statistics, the histogram data used in Section III-D was obtained by discarding any RIR displaying a significantly irregular energy decay pattern.

B. Model parameters

The accuracy of the proposed method typically depends on a suitable choice of numerical values for the various

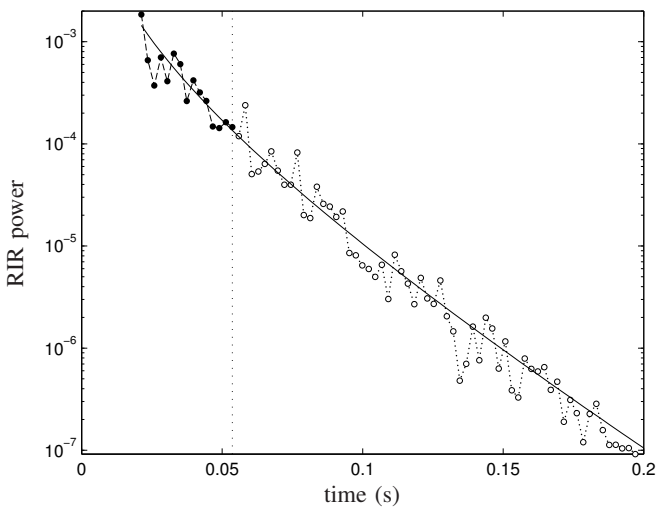

Fig. 4. RIR power measurements and predicted power envelope. *Black dots (dashed line):* measured RIR power during early reflections. *White circles (dotted line):* measured RIR power for late reverberation (for comparison purposes only). *Solid line:* prediction of the RIR power envelope for the considered simulation setup.

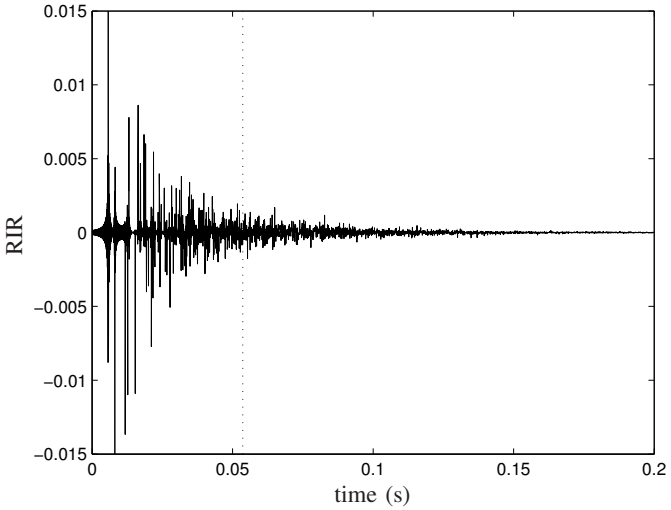

Fig. 5. Example of RIR simulated with the proposed IS-DRM method, using the same setup as in Figure 1 (ordinate axis cropped for display purposes). The vertical dotted line indicates the transition time t_c between early reflections (ISM-based simulation) and late reverberation (random noise model).

model parameters, which in turn depends on the targeted domain of application. The savings in simulation time, for instance, directly result from the chosen value of t_c, which also determines how much of the early reflections is cropped out in the resulting RIR. Real-time applications might thus require a low early-to-late transition time leading to some degradation of the early reflections. On the other hand, applications using offline simulations of a large number of RIRs will achieve good simulation results and a significant computational speed-up with only slightly larger values of the t_c parameter.

Extensive simulations of the IS-DRM method have also revealed that the accuracy of the resulting RIRs is rather insensitive to the specific numerical value selected for the window length T (as also reported in [10]). The same observation also applies to the number of frames K_λ. Section V provides some representative examples of RIRs simulated according to the proposed technique, using a typical setting of these model parameters.

C. Extension of the proposed method

The approach proposed in this work is not strictly limited to an ISM simulation of the early reflections. The early section of the RIR can be synthesized by any available method in practice, and the reverberant tail can then be readily generated as described in Sections III-C and III-D. This effectively allows the use of a different simulation model that might be more relevant in the frame of a given practical implementation, to account for additional effects such as early diffuse reflections, for instance. As mentioned earlier however, one potential advantage of the ISM over other methods is that it is able to find all valid specular reflections for a given environment, which may also be critical for the simulation of the early RIR section in the targeted application.

V. Experimental results

In this section, the performance of the IS-DRM approach is tested using the example of a reconstructed RIR as well as subjective listening tests. An assessment of the average simulation times is also presented to compare its computational requirements with respect to standard image-source simulations as well as other existing ISM-based variants.

A. Reconstructed RIR example

For comparison purposes, the IS-DRM method was applied to the example environment considered earlier in Section III-A, i.e., a $6.6 \times 6.1 \times 2.9\,\mathrm{m}^3$ room with $T_{60} \approx 0.3\,\mathrm{s}$. The experimental results shown in this section were obtained with the following parameter settings. The window length for the instantaneous power estimates was set to $T = 2.31\,\mathrm{ms}$, a total of $K_\lambda = 9$ windows were used when estimating $\hat{\lambda}$, and the early reflections were simulated until the RIR energy had decreased by $\Delta_c = 15\,\mathrm{dB}$, leading to a transition time of $t_c \approx 53.6\,\mathrm{ms}$. Additionally, the sampling frequency for the simulated RIRs was arbitrarily selected as $F_s = 16\,\mathrm{kHz}$.

Intermediate simulation results are shown in Figure 4. This figure displays the measured values of instantaneous acoustic power computed for the early reverberation part of the ISM-simulated RIR (black dots). The solid line represents the quantity $\hat{h}_P(\cdot)/\hat{\lambda}$, i.e., the predicted power envelope adjusted to fit the early reflection decay over K_λ frames. For comparison purposes, Figure 4 also shows the instantaneous power estimates (white circles) measured on the full-length ISM-simulated RIR for the considered environmental setup (shown in Figure 1); it is however emphasized that these values are not required in the process of simulating RIRs according to the IS-DRM model. A visual assessment of these results demonstrates that the expression in (3) provides a good prediction of the RIR's instantaneous acoustic power for the time interval of interest in the decay (late reverberation).

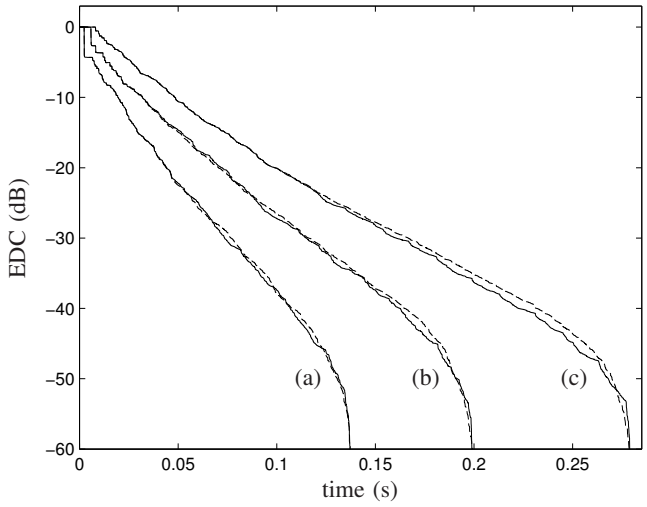

Fig. 6. EDC examples for different environmental setups with: (a) $T_{60} \approx 0.2\,\text{s}$, (b) $T_{60} \approx 0.3\,\text{s}$, (c) $T_{60} \approx 0.4\,\text{s}$. *Solid lines:* EDCs obtained for full-length ISM simulations. *Dashed lines:* EDCs computed from reconstructed RIRs synthesized with the IS-DRM method.

The final reconstructed RIR, simulated according to the proposed model, is displayed in Figure 5. The vertical dotted line in this plot indicates the time t_c where the simulation switches from the ISM computations to the late-reverberation model. This result can be directly compared to the full-length ISM simulation in Figure 1. A visual assessment reveals that both RIRs are very similar in shape, with the RIR in Figure 5 exhibiting a seamless transition from early to late reverberation, which highlights the accuracy of the developments presented in Section III. Furthermore, the computational time required for the simulation of the reconstructed RIR in Figure 5 was approximately 95 times lower compared to that required for the full-length ISM simulation in Figure 1.[5]

Finally, the accuracy of the IS-DRM method can be further investigated by computing the EDC for the reconstructed RIR, as shown in Figure 6 (dashed lines). This plot also contains the EDC obtained for the corresponding full-length ISM-simulated RIRs (solid lines). The energy decay computed for the RIR example considered so far corresponds to the curve labeled (b) in the plot. The curves (a) and (c) demonstrate the same results for different environments with $T_{60} \approx 0.2\,\text{s}$ and $T_{60} \approx 0.4\,\text{s}$, respectively. It is worth noting that these examples all exhibit a certain degree of nonexponential energy decay (i.e., nonlinear decay in a dB scale). The close match between these results again demonstrates the effectiveness of the EDC prediction technique used in this work, which is able to properly model the overall rate of the decay, as well as its intrinsic curvature in the case of RIRs not exhibiting a purely exponential energy decay.

B. Objective evaluation results

The literature contains the definition of many objective acoustic parameters that can be used as comparison criteria in order to assess the similarity between a reconstructed RIR and that obtained from standard ISM simulations. These include parameters such as early decay time *EDT*, clarity C_{50} or C_{80}, center time T_c, etc. Table I provides some typical numerical values computed for the two example RIRs considered in this work. A formal definition of the acoustic parameters used in this table can be found, e.g., in [43]. To put these results in the context of audibility, Table I also includes the just-noticeable difference (JND) thresholds for each objective parameter, as defined in [44].

TABLE I

COMPARISON OF OBJECTIVE ACOUSTIC PARAMETERS (WITH CORRESPONDING JNDS) COMPUTED FOR THE ISM AND IS-DRM RIRS SHOWN IN FIGURES 1 AND 5.

Parameter		ISM	IS-DRM	JND
Clarity	C_{50} (dB)	33.42	34.14	1.0
	C_{80} (dB)	51.02	51.84	1.0
Early decay time	*EDT* (ms)	187.68	185.83	9.3 (5%)
Center time	T_c (ms)	14.27	14.16	10.0
Reverberation time	T_{30} (ms)	239.11	243.88	23.9 (10%)

Most of the objective criteria found in the literature are defined as energy-based measures of the RIR properties. This implies that the similarity of two RIRs is ultimately assessed on the basis of the similarity of their EDCs. In this regard, the accurate approximation of ISM-simulated EDCs constitutes the main purpose of the EDC prediction method used in this work, and its performance has been demonstrated on several occasions in [10] as well as in Figure 6 in the present work. Consequently, these acoustic parameters lead to virtually identical values for both the ISM-simulated and DRM-reconstructed RIRs, as shown in Table I. This result is also to be expected since the early part of the RIR is indeed computed exactly with the IS-DRM technique.

In an attempt to further quantify any potential discrepancies between the proposed DRM approach and the ISM method, an assessment of the reconstructed RIRs based on subjective listening criteria is presented in the following subsections.

C. Subjective evaluation test

A listening test was set up in order to determine whether human listeners were able to detect any audible difference between the audio data generated with the ISM and the IS-DRM method. In addition, the test was designed to investigate which was the preferred method (if any) in cases where the listeners were able to detect a difference.

A total of 12 different scenarios were considered, with room volumes and reverberation times selected randomly in the range $V \in [30, 100]$ (m^3) and $T_{60} \in [0.2, 0.6]$ (s), respectively. One random source–sensor configuration (static positions) was chosen in each of these environments. For both the ISM and IS-DRM methods, the corresponding RIR was computed and a sample of reverberant audio data was obtained by convolution with an anechoic speech sample in Swedish. The resulting stimuli were then presented in random order to the listeners, who were asked to choose one of the following three answers at each trial:

[5]For both RIRs, the total impulse response length t_{RIR} here corresponds to the time for which the RIR energy has decreased by approximately 45 dB.

A1: first audio sample sounds more natural than second,
A2: second audio sample sounds more natural than first,
A3: I cannot hear any difference (samples are identical).

For the purpose of cross-validation, some of the 12 cases were selected to have identical stimuli instead of the two ISM and IS-DRM samples. In other words, both sentences were generated using the exact same RIR in these cases. Eight Swedish-speaking participants of varying age (20 to 60 years old), four male and four female, were asked to listen to the 12 sets of sentences using high-quality closed-cup headphones. The listening test was set up as a double-blind experiment.

The results and conclusions from this subjective test are as follows:

i) 62.5% of subjects were not able to hear any difference between the two sentences (i.e., replied with A3 in at least 11 of the 12 cases), indicating that the ISM and IS-DRM audio samples are identical to a majority of listeners,

ii) among the 37.5% of subjects who said they could hear differences, the average error rate (i.e., A3 selected when different RIRs were used, or A1/A2 selected when the two sentences were identical) was 47.2%, which is close to random selection,

iii) for the 37.5% of participants who said they could hear differences, when correctly answering that the two audio samples were different, they favored the IS-DRM method (i.e., found it more natural) in 80% of the cases on average, which suggests that listeners have no particular aversion to the DRM-based audio data.[6]

These results indicate that subjectively, the use of the proposed IS-DRM approach does not lead to any specific artifact in the corresponding audio data. This would further suggest that the human ear is not particularly sensitive to the exact (fine-scale) structure of a diffuse RIR tail, as long as the energy content remains the same. This outcome is also in line with the results from Section V-B (Table I) which indicate that the difference between objective parameter values for both methods are well within the limits of subjective perception.

Finally, another interesting observation that can be made from this test is that the 62.5% of participants who were not able to hear any difference between the sentences were all aged 30 and over, while the other 37.5% were all below 30. It is however unclear whether this effect is due to psychological factors rather than the physiological hearing ability of the listeners.

D. Subjective evaluation: moving sound source[7]

Further subjective tests were carried out according to the following setup. Two microphones were defined 20 cm from each other (simulating a crude stereo auralization system) in a virtual room with dimensions $4.0 \times 5.0 \times 2.7\,\mathrm{m}^3$. The vector of reflection coefficients was defined as $\boldsymbol{\beta} = [0.27\ 0.22\ 0.25\ 0.19\ 0.24\ 0.22]^{\mathrm{T}}$, leading to a reverberation time $T_{60} \approx 0.55$ s. A sound source was then simulated moving along a 2.64 m trajectory forming part of a circle around the two acoustic sensors.

The trajectory was first subdivided into 106 discrete points 2.51 cm apart, from which the RIRs were computed to each of the two microphones using the standard ISM technique [10] as well as IS-DRM simulations. The audio data at the sensors was obtained by convolution (overlap-add) of the computed RIRs with an 8.9 s sample of anechoic speech signal. Due to the relatively large distance increment between trajectory points, the audio signals for both methods contain a distinct clicking noise that can be heard throughout the simulation. This highlights a potentially significant issue that arises when using too few RIRs during the ISM-based simulation of audio data in a moving-source scenario.

In a second experiment, the same trajectory was discretized using a finer grid of 528 points, corresponding to a distance increment of 0.5 cm. The 1056 RIRs were then computed using the fast IS-DRM technique. Due to the finer spacing between the trajectory points, no audible artifact can be perceived in the computed microphone signals any longer. Furthermore, the resulting audio data does not appear to suffer from any unexpected issues related to spatialization or other subjective acoustic factors, and does not present any evidence of a potential spectral sound coloration effect either.[8] In the computation of these experimental results, the 1056 RIRs along the finely discretized trajectory were simulated with the proposed DRM approach in about 21 minutes on a standard desktop computer. On the other hand, a total of about nine hours was necessary for the computation of 212 RIRs (coarse trajectory) by means of the standard ISM algorithm.

E. Analysis of computational requirements

The computational load required by an ISM implementation results from several factors, including mainly the considered enclosure dimensions L_x, L_y and L_z, the sampling frequency F_s, the reverberation time T_{60}, and the desired total RIR length t_{RIR}. In practice, the combination of such factors can lead to considerable simulation times, especially for certain types of application requiring the computation of many distinct RIRs, as demonstrated in the previous subsection for instance.

Figure 7 demonstrates the significant savings in computational requirements achieved with the proposed RIR synthesis method compared to a standard ISM implementation. To this purpose, the following statistical average was carried out. For a given value of T_{60}, the CPU time (central processing unit) required for the computation of a RIR was recorded for both simulation methods in a total of 100 different environmental setups. These involved different source–receiver configurations, various nonuniform values of reflection coefficients, and random room volumes ranging from 20 to $250\,\mathrm{m}^3$. For both methods, the RIRs were simulated for a total length corresponding to the time for which the EDC has decreased by approximately 50 dB. With the IS-DRM technique, the early-to-late reverberation threshold was defined as $\Delta_c = 20$ dB. The CPU times were obtained on the basis of a MATLAB

[6]This result should however be treated with a certain degree of caution given the high error rate in the answers given by this group of respondents.

[7]The samples of audio data for the experiments described in this subsection are available online for download from [45].

[8]Spectral coloration issues were noted in [31] as a result of using dissimilar methods for the synthesis of the different sections of the impulse response.

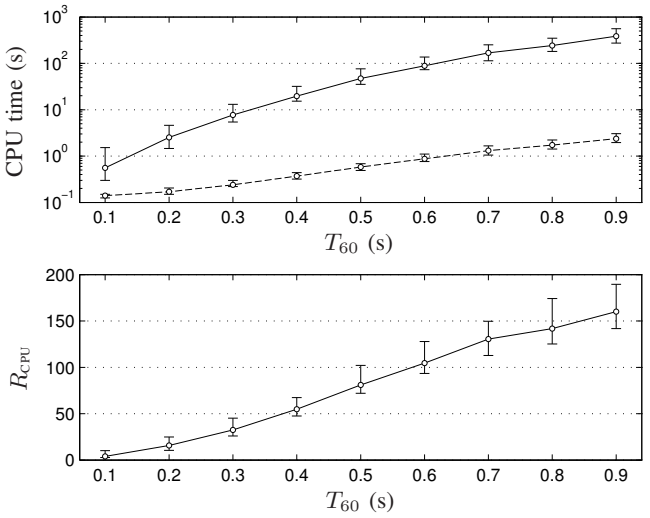

Fig. 7. Assessment of computational load for the ISM and DRM methods (median and IQR plots). *Top plot:* CPU times required for the simulation of ISM RIRs (solid line) and IS-DRM RIRs (dashed line). *Bottom plot:* execution time speed-up factor R_{CPU} between the ISM and the proposed method.

implementation executed on a standard $2.4\,$GHz desktop computer running under Linux, with the program code for these implementations specifically optimized for execution speed.

The distribution results for both methods are displayed in the top graph of Figure 7, in the form of median and interquartile range (IQR) plots. As expected, the ISM algorithm spends a significant computational effort on the simulation of the higher-order reflections. Modeling this diffuse RIR phase according to the IS-DRM technique thus allows an important reduction of the computational costs. Also, the computation times for full-length ISM simulations would rise even further if the total RIR length t_{RIR} is increased, whereas the computational load for the proposed method remains virtually unaffected by this parameter.

The savings in simulation times are further quantified in the bottom plot of Figure 7, which displays the ratio R_{CPU} of the CPU time required by the standard ISM implementation to that of the IS-DRM technique. This speed-up factor was computed for each individual RIR in the aforementioned simulations, and the resulting distributions are again plotted using a median and IQR representation. These results show that a reduction of the computation times by up to two orders of magnitude can be achieved with the proposed simulation method, in the frame of the considered experimental setup. These results also clearly illustrate how the IS-DRM approach becomes increasingly advantageous as the simulation times of the standard ISM implementation become increasingly prohibitive.

F. Comparison with existing ISM-based methods

As mentioned in Section I, several other methods have been presented in the literature with the aim to reduce the computational costs of the ISM algorithm. The authors of these works however do not always present explicit speed-up results, or present such results with respect to other factors such as the number of image sources or the length of the computed RIR, regardless of the reverberation time T_{60}. Table II summarizes the results reported in the literature for some of these existing ISM-based algorithms. These results are presented here solely as an indication of how the IS-DRM compares in terms of computational complexity. It is important to note that the other ISM-based methods considered here compute exact image-source results, whereas the IS-DRM algorithm provides a partial approximation thereof. Therefore, the results in Table II should not necessarily be seen as a measure that can be used to quantify the superiority of one method over another.

TABLE II

SPEED-UP FACTORS ACHIEVED BY EXISTING ISM ALGORITHMS. FOR IS-DRM, VALUES FOR R_{CPU} AND THE NUMBER OF SOURCES CORRESPOND TO AVERAGE RESULTS OVER 100 DIFFERENT ENVIRONMENTS.

Method	Number of image sources	T_{60} (s)	R_{CPU}
IS-DRM	$3.2 \cdot 10^5$	0.9	160
	$1.1 \cdot 10^5$	0.6	104
	$5.5 \cdot 10^4$	0.5	81
	$6.3 \cdot 10^3$	0.2	16
Ref. [33]	$1.1 \cdot 10^6$	–	< 10
Ref. [28]	$1.0 \cdot 10^4$	–	8.6
Ref. [27]	$7.2 \cdot 10^4$	–	17

The method presented in [33] is shown to achieve a reduction of the computation times by less than one order of magnitude for $1.1 \cdot 10^6$ image sources (fifth reflection order), although this result is obtained for a complex polyhedral environment with many invalid ray paths. The authors in [28] propose the use of look-up tables and sorting methods to prevent redundant ISM calculations. They provide the example of a RIR simulation with about 10^4 image sources, achieving a speed-up factor of $R_{\text{CPU}} \approx 8.6$. Finally, the work in [27] describes a multipole expansion technique capable of reducing the computational requirements by up to 17 times when simulating a total of about $7.2 \cdot 10^4$ image sources. This approach however involves a highly demanding pre-calculation phase (up to 38 hours of pre-computation time on a modern desktop computer, for any given environment).

For comparison, some of the simulation times presented in Section V-E (Figure 7) for the IS-DRM method are also reproduced in Table II. For this algorithm, the values of R_{CPU} and the number of image sources correspond to an average (median) obtained over the 100 different environments considered in Section V-E for the given values of T_{60}. These results show that the IS-DRM method typically achieves higher speed-up factors (when considering a similar number of image sources) for the computation of RIRs that have been shown to closely approximate image-source results.

VI. CONCLUSION

Using a recently proposed model of the diffuse decay in a virtual reverberant environment, the method presented in this work provides a tool allowing acoustics researchers and engineers to synthesize RIRs in an efficient manner, as computed on the basis of the image-source model. The resulting impulse responses were shown to truthfully represent the considered acoustic environment, statistically and in terms of energy

decay, while only requiring a fraction of the computation times compared to a standard ISM simulation. To the same extent as the original ISM algorithm itself, the proposed DRM method cannot necessarily be seen as a substitute for real-world acoustic simulations, e.g., when used as a validation technique for audio processing algorithms operating in reverberant conditions. However, it is able to provide valuable preliminary assessment results and significant computational advantages in several domains of application such as real-time audio processing algorithms and offline simulations involving a large number of RIRs.[9]

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for the valuable comments and suggestions provided during the preparation of this manuscript.

REFERENCES

[1] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, July 1968.

[2] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," *Journal of the Acoustical Society of America*, vol. 115, no. 2, pp. 739–756, February 2004.

[3] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 812–824, May 2008.

[4] A. Craggs, "Acoustic modeling: finite element method," in *Handbook of Acoustics*, M. J. Crocker, Ed. New York: Wiley, 1998, pp. 149–156.

[5] A. F. Seybert and T. W. Wu, "Acoustic modeling: boundary element methods," in *Handbook of Acoustics*, M. J. Crocker, Ed. New York: Wiley, 1998, pp. 157–167.

[6] D. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley, "Acoustic modeling using the digital waveguide mesh," *IEEE Signal Processing Magazine*, vol. 24, no. 2, pp. 55–66, March 2007.

[7] S. A. Van Duyne and J. O. Smith, "Physical modeling with the 2-D digital waveguide mesh," in *Proceedings of the International Computer Music Conference*, Tokyo, Japan, 1993, pp. 40–47.

[8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, April 1979.

[9] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527–1529, November 1986.

[10] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, July 2008.

[11] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja, "The room acoustic rendering equation," *Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1624–1635, September 2007.

[12] C. Lauterbach, A. Chandak, and D. Manocha, "Interactive sound rendering in complex and dynamic scenes using frustum tracing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1672–1679, November/December 2007.

[13] B. Kapralos, M. Jenkin, and E. Milios, "Sonel mapping: a stochastic acoustical modeling system," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, Toulouse, France, May 2006, pp. 421–424.

[14] C. T. Wolfe and S. K. Semwal, "Acoustic modeling of reverberation using smoothed particle hydrodynamics," in *Proceedings of the International Conference on Computer Graphics, Visualization and Computer Vision*, Plzen-Bory, Czech Republic, February 2008, pp. 191–198.

[9]To this purpose, the MATLAB code for the proposed IS-DRM implementation is made freely available for download from [46, 47].

[15] M. A. Akeroyd, S. Gatehouse, and J. Blaschke, "The detection of differences in the cues to distance by elderly hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1077–1089, February 2007.

[16] S. F. Poissant, N. A. Whitmal III, and R. L. Freyman, "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," *Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1606–1615, March 2006.

[17] J. Dmochowski, J. Benesty, and S. Affes, "An information-theoretic view of array processing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 2, pp. 392–401, February 2009.

[18] E. A. P. Habets, S. Gannot, I. Cohen, and P. C. W. Sommen, "Joint dereverberation and residual echo suppression of speech signals in noisy environments," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1433–1451, November 2008.

[19] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, October 2006.

[20] M. Kuster, "Reliability of estimating the room volume from a single room impulse response," *Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 982–993, August 2008.

[21] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, April 2005.

[22] N. H. Adams and G. H. Wakefield, "State-space synthesis of virtual auditory space," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 881–890, July 2008.

[23] T. Lokki, L. Savioja, R. Väänänen, J. Huopaniemi, and T. Takala, "Creating interactive virtual auditory environments," *IEEE Computer Graphics and Applications*, vol. 22, no. 4, pp. 49–57, July/August 2002.

[24] D. Necsulescu, W. Zhang, W. Weiss, and J. Sasiadek, "Room acoustics measurement system design using simulation and experimental studies," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 1, pp. 167–172, January 2009.

[25] D. Li and M. Hodgson, "Optimal active noise control in large rooms using a "locally global" control strategy," *Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3653–3661, December 2005.

[26] J. Rindel, "The use of computer modeling in room acoustics," *Journal of Vibroengineering*, vol. 3, no. 4, pp. 219–224, 2000.

[27] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Fast evaluation of the room transfer function using multipole expansion," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 565–576, February 2007.

[28] S. G. McGovern, "Fast image method for impulse response calculations of box-shaped rooms," *Applied Acoustics*, vol. 70, no. 1, pp. 182–189, January 2009.

[29] E. A. Lehmann and A. M. Johansson, "Particle filter with integrated voice activity detection for acoustic source tracking," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007, Article ID 50870, 11 pages.

[30] H. Hacıhabiboğlu and F. Murtagh, "Perceptual simplification for model-based binaural room auralisation," *Applied Acoustics*, vol. 69, no. 8, pp. 715–727, August 2008.

[31] R. Heinz, "Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail," *Applied Acoustics*, vol. 38, no. 2–4, pp. 145–159, 1993.

[32] U. R. Kristiansen, A. Krokstad, and T. Follestad, "Extending the image method to higher-order reflections," *Applied Acoustics*, vol. 38, no. 2–4, pp. 195–206, 1993.

[33] H. Lee and B.-H. Lee, "An efficient algorithm for the image model technique," *Applied Acoustics*, vol. 24, no. 2, pp. 87–115, 1988.

[34] D. R. Begault, *3-D sound for virtual reality and multimedia*. Boston: Academic Press, 1994.

[35] L. Dahl and J.-M. Jot, "A reverberator based on absorbent all-pass filters," in *Proceedings of the COST-G6 Conference on Digital Audio Effects*, Verona, Italy, December 2000, pp. 67–72.

[36] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, San Francisco, CA, USA, March 1992, pp. 221–224.

[37] M. Kompis and N. Dillier, "Simulating transfer functions in a reverberant room including source directivity and head-shadow effects," *Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2779–2787, May 1993.

[38] J. Borish, "Extension of the image model to arbitrary polyhedra," *Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, June 1984.
[39] J. Garas, *Adaptive 3D sound systems*. Norwell, Mass.: Kluwer Academic, 2000.
[40] B. Blesser, "An interdisciplinary synthesis of reverberation viewpoints," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 867–903, October 2001.
[41] K. Meesawat and D. Hammershøi, "An investigation on the transition from early reflections to a reverberation tail in a BRIR," in *Proceedings of the International Conference on Auditory Display*, Kyoto, Japan, July 2002.
[42] M. R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *Journal of the Audio Engineering Society*, vol. 35, no. 5, pp. 299–306, May 1987.
[43] N. Toma, M. D. Ţopa, V. Popescu, and E. Szopos, "Comparative performance analysis of artificial reverberation algorithms," in *Proceedings of the IEEE International Conference on Automation, Quality and Testing, Robotics*, vol. 1, Cluj-Napoca, Romania, May 2006, pp. 138–142.
[44] N. Prodi and S. Velecka, "The evaluation of binaural playback systems for virtual sound fields," *Applied Acoustics*, vol. 64, no. 2, pp. 147–161, February 2003.
[45] E. A. Lehmann, "Fast ISM simulation: audio data example," online: http://www.eric-lehmann.com/fast_ISM_audio, 2009.
[46] E. A. Lehmann, "Fast image-source method: Matlab code," online: http://www.eric-lehmann.com/fast_ISM_code, 2009.
[47] The MathWorks, "File exchange page for E. A. Lehmann," online: http://www.mathworks.com/matlabcentral/fileexchange/authors/31461, 2009.