UNIVERSITY *of York*

This is a repository copy of *Boundary Absorption Approximation in the Spatial High-Frequency Extrapolation Method for Parametric Room Impulse Response Synthesis*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/142520/

Version: Published Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Boundary absorption approximation in the spatial high-frequency extrapolation method for parametric room impulse response synthesis

Alex Southern, Damian T. Murphy, and Lauri Savioja

Boundary absorption approximation in the spatial high-frequency extrapolation method for parametric room impulse response synthesis

Alex Southern,[1,a)] Damian T. Murphy,[1,b)] and Lauri Savioja[2]

[1]AudioLab: Department of Electronic Engineering, University of York, York, YO10 5DD, United Kingdom
[2]Department of Computer Science, Aalto University, Espoo, 02150, Finland

The spatial high-frequency extrapolation method extrapolates low-frequency band-limited spatial room impulse responses (SRIRs) to higher frequencies based on a frame-by-frame time/frequency analysis that determines directional reflected components within the SRIR. Such extrapolation can be used to extend finite-difference time domain (FDTD) wave propagation simulations, limited to only relatively low frequencies, to the full audio band. For this bandwidth extrapolation, a boundary absorption weighting function is proposed based on a parametric approximation of the energy decay relief of the SRIR used as the input to the algorithm. Results using examples of both measured and FDTD simulated impulse responses demonstrate that this approach can be applied successfully to a range of acoustic spaces. Objective measures show a close approximation to reverberation time and acceptable early decay time values. Results are verified through accompanying auralizations that demonstrate the plausibility of this approach when compared to the original reference case.

I. INTRODUCTION

The spatial high-frequency extrapolation method (SHEM) uses time/frequency analysis of a band-limited spatial room impulse response (SRIR) to generate time-varying metadata defining the directionality/diffusivity of the modeled soundfield.[1] This metadata is then used to synthesize temporally weighted, directionally rendered, high-frequency energy to produce a new full-band room impulse response (RIR) suitable for auralization. The three-dimensional (3-D) finite-difference time domain (FDTD) method has often been applied to room acoustics simulation and auralization problems,[2,3] usually resulting in the synthesis of band-limited RIRs due to the high computational requirements for full audio bandwidth simulations.[4] The computational performance of FDTD methods has been improved through various parallel graphics processing unit (GPU) implementations,[5–7] extending to other architectures,[8] with the aim being full audio bandwidth simulations computed within what might be considered a reasonable time limit (e.g., minutes rather than hours). However, due to the memory required to host a FDTD grid capable of ensuring accuracy at high frequencies (multiple grid points per wavelength) and minimization of the further band-limiting effects of dispersion error,[9–11] full audio bandwidth FDTD simulations in close to real time are generally still not possible. By way of example, one of the acoustic spaces used as a simulation target in previous work,[1] and also later in this paper, is York Minster, one of the largest churches in Europe with a volume of approximately

150 000 m^3. Based on data from some of this prior work,[6,7] a FDTD simulation of this space would require in the region 1 Tb of GPU memory for a grid update rate of 44.1 kHz, and still only give a usable bandwidth of 7 kHz.

Hybrid acoustic models that combine simulation methods over defined time and/or frequency regions are an alternative/addition to highly parallel implementations for reducing computational load and memory requirements in FDTD and wave-based modeling more generally. A comprehensive overview of both geometric and wave-based acoustic simulation methods, together with hybrid models based on a combination of these approaches that have arisen as a consequence, has been presented previously.[4] The application of SHEM to band-limited SRIRs arising from 3-D FDTD room acoustic simulations therefore provides an alternative to geometric acoustic/wave-modeling combinations, based as it is on a parametric/signal-modeling approach. However, SHEM gives best results for large volume spaces with long reverberation times, where air absorption at higher frequencies dominates over boundary absorption losses. In mid-sized spaces, such as concert halls, boundary absorption effects are much more critical and in these examples SHEM gives less accurate results, revealing the limitations of the currently implemented boundary absorption model.[1]

This paper refines SHEM through the introduction of a new, time-varying frequency dependent boundary absorption weighting function based on an energy decay relief (EDR) extrapolation of the existing low-frequency SRIR information. This new method is tested across a range of measured data from spaces of different size and varying acoustic character, and offers an improvement over the previously

a)Also at: AECOM, Glasgow, UK.
b)Electronic mail: damian.murphy@york.ac.uk

adopted boundary approximation function. On this basis SHEM is then applied and tested in the context of hybrid impulse response synthesis for room acoustics simulation. The remainder of this paper is arranged as follows: additional background context for SHEM is presented in Sec. II; in Sec. III the basis for SHEM is reviewed including a detailed consideration of the analysis and synthesis stages, together with the approach used to determine the new boundary absorption weighting function; Sec. IV presents case studies that outline how the SHEM algorithm works, testing it against a number of real-world room acoustic measurements before applying it as part of a hybrid room acoustics simulation; finally, Sec. V summarizes what has been observed and considers further directions for this research.

II. PARAMETRIC MODELS OF REVERBERATION

The use of SHEM to synthesize or model a RIR is similar to methods used in digital reverberation design for virtual acoustics where convolution (based on RIR measurements) and parametric approaches (based on feedback delay networks or similar) are combined.[14–16] Reverberation for spatial audio applications can also be considered from an object-based audio perspective, where a scene is composed of audio content and associated metadata that can be interpreted in different ways by a renderer at playback according to need, such as number of loudspeakers available. RIR convolution-based reverb has been categorised as a signal-based approach[15] with limitations for adaption to object-based representations. SHEM therefore develops a purely computational acoustic room rendering method in this paper based around FDTD modeling toward a flexible semi-parametric implementation suitable for more diverse, object-based applications.

With SHEM it is assumed that there is sufficient information in the existing 3-D FDTD low-frequency SRIR, together with additional acoustic information about the space being modeled, to enable the remaining high-frequency part to be synthesized, resulting in a complete, full bandwidth RIR ready for audio convolution (or rendering). Each time/frequency analysis/synthesis frame is determined as being either directional (containing a distinct reflection) or diffuse (containing reverberant energy) based on this low-frequency input to the SHEM-analysis stage, and so acts similarly to the use of mixing time in parametric reverb algorithms[17] in that strong reflections are treated separately to stochastic reverberation. In previous work[1] the high-frequency energy added at the SHEM-synthesis stage is shaped in time and frequency to approximate boundary absorption losses using a simple, non-time-varying weighting function with a high-frequency roll-off characteristic. Losses due to air absorption over time and frequency are also accounted for, in this case, using an analytical model.[18] SHEM is therefore similar in concept to Jot's EDR-shaped white noise model of reverberation[19] where it is possible to determine a perceptually relevant model of late reverberant decay based on an estimate of the initial spectrum, the reverberation time, and the EDR of a target RIR. These parameters are used to define a time-

varying filter that can synthesize this late reverberant decay based on a Gaussian white noise input.

The introduction and shaping of high-frequency content based on existing low-frequency information means that SHEM also shares its background with audio bandwidth extension (BWE) methods. For instance, spectral band replication (SBR)[20–24] is a form of BWE used in a number of audio codecs and applications to extrapolate high-frequency content of a band-limited signal based on the spectral envelope and related metadata sourced from the original full bandwidth signal, or predicted from the known bandwidth in the absence of full spectrum information. With SHEM, prior information is also assumed to be limited with the low-frequency SRIR being the main source of information for analysis. Given that SHEM is proposed as a method for extending the results obtained from a FDTD acoustic model, it can also be assumed that boundary absorption parameters (e.g., absorption coefficients) are known, leading to an estimation of frequency dependent reverberation times for the given space.

III. THE SHEM

A. Overview

The SHEM is divided into SHEM-analysis and SHEM-synthesis stages, as presented in the block diagram overview shown in Fig. 1. The input to the analysis stage is a SRIR, that is, a RIR captured over multiple channels suitable for
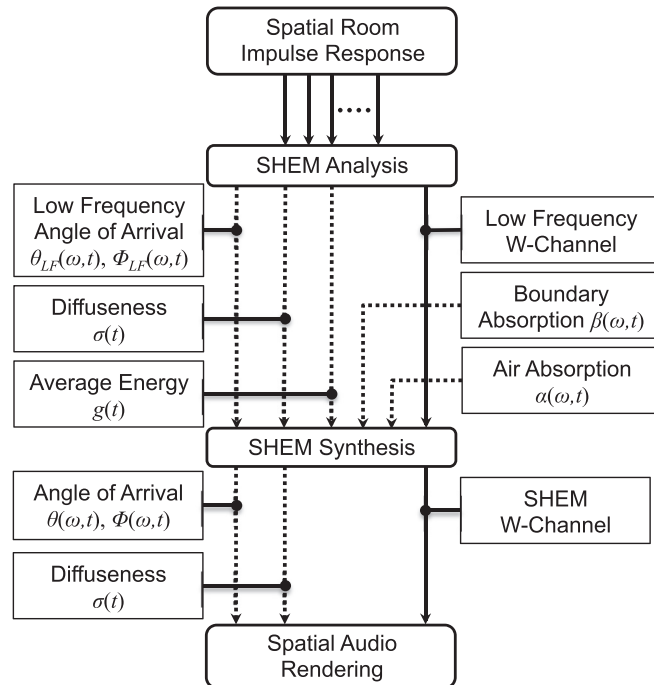
FIG. 1. Block diagram overview of SHEM where dashed and solid lines represent metadata and audio, respectively. The analysis phase estimates directionality, diffuseness, and average energy based on a time-frequency representation of the low-passed, multiple channel SRIR. This metadata, together with an approximation of boundary and air absorption for the space, is passed to the synthesis stage that reconstructs the high-frequency part of the impulse response. The final, post-SHEM-synthesis signal and metadata is then passed to an appropriate spatial audio rendering method.

J. Acoust. Soc. Am. **145** (4), April 2019

Southern *et al.* 2771

obtaining data relating to spatial content such as angle of arrival for direct and reflected sound components.

In the SHEM-analysis stage, the existing low-frequency part of the SRIR is used to generate four time dependent metadata parameters, *low-frequency angle of arrival*, *low-frequency average angle of arrival*, *diffuseness*, and *average energy*. It is assumed that this metadata should be obtainable from any appropriately captured band-limited SRIR.

In the SHEM-synthesis stage the aim is to retain the existing low frequencies of the RIR and introduce temporally weighted, frequency dependent, high-frequency energy to produce a new full audio bandwidth RIR. Corresponding high-frequency metadata is also generated for directional components to enable them to be correctly rendered into the final auralization. The high-frequency energy weighting is determined based on the SHEM-analysis metadata together with losses introduced due to air absorption from an analytical model and boundary absorption losses in this paper determined using an EDR based on an analysis of the available low-frequency part of the RIR and estimated frequency dependent reverberation times. The final, full bandwidth RIR, together with associated metadata, can then be passed to an appropriate spatial audio rendering scheme capable of deriving the required speaker feed signals suitable for, e.g., mono, stereo, or multi-speaker surround-sound playback.

Although SHEM is conceived without recourse to any specific SRIR or spatial audio rendering format, the method used previously and that which is expanded on here is based on SRIRs captured using a coincident B-format microphone.[1] A B-format signal represents the soundfield as captured by a coincident array of three pressure gradient microphones orientated along the Cartesian *xyz* axes, known as X,Y,Z channels, together with the W-channel pressure signal. B-format SRIRs are readily available online from actual room measurements[25] or can be derived from SRIRs obtained from numerical room acoustics simulations.[26] In this work, the analysis and rendering are based on spatial impulse response rendering (SIRR)[27] and the related method, directional audio coding (DirAC),[28] as both utilize B-format SRIRs. Alternative soundfield analysis/rendering methods based on a B-format (or similarly derived) source signal also exist,[29,30] and could be applied as an alternative.

B. SIRR

SIRR is a technique for reproducing a measured soundfield over an arbitrary number and positioning of loudspeakers based on a short-time Fourier transform (STFT) analysis of a B-format RIR.[27] For each time window of the segmented SRIR, the direction of arrival and diffuseness are analyzed across frequency bins. The result of this analysis gives metadata defining diffuseness and directionality (in terms of angles of azimuth and elevation) for each time-frequency component. The directional information is derived from the instantaneous sound intensity, defined as the product of a particles' instantaneous acoustic pressure at a point and its associated velocity through that point in a given direction.[31] By measuring the sound intensity in the

directions of the Cartesian coordinate system after,[27] the average direction of arrival can be estimated,

$$\theta(\omega) = \tan^{-1}\left[\frac{-I_y(\omega)}{-I_x(\omega)}\right], \tag{1}$$

$$\phi(\omega) = \tan^{-1}\left[\frac{-I_z(\omega)}{\sqrt{I_x^2(\omega) + I_y^2(\omega)}}\right], \tag{2}$$

where $\theta(\omega)$ and $\phi(\omega)$ denote the azimuth and elevation of the arrival direction as a function of radial frequency and $I_x(\omega)$, $I_y(\omega)$, and $I_z(\omega)$ are the measured sound intensity vectors in the frequency domain. In practice these are obtained from the Fourier transform of the B-format signals $W(t)$, $X(t)$, $Y(t)$, and $Z(t)$. The diffuseness estimate $\psi(\omega)$ provides an indication of when the arrival directions in Eqs. (1) and (2) can be considered as directional or diffuse energy and, again, after[27] is defined as the ratio of the sound intensity with the energy density

$$\psi(\omega) = 1 - \frac{\sqrt{2}\|\Re\{W^*(\omega)\mathbf{X}'(\omega)\}\|}{|W(\omega)|^2 + |\mathbf{X}'(\omega)|^2/2}. \tag{3}$$

For $\psi(\omega) = 0$, the given time-frequency component is ideally non-diffuse (directional) and for $\psi(\omega) = 1$ the given time-frequency component is considered ideally diffuse. In this work SIRR essentially provides the framework for the SHEM analysis stage, although SHEM uses an alternative means for deriving diffuseness, based on spherical variance as outlined in the following.

C. SHEM analysis

The SHEM-analysis stage is summarized in more detail in Fig. 2 and has two main parts, SIRR-based estimation of directional/diffuse energy and average energy estimation. The input signals are first processed in the time frame windowing block using the STFT. The fast transient nature of a SRIR dictates that the reflection detection must be carried out using high resolution analysis frames in the time domain. However, short-time windows result in a low number of sound intensity vectors in the frequency domain, which will

FIG. 2. A block diagram representation of the SHEM-analysis stages, starting with a B-format SRIR from which is extracted the required low-frequency information from the W-channel, together with the metadata needed for the synthesis stage. The dashed and solid lines represent metadata and audio, respectively.
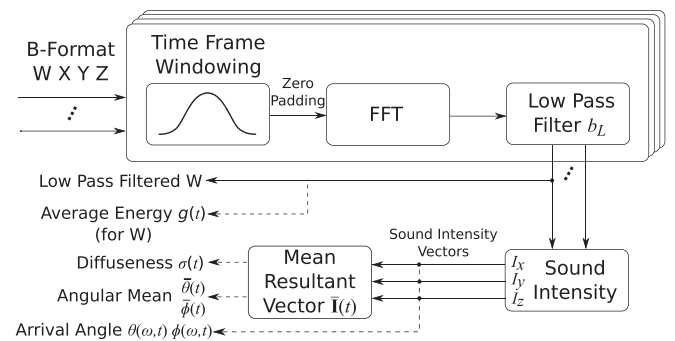
affect the length of the resultant vector used to quantify the diffuseness of the frame. Hence for first part estimation of directional/diffuse energy, the metadata frame and its associated fast Fourier transform (FFT) size are set to 512 samples, 75% overlap Hann windows. The larger time window ensures that more vectors are available in each frame to estimate diffuseness with the hop size ensuring that the directional vectors will change direction only gradually over time. The second part average energy estimation and its associated FFT size are 32 sample, 50% overlap Hann windows. This frame size will also determine how often a directional reflection can be added at the SHEM-synthesis stage. Note that the first part directional/diffuse energy estimation is interpolated to match the second part hop size of 16 samples. Each frame is then convolved with a finite impulse response (FIR) low-pass filter, b_L, to remove any aliasing or higher frequency artefacts (e.g., dispersion error if SRIRs have been sourced from a FDTD method) in the case of using a modeled SRIR, or to provide a low-frequency reference in the case of using a measured SRIR.

An appropriate value for the average energy, $g(t)$, in each frame is taken for each input channel and, in this case, the mean energy over the frequency range considered, summed across both current and past analysis frames is used. Consider a single time-domain reflection, with its total energy spread over a number of sample points (that is, not a single-sample impulse), which is then windowed over two successive analysis frames 16 samples apart (using SHEM-analysis stage, 32 sample, 50% overlapped Hann windows). The peak value in each analysis frame will typically be less than that obtained for the reflection centred and captured within one frame only. Therefore, basing $g(t)$ on only one analysis frame leads to the potential for an underestimation of this value and a locally maximal reflection not being identified as part of the SHEM-synthesis stage.

In the next stage of Fig. 2 the sound intensity vectors are used to calculate the direction of arrival metadata using Eqs. (1) and (2). Ideally, if one or more coherent reflections arrive at the B-format microphone/receiver array, the resulting sound intensity vectors across the frequency bins of a single analysis frame will all point in the same direction. As the soundfield, and hence the SRIR, become more diffuse with the arrival of high order reflections, the directions of these vectors will be distributed more randomly. This is a property that is exploited in SIRR through the diffuseness estimate (3), although SHEM makes use of a similar, alternative quantity, *spherical variance*, calculated using the magnitude of the mean resultant vector, and that has been shown to give improved diffuseness estimation results.[32]

It is assumed that the magnitude of each sound intensity vector in each frequency bin is unity although the associated azimuth and elevation angles are maintained as

$$\mathbf{I_i} = \begin{bmatrix} \cos\theta_i \cos\phi_i & \sin\theta_i \cos\phi_i & \cos\phi_i \end{bmatrix}^T, \qquad (4)$$

where $\mathbf{I_i}$ is the unit vector for θ_i and ϕ_i for the ith frequency bin. In the ideal case, for a coherent reflection, the sound intensity vector for each bin will be aligned to the same angle of arrival, but this may not always be the case, perhaps

due to the limits of accuracy in calculating Eq. (4), diffusely reflecting high-frequency components or directional analysis errors occurring from the physical limitations of the B-format microphone or probe. Hence, the mean resultant vector $\bar{\mathbf{I}}$, the average of all multiple sound intensity vector angles for a given analysis frame, as represented in Fig. 2, is calculated from

$$\bar{\mathbf{I}}(t) = \frac{1}{X} \sum_i^X \mathbf{I_i}(t), \qquad (5)$$

where X is the total number of frequency bins in the passband of the low-pass filter b_L. Note that this could also be achieved by first normalizing the length of each sound intensity vector in each frequency bin to unity and then applying Eq. (5) to the result. The mean angular direction $[\bar{\theta}(t), \bar{\phi}(t)]$ for the given analysis frame is then computed by converting $\bar{\mathbf{I}}$ back to its polar coordinate representation using Eqs. (1) and (2) by substituting $[I_x \, I_y \, I_z]$ with $[\bar{I}_x \, \bar{I}_y \, \bar{I}_z]$ and disregarding the frequency dependence. The spherical variance is defined using the magnitude of the resultant vector

$$\sigma(t) = 1 - \|\bar{\mathbf{I}}(t)\|, \qquad (6)$$

where $\sigma(t)$ is an indicator of the reliability of the angular mean $[\bar{\theta}(t), \bar{\phi}(t)]$. The operation $\| \cdot \|$ determines the magnitude/Euclidean norm of the enclosed vector. When $\sigma(t) = 1$ the sound intensity vectors are evenly distributed implying a diffuse field, and when $\sigma(t) = 0$ it implies that all vectors are pointing in the same direction and the time frame contains a clear reflection.

Once the analysis has been completed in this way for the SRIR, the resulting signals and metadata, as shown in Fig. 2, can be passed to the synthesis stage for the insertion and manipulation of the missing high-frequency information.

D. SHEM-synthesis

The SHEM-synthesis stages are summarized in block diagram form in Fig. 3, and can be considered generally as a crossfade in the frequency domain between the existing low-frequency W-channel RIR with an extrapolated high-frequency RIR. In the SHEM-analysis stage, the low-frequency RIR is determined by the low-pass filter b_L, and so a matching high-pass filter b_H is used in the SHEM-synthesis stage to determine from where this extrapolation starts, and is represented in the input to the lower, audio signal processing stage of Fig. 3.

The key SHEM-synthesis stage is the introduction of temporally weighted, directionally rendered, high-frequency energy that is matched to the existing low-frequency W-channel RIR. This is done differently depending on whether each analysis frame in the low-frequency RIR is determined as being directional (that is, a reflection) or not (diffuse energy) based on the calculation of spherical variance for that analysis frame across all frequency bins from Eq. (6). This is represented by the diffuse/non-diffuse switch in the audio signal processing stage of Fig. 3. A preset threshold σ_t

J. Acoust. Soc. Am. **145** (4), April 2019
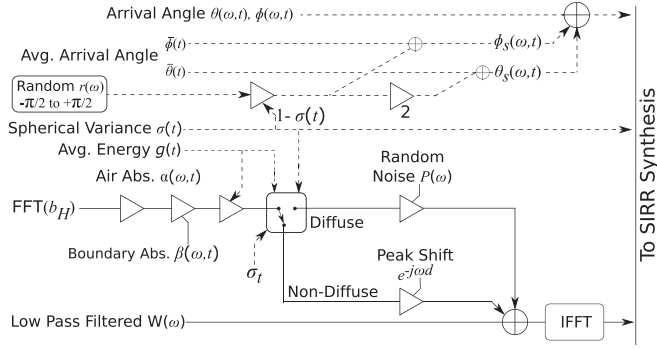
Southern *et al.* 2773

FIG. 3. A block diagram representation of the SHEM-synthesis stages. The dashed and solid lines represent metadata and audio, respectively. In the top part of the figure, high-frequency direction of arrival metadata is generated and added to the existing low-frequency metadata. In the bottom half of the figure, the high-frequency energy is synthesized and added to the existing low-frequency band-limited signal.

is determined such that when $\sigma(t) \leq \sigma_t$ the frame is considered directional. In this work $\sigma_t = 0.35$, and was selected experimentally. A further condition is applied before a frame is considered directional rather than diffuse—the metadata for average energy, $g(t)$, must be locally maximal when compared to the immediately preceding or subsequent frames Hence, in Fig. 3, the diffuse/non-diffuse switch is determined by both metadata variables, $\sigma(t)$ and $g(t)$.

If these two conditions are met, a single sample impulse is high-pass filtered according to b_H, weighted and added to the low-passed input signal in the centre of the current time frame, t_n, the nth time frame of the input signal with frame length N according to

$$\hat{Z}(\omega, t_n) = Z(\omega, t_n) + g(t_n)\alpha(\omega, t_n)\beta(\omega, t_n)B_H(\omega)e^{-j\omega d}, \quad (7)$$

where $\hat{Z}(\omega, t_n)$ is the SHEM-synthesized complex frequency domain signal, $Z(\omega, t_n)$ is the low-frequency input signal and $g(t_n)$ is the averaged energy weighting, all for the given analysis frame t_n. Note that it is assumed that this single sample impulse corresponds to an omnidirectional sound source (high-pass filtered) that emits the same power at all frequencies. Generally, $\alpha(\omega,t)$ is the analytically predicted normalized air absorption function.[18] It has been adapted here so that the absorption weighting changes as a function of the current time frame as this implies a distance traveled. This may be optimized further according to a given humidity value, assumed in this paper to be 50%.

The general term $\beta(\omega, t)$ is a time and frequency dependent approximation representing the high-frequency energy loss due to boundary absorption. Previously this was an arbitrarily chosen non-time-varying value, tapering with frequency such that higher frequencies (within the high-pass band) undergo greater absorption.[1] A new and more flexible method is proposed in Sec. III E.

Finally, with respect to Eq. (7), $B_H(\omega)$ is the frequency domain transform of the high-pass FIR coefficients b_H, and $d = 0.5N$, where N is the length of the time frame, shifting the inserted pulse to the centre of the synthesis frame being considered.

When $\sigma(t) > \sigma_t$ a frame t_n is defined as being diffuse and the amount of high-frequency diffuse energy that is added is determined according to Eq. (8),

$$\hat{Z}(\omega, t_n) = Z(\omega, t_n) + g(t_n)\alpha(\omega, t_n)\beta(\omega, t_n)B_H(\omega)P(\omega), \quad (8)$$

where $P(\omega)$ is a frequency domain generated random pulse train, equivalent to the frame length and whose amplitudes are uniformly distributed between -1 and 1. Note also that for frames defined as being directional, where $\sigma(t) \leq \sigma_t$, but that do not exhibit a locally maximal value for $g(t)$, will not have any diffuse energy will not be applied.

E. EDR modeled boundary absorption

In this implementation, boundary absorption is modeled from the EDR of the existing B-format W-channel RIR to ensure that the decay rate of the high-frequency energy introduced in the SHEM-synthesis stage meets pre-determined frequency dependent reverberation time criteria. Reverberation time data may be obtained from either existing RIR measurements or predictions based on boundary absorption coefficient data. Total boundary absorption is related to overall reverberation time in a given space, and reverberation time can be derived from the Schroeder energy decay curve (EDC) with the EDR being a frequency dependent interpretation of the EDC[19] defined as

$$\text{EDR}_h(\omega, t) = \left| \int_t^\infty h(\tau)e^{-j\omega\tau}\, d\tau \right|^2, \quad (9)$$

where $\text{EDR}_h(\omega, t)$ is the EDR of $h(t)$, the RIR being considered, and can be calculated through the backward integration of the STFT of $h(t)$ and displayed as a 3-D surface. An ideal EDC can be analytically defined using an exponential decay model,[19] and the associated frequency dependent EDR model is therefore given by

$$\begin{aligned}
\text{EDR}(\omega, t) &= A(\omega)e^{k(\omega)t}, \\
A(\omega) &= 10^{A_{\text{dB}}(\omega)/20}, \\
k(\omega) &= \frac{1}{\text{RT60}(\omega)} \ln\left(10^{-60/20}\right), \quad (10)
\end{aligned}$$

where $A_{\text{db}}(\omega)$ is the initial level of the decaying signal in dB at each frequency, and reverberation time $\text{RT60}(\omega)$ is the known (or predicted) frequency dependent reverberation time in seconds.

Note that the average energy weighting function $g(t_n)$ also has an associated EDC as it also shapes the SHEM-synthesis RIR over time, and so when combined with the $\text{EDR}(\omega,t)$ boundary absorption model would result in the final RIR having a shorter than required decay time. To compensate for this double decay rate, a decay rate coefficient k_g must first be calculated based on double the measured RT60 value for $g(t_n)$,

$$k_g = \frac{1}{2\text{RT60}_g} \ln\left(10^{-60/20}\right), \quad (11)$$

where $RT60_g$ is the measured RT60 of $g(t_n)$. The new decay rate coefficients $k_c(\omega)$ for the compensated EDR(ω,t) function can then be calculated,

$$k_c(\omega) = k(\omega) - k_g. \qquad (12)$$

With the correct frequency dependent decay rates determined based on their respective target RT60 values, it remains to define the initial values for $A(\omega)$. This is obtained from the direct sound component of $h(t)$, the spectrum of which characterises the source-receiver pair for a given RIR,[19] and is used to set the relative magnitude values for each frequency in the EDR. In this implementation a straight line approximation of the magnitude spectrum of the Hann windowed direct sound component is used to define $A(\omega)$. Hence, it is now possible to define $\beta(\omega,t)$ representing the time and frequency dependent approximation to the SHEM-synthesis high-frequency energy loss due to boundary absorption as

$$\beta(\omega, t) = A(\omega)e^{k_c(\omega)t}. \qquad (13)$$

Given that the EDR can be calculated accurately up to the low-frequency crossover point of the existing B-format W-channel RIR, this 3-D surface information can then be extended along the frequency axis for each time frame t_n using Eq. (13) and so synthesize the EDR for the unknown high-frequency part.

F. Generating new directional metadata

After SHEM-synthesis the complete RIR plus metadata is passed to an appropriate spatial audio rending algorithm. However, whereas the metadata exists for the original low-frequency W-channel RIR, the extrapolated high-frequency data that has just been synthesized has no such information, and this must also be created based on spherical variance $\sigma(t)$ and average direction of arrival $(\bar{\theta}(t), \bar{\phi}(t))$

$$\theta_s(\omega, t_n) = 2r(\omega)[1 - \sigma(t_n)] + \bar{\theta}(t_n), \qquad (14)$$

$$\phi_s(\omega, t_n) = r(\omega)[1 - \sigma(t_n)] + \bar{\phi}(t_n), \qquad (15)$$

where $\theta_s(\omega, t_n)$ and $\phi_s(\omega, t_n)$ are the randomly generated azimuth and elevation angles for frequency ω in time frame t_n. Note that r is a uniformly distributed random number between $-\pi/2$ and $\pi/2$. The $2r$ in Eq. (14) ensures the random r values have the range $-\pi$ and π. It can be observed that for ideally directional time frames with $\sigma(t_n) = 0$, $\theta_s(\omega) = \bar{\theta}$ and $\phi_s(\omega) = \bar{\phi}$. When $\sigma(t_n) = 1$ the angles are randomly distributed in all directions. This new high-frequency directional metadata is also represented in the upper metadata part of the SHEM-synthesis stages in Fig. 3.

IV. VALIDATION AND RESULTS

This section presents a series of examples and related results that help to demonstrate and validate the ability of SHEM to deal with a varied range of impulse responses, both measured and modeled. The main metrics used to judge the success or otherwise of the method are the acoustic parameters reverberation time (RT60) and early decay time (EDT), both determined from the Schroeder decay curve of the synthesized or reference impulse response. In the former case, RT60 is calculated according to the ISO 3382 parameter T20, that is, based on a linear regression through the $-5\,dB$ and $-25\,dB$ points relative to the maximum value of the given Schroeder decay curve. EDT calculates a value for RT60 based on the first $10\,dB$ of the Schroeder decay curve.[33,34] This implies that as EDT incorporates the early sound decay, it is more susceptible to changes in the RIR direct sound and early reflections, and so acts as a more critical objective metric for the SHEM algorithm. Results are presented in third-octave bands and in addition, for each example, a set of auralizations are prepared to allow the reader a more direct comparison.

A. Example 1: Shoebox room

In this case a B-Format SRIR is calculated for a $10\,m \times 6\,m \times 3\,m$ shoebox room using the FDTD method.[26] With reference to a bottom corner origin point, the source is located at $(7.0,5.0,1.5)$ and receiver at $(5.0,3.0,1.5)$, and a uniform reflection coefficient of $R = 0.98$ is used across all surfaces. The FDTD grid is spatially sampled to give an output rate of $33.3\,kHz$ and the SRIR obtained is high-pass filtered to remove DC-related (or zero frequency) effects[12,13] and low-pass filtered to remove the effects of dispersion error, giving a bandwidth assumed valid up to $1.6\,kHz$.

The results that follow for this relatively simple example enable the various SHEM stages to be illustrated, and given that a full bandwidth response is not available for this typical example of a FDTD simulation, a SRIR obtained from the image-source method (ISM) is used instead based on the same input parameters. The RIR obtained from the ISM example also serves as a calibration source from which the target RT60(ω) values used to calculate $k_c(\omega)$ in Eq. (13) are obtained. The direct sound, in this case a windowed impulse signal, is used to define the initial values of $A(\omega)$, also in Eq. (13), so leading to the estimation of the SHEM-synthesis $\beta(\omega,t)$ parameter.

The results produced from three stages in the SHEM-synthesis process are presented in Fig. 4 in the form of EDR plots. Figure 4(a) shows the EDR of the post-processed B-format SRIR W-channel obtained from the FDTD simulation, demonstrating the low-frequency part available as the input to the SHEM-analysis stage, and the missing region required to be synthesized using SHEM. Figure 4(b) shows the result produced directly from the SHEM-synthesis stage where temporally weighted high-frequency energy has been added to the existing low-frequency signal. Note that the synthesis gives results that vary with frequency according to the estimations of $A(\omega)$ and RT60(ω) that define $\beta(\omega,t)$, obtained from the reference ISM result. Figure 4(c) presents the EDR of the final SHEM impulse response. In this example, the air absorption function $\alpha(\omega,t)$ has been applied to the result shown in Fig. 4(b), resulting in an overall smoothing of the response and faster attenuation at higher frequencies with increasing time.

J. Acoust. Soc. Am. **145** (4), April 2019
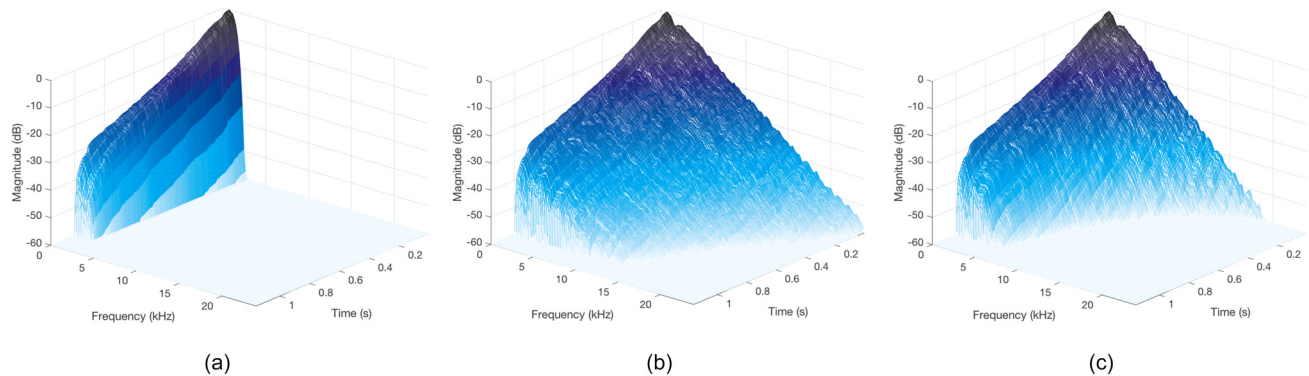
Southern *et al.* 2775

FIG. 4. (Color online) Energy decay reliefs showing three stages of SHEM-synthesis. (a) B-format W-channel SRIR obtained from the FDTD simulation of the shoebox room with crossover at 1.6 kHz; (b) post-SHEM-synthesis; (d) the final result after air absorption $\alpha(\omega,t)$ has been applied.

Figure 5 shows the first 1000 samples of the overlaid time-domain responses obtained as part of the SHEM-synthesis process. The B-format W-channel SRIR obtained from the ISM, used as the reference case for this example, is presented as a grey solid line. The corresponding FDTD result is shown as the blue dotted line, and this relates directly to the EDR plot shown in Fig. 4(a). As part of the SHEM-analysis stage, each (low-frequency) frame is defined as either being directional or diffuse according to the spherical variance parameter, $\sigma(t)$, from Eq. (6) and processed according to the SHEM-synthesis processing steps defined by Eqs. (7) and (8), respectively. The result of this process can be seen in the black solid line that corresponds to the EDR plot shown in Fig. 4(c). It can be observed that the early reflections have been correctly identified with diffuse energy added where clear directional components are less evident. Note that the first reflection has a greater relative amplitude compared with the direct sound, and this can also be observed in the FDTD input signal and is due to multiple reflections correctly summing at the receiver location.

Finally, for this example, third-octave room acoustic parameters RT60 and EDT are considered as shown in Figs. 6(a) and 6(b), respectively. The black solid line with error bars shows results for the original FDTD outcome up to 1.6 kHz and above this for the ISM reference case, which is used to calibrate the SHEM-synthesis according to Eq. (13),

noting that although the simulation and results generated are comparable to the FDTD model, the two methods are not exact. The error bars represent the quoted just noticeable difference (JND) values for RT60 and EDT,[33] and give a measure of the expected perceptual similarity of the original reference result to the SHEM output. The blue solid line with circle markers is the result from the FDTD-SHEM example with only boundary absorption applied and so correspond to the EDR plot in Fig. 4(b). The final FDTD-SHEM result is represented by the red line with dot markers, corresponding to the EDR plot in Fig. 4(c). Results are presented above 800 Hz only, given that the crossover point above which SHEM is applied is generally greater than this and, in this example, is fixed at 1.6 kHz.

Hence, it can be seen that in comparing the FDTD-ISM reference case and the FDTD-SHEM results without air absorption in terms of the target RT60 values, all results across all third-octave bands are within the limits of the JND, indicating no perceptual difference between the two. This should be expected as the ISM RT60 values are used to calibrate the SHEM-synthesis. For EDT, all results are within the limits of the JND apart from 1.6 kHz and 8 kHz third-octave bands, and it is important to note that the EDT is not part of the calibration process and tends to be a more sensitive or variable parameter than RT60 due to its increased dependence on the early part of the RIR.
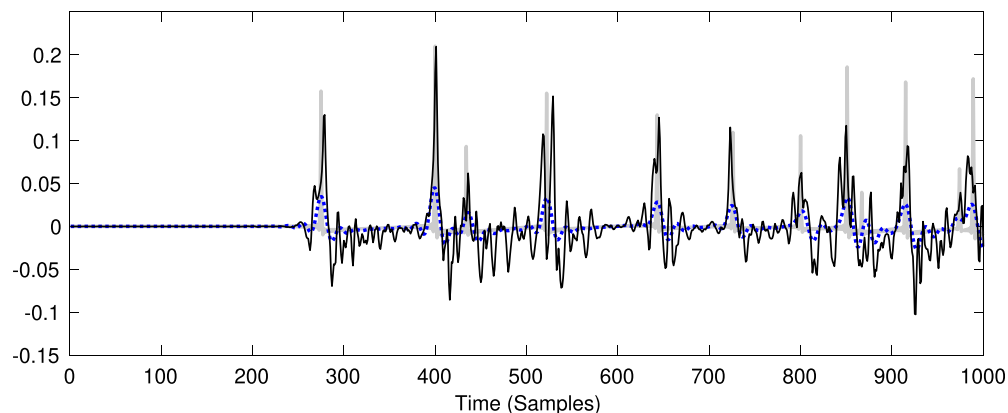

FIG. 5. (Color online) The first 1000 sample overlaid time domain impulse responses highlighting the results of the SHEM-synthesis process: (grey solid line) B-format W-channel SRIR obtained from the ISM simulation of the shoebox room used as the full bandwidth reference case; (blue dotted line) B-format W-channel SRIR obtained from the FDTD simulation of the shoebox room, valid up to 1.6 kHz; (black solid line) the final SHEM result.
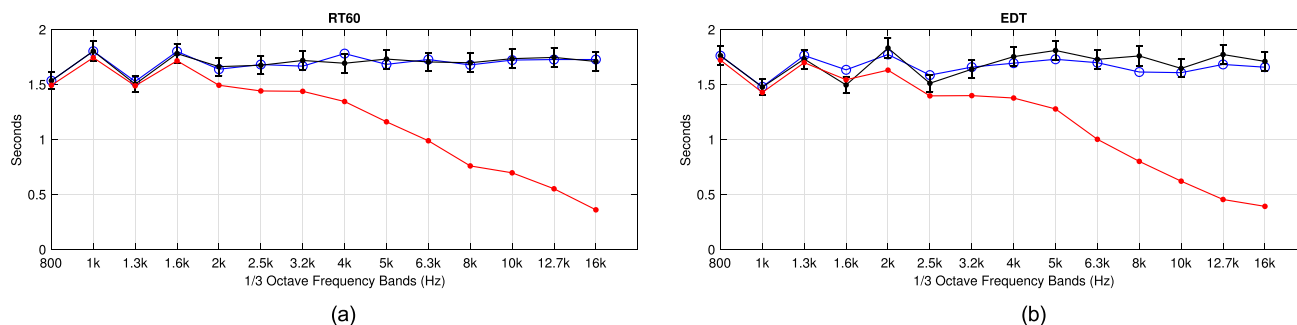
FIG. 6. (Color Online) Third-octave room acoustic parameters for the shoebox room example. The black solid line with error bar is the FDTD result up to 1.6 kHz and the ISM reference case above, the blue solid line with circle markers is the SHEM example with only boundary absorption having been applied, and the red solid line with dot markers the final SHEM result with both boundary and air absorption applied. (a) Reverberation time; (b) EDT.

B. Example 2: OpenAIR measurements

With the SHEM method established in the results presented in Sec. IV A it is required to test the potential of this approach across a more varied range of spaces, preferably with a clear reference to give a better indication of the success or otherwise of the algorithm. An earlier version of SHEM was applied to three Finnish concert halls and a large cathedral,[1] with much better results obtained for the cathedral example due to the dominance of air absorption over boundary absorption in a very large space, and the limited adaptability of the boundary absorption parameter used in that study. It is hypothesized that the improved boundary absorption parameter $\beta(\omega,t)$ presented

in this paper should be able to deal with a broader range of spaces. The open acoustic impulse response (OpenAIR) library[25,35] is an online repository of B-format RIRs available for research or creative use. Three contrasting examples have been selected in order to test the ability of SHEM to approximate the high-frequency part of the reference impulse response. In each case the crossover point for the algorithm is defined as 2.2 kHz, as used in the original SHEM work.[1]

1. York Minster, York, UK

York Minster is the largest medieval gothic cathedral in Northern Europe and is approximately 160 m long, 76 m
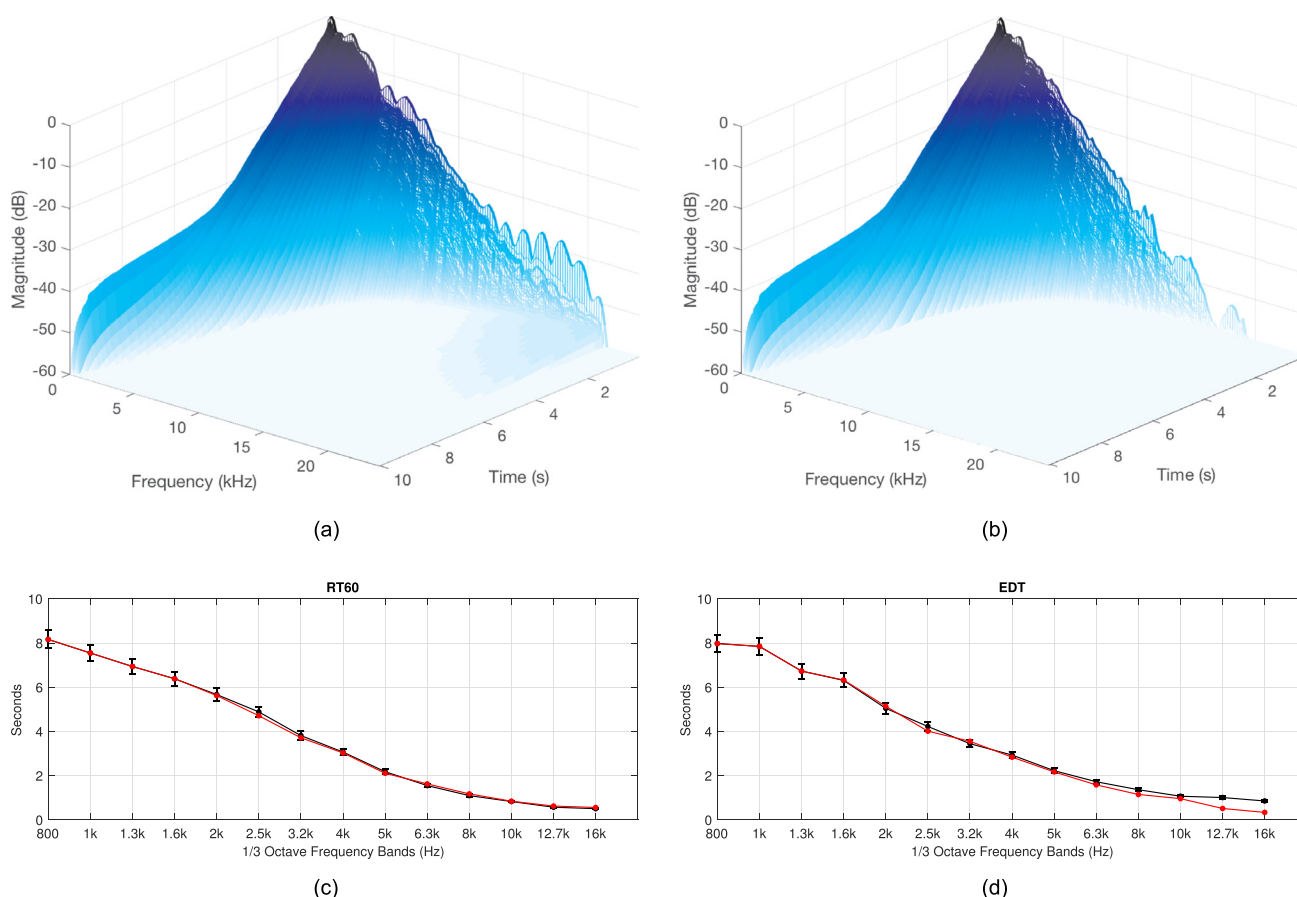

FIG. 7. (Color online) York Minster (large cathedral): (a) EDR of B-format W-channel SRIR; (b) EDR of final SHEM output; (c) third-octave band RT60 values, black line is the reference case with error bars representing JND, red line with dot markers is the SHEM-result obtained using only boundary absorption; (d) third-octave band EDT values.

wide, and 27 m high to the vaulted ceiling and is constructed predominantly of stone. As it is also the best-case example from the original SHEM paper, it is important to ensure that the new $\beta(\omega,t)$ used here does not have a negative impact on the result obtained previously. Results are presented in Fig. 7, showing the original input and post-SHEM EDRs, together with third-octave band values for RT60 and EDT. In this example a measured RIR is used as the SHEM input, and so this is also used as the calibration target for SHEM-synthesis in terms of one-third octave band RT60 values, RT60(ω), and initial EDR magnitude estimation, $A(\omega)$, from the windowed direct sound component. The RT60 values estimated from such a measured RIR will already incorporate the effects of air absorption, and so the $\alpha(\omega,t)$ term is not used as part of the SHEM-synthesis in this case.

From Figs. 7(a) and 7(b) it can be noted that the overall post-SHEM EDR profile visually provides a close match to the original. In terms of RT60 and EDT results, when comparing to the reference case, all values are within the limits of the JND apart from EDT values for 12.7 kHz and 16 kHz third-octave bands.

2. St. Margaret's Church, York, UK

This building also dates from the early medieval period and has been extensively repurposed as a modern music performance venue, which includes variable acoustic panels and drapes as part of its design. It has a flagstone floor, lime-wash on stone walls with some wood paneling, and a wooden roof. Its approximate dimensions are 24 m × 12.5 m × 11.2 m in height. Results are presented in Fig. 8 as for the York Minster case. In terms of RT60, only the result for the 3.2 kHz third-octave band lies outside of the limits of the JND compared to the reference case. For EDT, the 3.2 kHz, 4 kHz, and 12.7 kHz results are not within the limits of the JND. Although St. Margaret's Church has been optimized for musical performance, it still demonstrates architectural features relating to its original use (e.g., supporting columns, a coupled volume to what was the church tower), and it is these aspects that influence variation in EDT while the reverberant features are captured and approximated well by the SHEM analysis/synthesis process. This also demonstrates the successful application of SHEM for spaces that are not overly dominated by their reverberant field, such as York Minster with an RT60 of almost 8 s at 1 kHz, and even the shoebox room case with an RT60 of ~1.8 s at 1 kHz.

3. Maes Howe, Orkney, UK

This is an example of a chambered cairn dating from 3000 BC, almost cubic in shape, of dimension 4.6 m with walls made from large, flat slabs of stone, resulting in smooth reflecting surfaces and a very resonant sound. It has a long, narrow entrance passageway to the main chamber
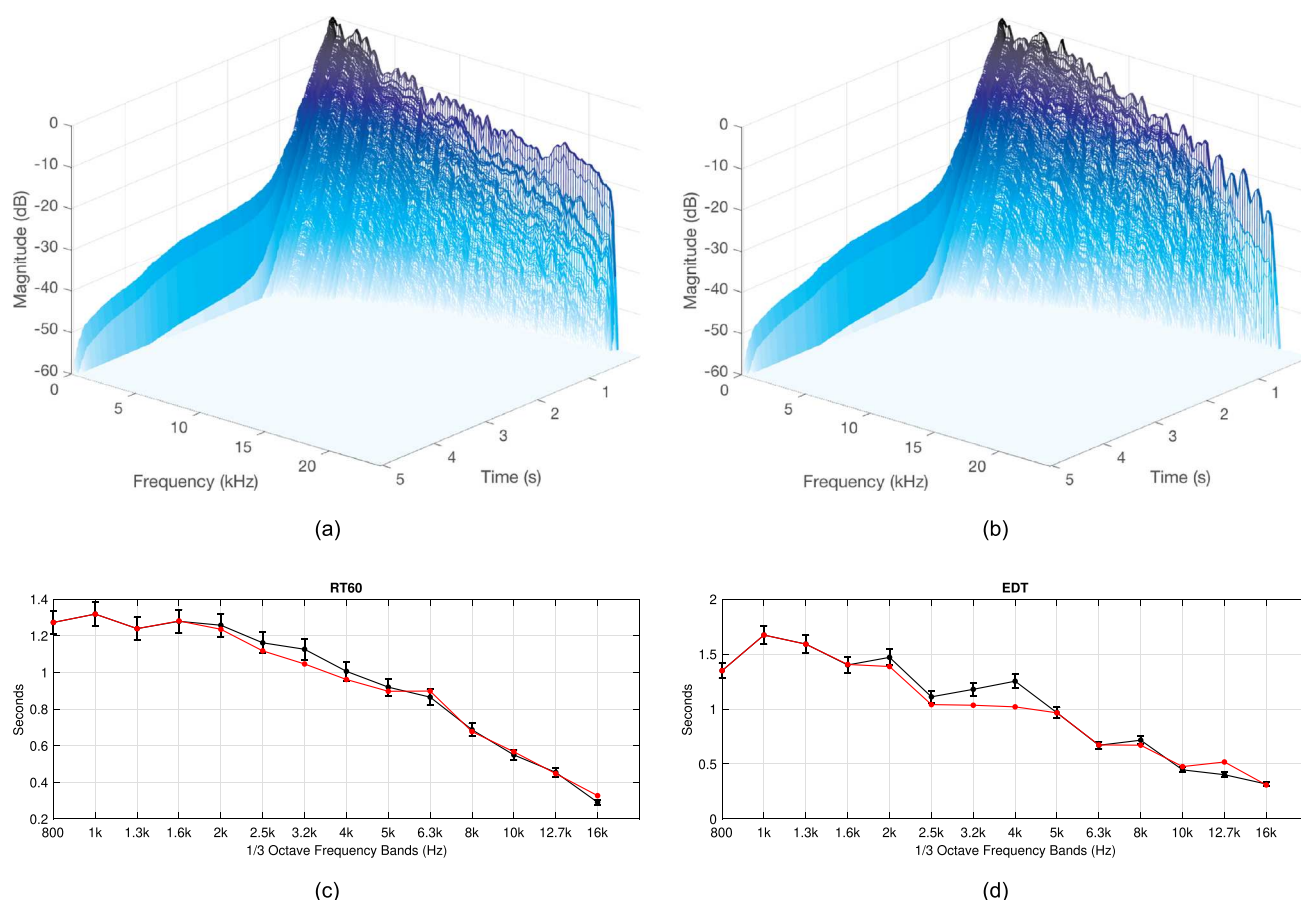

FIG. 8. (Color online) St. Margaret's (medium-sized church): (a) EDR of B-format W-channel SRIR; (b) EDR of final SHEM output; (c) third-octave band RT60 values, black line is the reference case with error bars representing JND, red line with dot markers is the SHEM-result obtained using only boundary absorption; (d) third-octave band EDT values.

from which three much smaller, approximately cubic antechambers are arranged. Results are presented in Fig. 9 as in previous examples. In terms of RT60, all results except the 5 kHz third-octave band are within the limits of the JND when compared to the reference case. However, EDT demonstrates greater variability with all results above 2 kHz, except for the 6.3 kHz third-octave band being outside of the limits of the JND. In particular, it appears that there is some feature in the actual space resulting in a peak in RT60 at 3175 Hz, and a corresponding drop in EDT, possibly due to the three coupled volumes off the main space or the effect of multiple strong reflections. As the crossover point is at 2.2 kHz, there is no way of capturing this feature in the SHEM approximation and so it is unsurprising that the algorithm performs less well in this region.

C. Example 3: Hybrid concert hall simulation

A hybrid room acoustics simulation method based on 3-D FDTD for low frequencies, combined with beam-tracing and acoustic radiance transfer has enabled FDTD-based RIRs to be used as the foundation for full bandwidth auralization.[4] The results presented in Sec. IV A demonstrate how SHEM can be used as an alternative means of complementing a low-frequency FDTD simulation, resulting in a full audio bandwidth RIR, and these results are now extended for this example case study. A fictional concert hall is used[4]

with dimensions 24.4 m × 17.8 m × 10.5 m and a volume of 3802 m^2. It has a curved stage ceiling and different absorption coefficients attributed to each boundary over octave bands from 32 Hz to 16 kHz. The FDTD grid is spatially sampled to give an output rate of 10 kHz, and a valid low-frequency response is assumed up to around 1.5 kHz after appropriate low-pass/high-pass filtering to remove dispersion error and DC-related (or zero frequency) effects. A beam-tracing/acoustic radiance transfer (BT/ART) model was used for the 2 kHz octave band and above and a B-format SRIR obtained across the three simulation methods to arrive at a final result. Additional details relating to the implementation of this model are available.[4,36] For the SHEM implementation, the crossover point is defined as 2 kHz and the results, in the same format as those presented in Sec. IV B, are presented in Fig. 10. In this case both RT60 and EDT results are within the limits of the JND when compared with the reference case, apart from the 2.5 kHz, 5 kHz, 6.3 kHz, and 16 kHz third-octave bands.

D. Audio examples

A full set of auralizations for each of the examples presented in this section are available online.[37] Two source sounds are used—one being a synthesized drum loop with the sharp transients providing a good means of highlighting details in the resulting auralization; the other is an anechoic
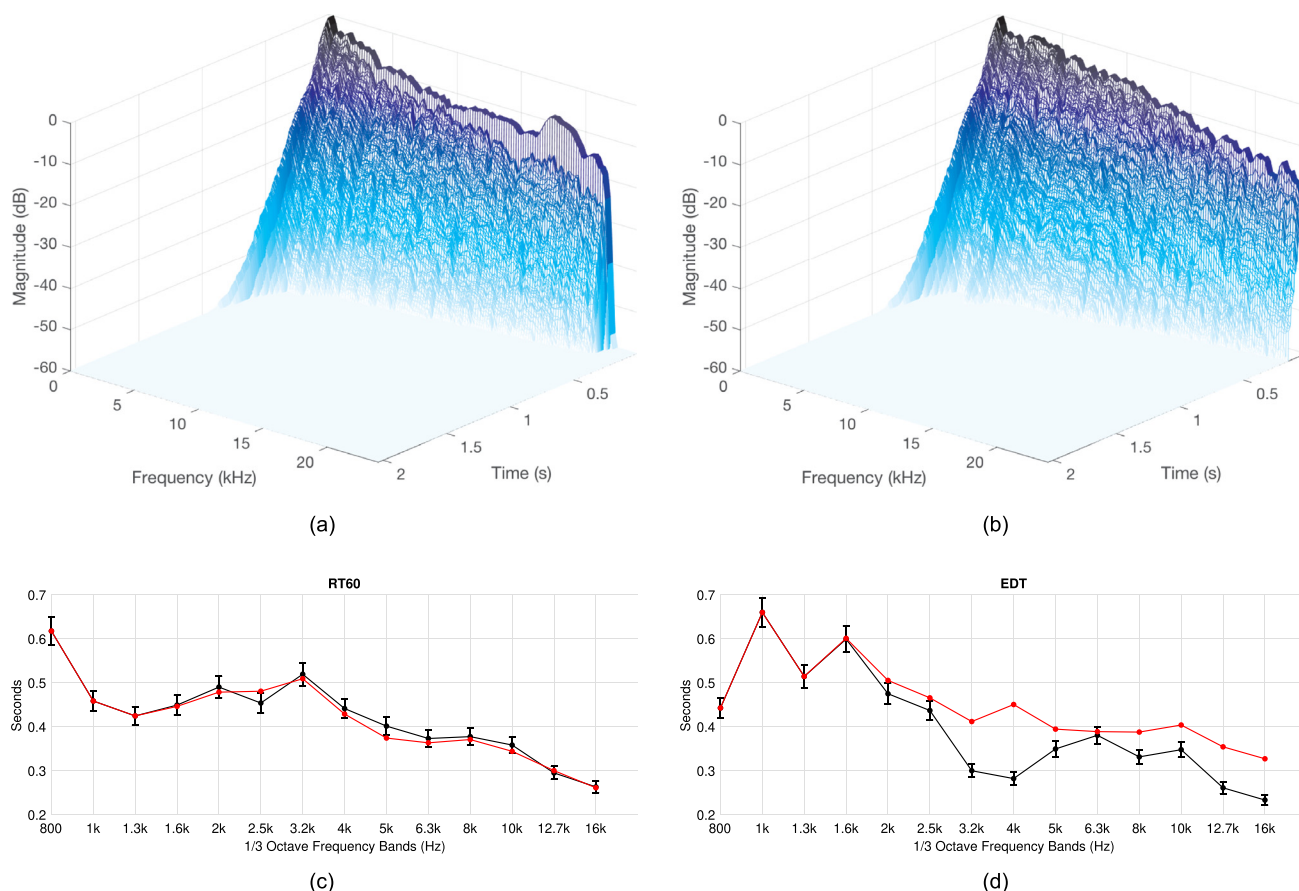

FIG. 9. (Color online) Maes Howe (small stone chambered cairn): (a) EDR of B-format W-channel SRIR; (b) EDR of final SHEM output; (c) third-octave band RT60 values, black line is the reference case with error bars representing JND, red line with dot markers is the SHEM-result obtained using only boundary absorption; (d) third-octave band EDT values.
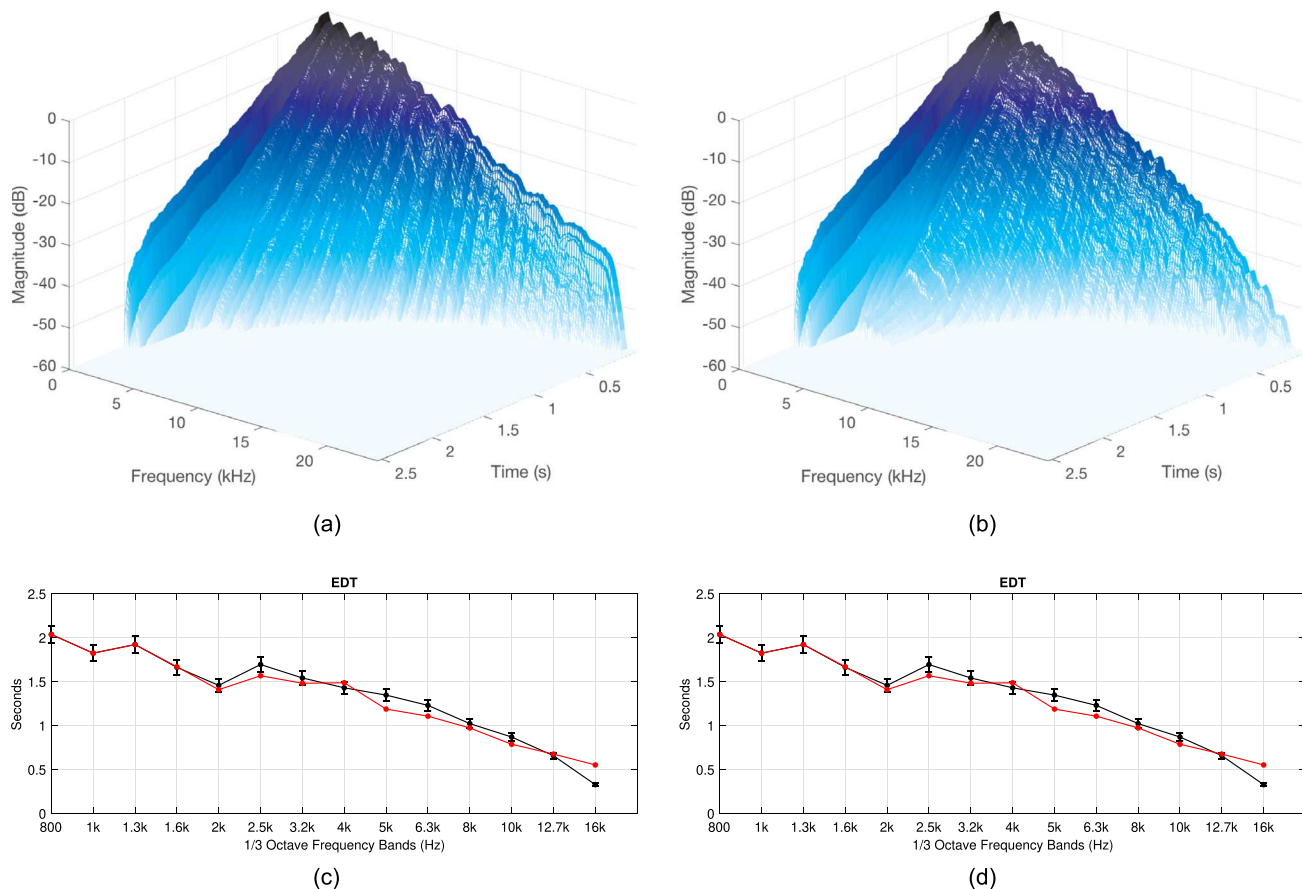
FIG. 10. (Color online) SHEM-synthesis of a FDTD/BT/ART hybrid model of a medium sized concert hall: (a) EDR of B-format W-channel SRIR; (b) EDR of final SHEM result; (c) third-octave band RT60 values, black line is the reference case with error bars representing JND, red line with dot markers is the SHEM-result obtained with only boundary absorption; (d) third-octave band EDT values.

recording of a four-part vocal performance. All examples are in mono only as they are designed to highlight the success or otherwise of SHEM, rather than any one specific rendering technique. For each of the five spaces considered (shoebox room, York Minster, St. Margaret's Church, Maes Howe, hybrid concert hall) there are two sets of examples based on these two source sounds. Each example set has three auralizations—the original reference case, the low-frequency-only result on which the SHEM analysis/synthesis is based, and the final SHEM output.

In informal listening, SHEM gives very good results. All examples present a plausible and natural sounding extension to the low-frequency-only case, maintaining the characteristics of the original with the Maes Howe case being an excellent case in point given that the objective results for EDT in Sec. IV B were the least well matched to the reference data in terms of being within the limits of the JND. In fact, it is only when the reference case and the SHEM example are compared directly that some differences become more obvious, and this is perhaps most noticeable with the shoebox room where the reference case auralization has been sourced from an ISM-only-based example rather than the FDTD method.

It is proposed that there may be some additional refinement required in the addition of the high-frequency diffuse energy in the SHEM-synthesis stage, and some further comments on this are offered in the consideration of

possible future work. However, what has been clearly demonstrated here is the ability of the newly proposed EDR-based SHEM boundary absorption function $\beta(\omega,t)$ to deal with a wider and more varied range of acoustic spaces, giving auralizations based on an approximation only, which are highly plausible in terms of what they deliver perceptually.

V. CONCLUSIONS

This paper has considered in more detail the application of the SHEM as a means to extend low-frequency band-limited SRIRs obtained from FDTD wave-based acoustic-modeling methods to a natural sounding, full bandwidth auralization. In particular, a new method has been presented for approximating the required boundary absorption weighting function that allows SHEM to be applied successfully to a broader range of spaces than was possible in earlier work. This is based on an approximation of the EDR for the source SRIR with the EDR produced by the final SHEM output giving good agreement to the target original when available as a reference. Reverberation time (RT60), and EDT room acoustic parameters provide a more objective measure of the subjective reliability of the method and indicate that the technique can deal with spaces of differing type and volume, whether from measurement or simulation. Auralizations help to verify and contextualize these objective results and give

audible examples demonstrating that in the majority of cases, RT60 values for the reference case and SHEM output are within the quoted JND. EDT results show some more variation, compared to RT60, and indicate that there is some further fine-tuning that might be applied. However, SHEM demonstrates significant promise as a ready means of synthesizing the high-frequency part of a low-frequency-only room acoustics simulation using parametric methods for auralization purposes. SHEM therefore has potential application in both creative sound and architectural design, as well as for real-time, interactive room acoustics rendering for gaming and virtual reality.

Further work will consider the nature of the diffuseness estimation and diffuse energy synthesis, making a comparison with other diffuseness estimators, and also consider the diffuse nature of the source SRIR as a means of gathering additional information to apply in the SHEM-synthesis stage. The underlying assumptions on which this method is based should also be considered in some more depth as there are potential limitations. For instance, given that this method is synthesizing a target reverberant field, the diffuse nature of this field is assumed, and unexpected high-frequency behaviour cannot be predicted, as evidenced in the Maes Howe example, in particular. Furthermore, SIRR, as implemented in this work, is only able to determine a single reflection within a given time window, and multiple coinciding reflections will be merged. The EDR might also be calibrated differently, for instance, approximating $\beta(\omega,t)$ using EDT rather than RT60 values in Eq. (13) may help to refine the SHEM-synthesis stage further. Other acoustic parameters could also be considered. Testing is also required to determine the crossover point (noting that three different values have been used across the examples presented in this paper), in either objective or perceptual terms, as to where SHEM could be optimally applied. Listening tests will be used to determine the success or otherwise of any further refinements. Finally a real-time implementation of SHEM should also be considered, leading to the potential for interactive auralization based on this approach.

ACKNOWLEDGMENTS

[1]A. Southern and L. Savioja, "Spatial high frequency extrapolation method for room acoustic auralization," in *Proc. 15th Int. Conf. on Digital Audio Effects (DAFx-12)*, York, UK (Sept. 17–21, 2010), pp. 145–149.

[2]D. Botteldooren, "Finite difference time domain simulation of low frequency room acoustic problems," J. Acoust. Soc. Am. **98**(6), 3302–3308 (1995).

[3]K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes," IEEE Trans. Audio Speech Lang. Process. **19**(1), 34–46 (2011).

[4]A. Southern, S. Siltanen, D. Murphy, and L. Savioja, "Room impulse response synthesis and validation using a hybrid acoustic model," IEEE Trans. Audio Speech Lang. Process. **21**(9), 1940–1952 (2013).

[5]L. Savioja, "Real-time 3D finite-difference time-domain simulation of low- and mid-frequency room acoustics," in *Proc. 13th Int. Conf. Digital Audio Effects (DAFx-10)*, Graz, Austria (Sept. 6–10, 2010), pp. 43–50.

[6]C. J. Webb and A. Gray, "Large-scale virtual acoustics simulation at audio rates using three dimensional finite difference time domain and multiple GPUs," in *Proc. Int. Congress on Acoustics*, Montréal, Canada (June 2–7, 2010).

[7]S. Bilbao, B. Hamilton, C. J. Webb, and N. Fletcher, "Finite difference room acoustics simulation with general impedance boundaries and viscothermal losses in air: Parallel implementation on multiple GPUs," in *Proc. Int. Symp. on Musical and Room Acoustics*, Buenos Aires, Argentina (Sept. 11–13, 2016).

[8]L. Stoltzfus, A. Gray, C. Dubach, and S. Bilbao, "Performance portability for room acoustics simulations," in *Proc. 20th Int. Conf. Digital Audio Effects (DAFx-17)*, Edinburgh, UK (Sept. 5–9, 2017), pp. 367–374.

[9]L. Savioja and V. Välimäki, "Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques," IEEE Trans. Speech, Audio Process. **8**(2), 184–194 (2000).

[10]A. Southern, T. Lokki, L. Savioja, and D. T. Murphy, "The perceptual effects of dispersion error on room acoustic model auralization," in *Proc. of Forum Acusticum*, Aalborg, Denmark (June 27–July 1, 2011).

[11]J. Saarelma, J. Botts, B. Hamilton, and L. Savioja, "Audibility of dispersion error in room acoustic finite-difference time-domain simulation as a function of simulation distance," J. Acoust. Soc. Am. **139**(4), 1822–1832 (2016).

[12]D. T. Murphy, A. Southern, and L. Savioja, "Source excitation strategies for obtaining impulse responses in finite difference time domain room acoustics simulation," Appl. Acoust. **82**, 6–14 (2014).

[13]J. Botts and L. Savioja, "Spectral and pseudospectral properties of finite difference models used in audio and room acoustics," IEEE/ACM Trans. Audio Speech Lang. Process. **22**(9), 1403–1412 (2014).

[14]V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," IEEE Trans. Audio Speech Lang. Process. **20**(5), 1421–1448 (2012).

[15]P. Coleman, A. Franck, P. J. B. Jackson, R. J. Hughes, L. Remaggi, and F. Melchior, "Object-based reverberation for spatial audio," J. Audio Eng. Soc. **65**(1/2), 66–77 (2017).

[16]R. Väänänen and J. Huopaniemi, "Advanced audiobifs: Virtual acoustics modeling in mpeg-4 scene description," IEEE Trans. Multimedia **6**(5), 661–675 (2004).

[17]A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses," J. Audio Eng. Soc. **60**(11), 887–898 (2012), available at http://www.aes.org/e-lib/browse.cfm?elib=16633.

[18]H. E. Bass, H. J. Bauer, and L. B. Evans, "Atmospheric absorption of sound: Analytical expressions," J. Acoust. Soc. Am. **52**(3B), 821–825 (1972).

[19]J. M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. of the Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, San Francisco, CA (March 23–26, 1992), pp. II.221–II.224.

[20]M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral band replication, a novel approach in audio coding," in *Proc. of the 112th AES Convention*, Munich, Germany (May 10–13, 2002), paper no. 5553.

[21]C.-M. Liu, W.-C. Lee, and H.-W. Hsu, "High frequency reconstruction for band-limited audio signals," in *Proc. 6th Int. Conf. on Digital Audio Effects (DAFx-03)*, London, UK (Sept. 8–11, 2003).

[22]E. Larsen, R. M. Aarts, and M. Danessis, "Efficient high-frequency bandwidth extension of music and speech," in *Proc. of the 112th AES Convention*, Munich, Germany (May 10–13, 2002), paper no. 5627.

[23]C.-C. Bao, X. Liu, Y.-T. Sha, and X.-T. Zhang, "A blind bandwidth extension method for audio signals based on phase space reconstruction," EURASIP J. Audio, Speech, Music Process. **2014**(1), 661–675 (2014).

[24]B. Bernschütz, "Bandwidth extension for microphone arrays," in *Proc. of the 133rd AES Convention*, San Francisco, CA (Oct. 26–29, 2012), paper no. 8751.

[25]S. Shelley and D. T. Murphy, "Openair: An interactive auralization web resource and database," in *Proc. of the 129th AES Convention*, San Francisco, CA (Nov. 4–7, 2010), paper no. 8226.

[26]A. Southern, D. T. Murphy, and L. Savioja, "Spatial encoding of finite difference time domain acoustic models for auralization," IEEE Trans. Audio Speech Lang. Process. **20**(9), 2420–2432 (2012).

[27]J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis," J. Audio Eng. Soc. **53**(12), 1115–1127 (2005), available at http://www.aes.org/e-lib/browse.cfm?elib=13401.

[28]C. Faller and V. Pulkki, "Directional audio coding: Filterbank and STFT-based design," in *Proc. of the 120th AES Convention*, Paris, France (May 20–23, 2006), paper no. 6658.

[29]S. Berge and N. Barrett, "High angular resolution planewave expansion," in *Proc. of the 2nd Int. Symp. on Ambisonics and Spherical Acoustics*, Paris, France (May 6–7, 2010).

[30]S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," J. Audio Eng. Soc. **61**(1/2), 17–18 (2013), available at http://www.aes.org/e-lib/browse.cfm?elib=16664.

[31]F. J. Fahy, *Sound Intensity* (Elsevier Applied Science, London, 1989).

[32]S. Tervo, T. Korhonen, and T. Lokki, "Estimation of reflections from impulse responses," J. Build. Acoust. **18**(1), 159–174 (2011).

[33]ISO 3382-1, "Acoustics, measurement of room acoustic parameters. Part 1: Performance spaces" (International Organization for Standardization, Brussels, 2009).

[34]ISO 3382-2, "Acoustics, measurement of room acoustic parameters. Part 2: Reverberation time in ordinary rooms" (International Organization for Standardization, Brussels, 2009).

[35]D. T. Murphy and S. B. Shelley, "Openair—The open acoustic impulse response library" (University of York, York, UK, 2014), available at http://www.openairlib.net/ (Last viewed July 31, 2018).

[36]A. Southern, "Hybrid model supporting material and sound examples" (Aalto University, Espoo, Finland, 2012), available at http://mediate-ch.aalto.fi/publications/acoustics/ham/index.html (Last viewed July 31, 2018).

[37]D. T. Murphy, "Impulse response synthesis for hybrid parametric room acoustics simulation using the spatial high-frequency extrapolation method sound examples" (University of York, York, UK, 2018), available at http://www-users.york.ac.uk/~dtm3/SHEM/index.html (Last viewed July 31, 2018).