Room Impulse Response Synthesis and Validation Using a Hybrid Acoustic Model

Alex Southern, Samuel Siltanen, Damian T. Murphy, and Lauri Savioja, *Senior Member, IEEE*

***Abstract*—Synthesizing the room impulse response (RIR) of an arbitrary enclosure may be performed using a number of alternative acoustic modeling methods, each with their own particular advantages and limitations. This article is concerned with obtaining a hybrid RIR derived from both wave and geometric-acoustics based methods, optimized for use across different regions of time or frequency. Consideration is given to how such RIRs can be matched across modeling domains in terms of both amplitude and boundary behavior and the approach is verified using a number of standardised case studies.**

***Index Terms*—FDTD, image-source, auralization, hybrid acoustic modeling.**

I. Introduction

PREDICTING wave propagation behavior within an arbitrary enclosure is beneficial in many areas of science. The simulation of sound wave propagation is particularly important as part of the architectural design process without which it is not possible to audition the acoustic qualities of a building before it has been constructed. Acoustic modeling techniques therefore provide an early indication of the objective acoustic properties of non-existent spaces through the analysis of the synthesized *room impulse response* (RIR) and subsequent derivation of standard acoustical parameters [1], [2]. Such methods can then also be extended with some confidence to other virtual reality type applications, as found in first person gaming environments [3], [45], or in the reconstruction of historic buildings that no longer exist, or exist only in part [4].

Fig. 1 is an overview of the typical acoustic design and evaluation workflow for a given environment that supports the analysis of established objective metrics with subjective analysis based on critical listening. In this situation the designer is presented with a rendered soundfield derived directly from the synthesized RIRs. Hence, the *RIR synthesis and spatial encoding* stage combined with the subsequent representation in a format suitable for audible rendering, result in an *auralization* of the modeled enclosed space. In this context auralization allows the listener to experience auditory cues that would occur if the listener was present in the equivalent real-world listening environment. However in practice the auditory cues are only representative of the listening situation rather than being absolutely identical. Realizing this subjectively informed method of design, particularly with respect to concert hall acoustics, is the overarching concept that motivates the contributions of this article.

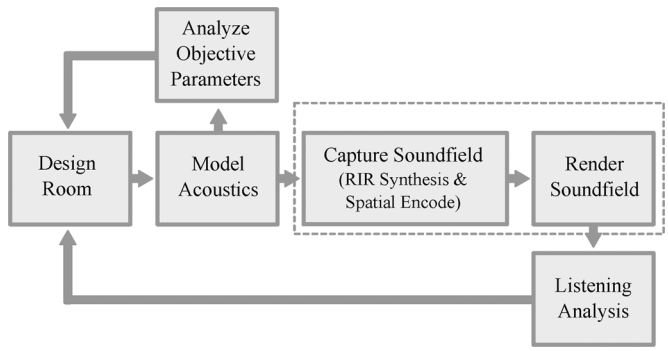


Fig. 1. An overview of a typical acoustic design and evaluation workflow for a given environment based on the analysis of both objective measures obtained from RIR synthesis, and subjective listening via auralization.

Currently established and accepted methods of predicting room acoustic characteristics are based on a *geometric-based* model of ray-like sound propagation. This approach is valid when representing wavelengths that are small in comparison to the dimensions of the bounding enclosure and internal objects. The main difference between these methods is in the discovery or approximation of reflection paths [5]. Each geometric method has its associated strengths and limitations although generally all rays that intersect a chosen listening region are recorded in a list for their energy, path length/time-of-arrival and angle-of-arrival, providing sufficient information to build a reflection *echogram*. At larger wavelengths the ray-like assumption no longer holds and this leads to a misrepresentation of the low frequency response of the acoustic model [6].

An alternative modeling approach is offered through *wave-based* methods that have been shown to be more appropriate in the low frequency range, e.g., [7]–[9]. These methods are based on a discrete numerical solution to the wave equation and are therefore a proper treatment of physical wave motion. However, they are computationally intensive, especially when modeling large volumes and/or a wide frequency bandwidth such as the range of hearing. This is due to the large number of elements/calculations required to approximate continuous wave propagation with sufficient accuracy using a discrete system.

Most previous research in attempting to develop a complete and accurate solution to synthesizing a RIR have been based

Manuscript received August 21, 2012; revised December 19, 2012, April 16, 2013; accepted April 18, 2013. Date of publication May 14, 2013; date of current version July 12, 2013. This work was supported in part by the Academy of Finland under Project No. 138780 and in part by EPSRC Grant EP/J000108/1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Woon-Seng Gan.

A. Southern, S. Siltanen, and L. Savioja are with the Department of Media Technology, Aalto University School of Science, 02150 Espoo, Finland (e-mail: mrapsouthern@gmail.com; saasilta@tml.hut.fi; lauri.savioja@aalto.fi).

D. T. Murphy is with the AudioLab, Department of Electronics, University of York, North Yorkshire YO10 5DD, U.K. (e-mail: dtm3@ohm.york.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2013.2263139

on a combination of geometric methods in order to increase the computational efficiency. Some limited work has explored combinations of wave and geometric methods, and it is this approach that is adopted in this article based on a combination of the Finite Difference Time Domain (FDTD), Beam-Tracing (BT) and Acoustic Radiance Transfer (ART) methods. It should be noted that the main priority of this hybrid model is to synthesize accurate impulse responses over the entire audible spectrum rather than minimize computation time.

The contributions of this article are, 1) an improved method of correctly balancing the relative energy in synthesized RIRs produced by the geometric and the FDTD methods, namely *calibration*. A generic calibration parameter is presented and derived for the employed FDTD scheme, which has clear advantages over the previously discussed methods. 2) the complete validation of a new *Hybrid Acoustic Model* (HAM) first presented in [10]. In particular this article considers the optimal combination of the synthesized acoustic energy responses obtained from geometric methods with the acoustic pressure responses of wave-based methods. Reflection coefficients in the two modeling domains are characterized and compared based on the proposed calibration method. The results are tested against established methods for typical room geometry examples of increasing complexity. Each acoustic modeling component of the HAM is shown to give comparable results within the *just-noticeable-difference* JND of a number of standard acoustic parameters. In particular, the late reverberation modeling is shown to be a clear improvement over the most recent HAM contribution. Finally, hybrid room impulse responses are synthesized using the standard dataset presented in [40] allowing predicted reverberation times to be compared with that of real measurement data.

The structure of this paper is as follows. Section II considers the general problem of acoustic modeling for RIR synthesis, and previously presented hybrid methods. In Section III a novel HAM is presented followed by Section IV introducing and validating the newly proposed calibration technique for mixing geometric and wave-based RIRs. Section V assesses the accuracy of the proposed HAM against established methods using a series of test case scenarios. Section VI considers the results of synthesizing hybrid RIRs based on this new approach and finally Section VII gives a final discussion and concluding remarks.

II. Strategies in Hybrid Room Acoustic Modeling

Geometric and wave based methods are both potentially capable of modeling a given room geometry and synthesizing results valid across the whole audible spectrum. More recently, increases in computational power have made it more practical and realistic to apply computationally intensive wave-based methods to real room acoustic modeling problems with a subsequent increase in related research in a number of areas (e.g., [7], [11]–[16]). In addition there has been a notable increase in the number of publications concerning acoustic modeling with parallel computation in recent years (e.g., [17]–[20]). However, when considering large room geometries, such as concert halls or places of worship, system memory constraints and total computation time currently make this an impractical task for a single modeling method.

To alleviate this issue previous work has suggested using combinations of different acoustic modeling methods. These HAM approaches are reported to offer a more computationally efficient solution without compromising acoustic accuracy. Combinations of image-source and ray-tracing methods have been proposed for calculating RIRs, with [21] presenting a combination that identified valid image sources from the exponentially increasing number of possible candidates using ray-tracing that recorded the reflection history of each ray. This approach, as the author noted, was however restricted to specular reflection and neglected diffraction and low frequency wave phenomena entirely. In [22] an image-source and ray-tracing hybrid model is presented. This differed from that of [21] as the exponentially growing number of image-source visibility checks was not directly reduced. Instead the early and late reflections are modeled by the image-source and ray-tracing methods respectively, where, in the latter case, rays interacting with a surface are treated as radiating secondary sources in order to synthesize the build up of the late reverberant field at the receiver point.

A HAM based on a combination of ray-tracing for early reflections and an energy transition method similar to the radiosity approach for late reflections was presented in [23]. Energy transition methods make the assumption that the soundfield is diffuse, subsequently removing the need to calculate specular reflection angles. Instead, the energy at the source is transferred to those visible surfaces, is appropriately attenuated by each surface and then distributed to all other visible surfaces iteratively. The visibility calculation is only performed once for each surface and stored, resulting in a relatively fast computation of the late reverberation when compared to specular ray-tracing and image-source methods.

Other approaches to hybrid acoustic modeling attempt to combine some geometrical approach with a statistically derived late reverberation stage, such as [24], where the image-source method (ISM) is used both for early reflections, and to derive the energy decay curve used to synthesize the late part of a RIR. However, perhaps some of the most promising recent work is based on a hybridization of wave and geometric approaches, as first presented in [25] and [9]. Although wave based methods are still highly demanding in terms of computation for a result that is valid across the whole audible spectrum, results in the low frequency region can now be computed at reasonable speed [17], and offer improvements over geometric methods, particularly in terms of accurate modeling of diffraction and occlusion effects.

In [25] the Finite Element Method (FEM) was used up to 350 Hz and combined with a geometric acoustic algorithm. FEM (up to 300 Hz) was compared with the ISM in [26], but the first more comprehensive hybrid method was considered in [27] with the FEM being used for low frequencies combined with image-source and stochastic ray-tracing (including scattering) at high frequencies, with the crossover being determined by the Schroeder frequency of the space being considered. Noting that appropriate care was required in setting simulation parameters across the methods used, a preliminary listening test revealed that subjects were unable to identify artefacts in audio examples based on the hybridized impulse responses obtained.

FEM was combined with the FDTD method in [28] with a view to modeling frequency dependent reflection and absorption at a surface, as might be used in architectural acoustics and auralization applications. However, it was in [29] that a more comprehensive hybrid solution was proposed, comprising a computationally intensive wave-based 3-D digital waveguide mesh (DWM) combined with a simple specular ray-tracer for early reflections and a less computationally intensive 2-D DWM with corrected reverberation time [30] model for late reflections. The simulation strategy adopted could be optimized according to the computational resources available, and the required maximum estimated time for calculating a RIR, with the 3-D DWM being favoured over 2-D DWM wherever practical. Synthesized and measured RIRs were obtained and compared for a small room of 27 m^3 and exhibited good agreement, especially when comparing the low frequency response of an established purely geometric model. Reverberation time measurements obtained from synthesized RIRs and averaged over receiver position gave a good match to those similarly obtained from measurements and established methods. However, some variation was observed when considering non-averaged results at specific positions.

The proposed hybrid modeling framework is closest to that presented in [27] although the utilized techniques are different in each frequency region. The approach presented in this paper utilizes FDTD modeling at low frequencies while high frequencies are simulated using a combination of beam-tracing and acoustic radiance transfer. All of these techniques are accurate and efficient in the frequency region over which they are intended to operate making the proposed hybrid technique more accurate than any of the single methods if applied for the whole response. There are also close similarities to the approach presented in [29], although late reverberation is approximated using a 2D FDTD method in this work rather than obtained more directly using an appropriate model (such as BT or ART), and a direct comparison is presented in the results which follow. However, the main emphasis of this paper is in the description of how the results from different simulation techniques should be calibrated such that full-bandwidth impulse responses can be formed. Similarly, treatment of boundary conditions in each modeling technique is described in detail. These two areas are crucial in making hybrid room acoustic models practical, and in delivering results that are accurate both within the limitations of the parameters supplied (e.g., boundary conditions), and across the applicable frequency region.

III. Defining the Hybrid Acoustic Model

The approach taken in this paper combines 3-D FDTD for low frequency modeling with Beam Tracing (BT) for low-order reflections combined with the Acoustic Radiance Transfer (ART) method for the late reflection stage, and can be seen to offer a more complete and optimized solution over those methods previously presented. This section introduces each method in turn with reference to Fig. 2 and demonstrates their use in the proposed hybridization scheme.

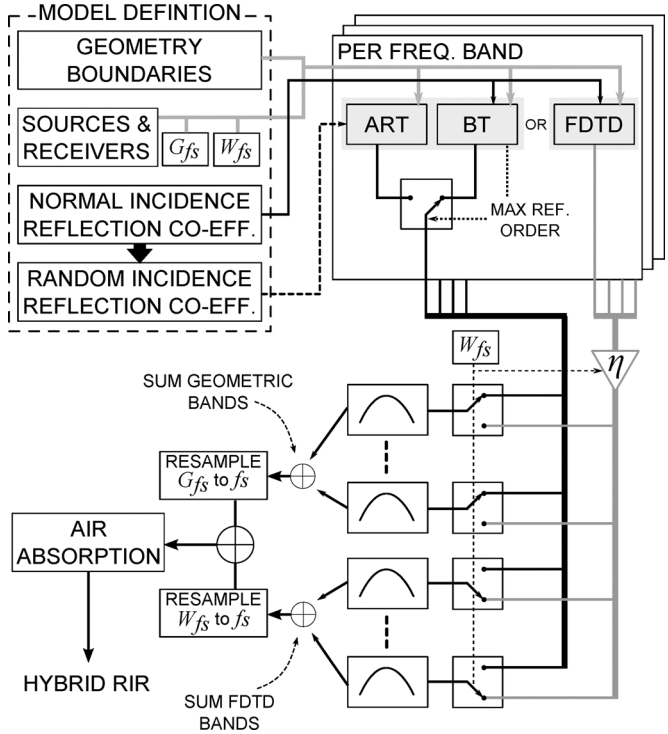


Fig. 2. An overview of the hybrid acoustic model and the supporting processing network proposed in this work. The modeled responses of beam tracing, acoustic radiance transfer and the FDTD methods are combined in a final band pass filter-bank. Ideally all bands should be modeled with the FDTD method, but the chosen sampling frequency W_{f_s} imposes an upper frequency limit that determines when the geometric methods should be used instead. The calibration parameter η correctly balances the energy between the modeling methods. In this work a method of determining η based on the FDTD sampling frequency W_{f_s} is presented. G_{f_s} is the sampling rate of the synthesized RIR output from the BT and ART models.

A. Finite Difference Time Domain Method

This method is well established for acoustic modeling of bounded room geometries. The volume of air enclosed by the geometric description of the space is represented by a 2-D/3-D grid of $\{i\ j\}/\{i\ j\ k\}$ indexed points or *nodes*. The metric spacing between the nodes is related to the sampling rate of the grid W_{f_s} by (1):

$$d_x = \frac{c}{\lambda W_{f_s}} = \frac{c \cdot \sqrt{D}}{W_{f_s}} \tag{1}$$

where d_x is the distance in metres between two nodes along the axial directions. Parameter $\lambda = \frac{cT}{X}$ is the Courant number and X is the unit grid spacing, T the unit time step and $c = 344$ m/s is the speed of sound. In this work the Courant number is set as $\lambda = 1/\sqrt{D}$ where $D = 3$ is the dimensionality of the grid used in this work.

This implementation is formulated by representing the wave equation using a second-order central finite difference approximation of its temporal and spatial derivatives. The 3-D standard rectilinear (SRL) formulation is used in this work (2) after [14]:

$$p^{n+1}_{i,j,k} = \lambda^2 \left(p^n_{i+1,j,k} + p^n_{i-1,j,k} + p^n_{i,j+1,k} + p^n_{i,j-1,k} + p^n_{i,j,k+1} + p^n_{i,j,k-1}\right) - p^{n-1}_{i,j,k} \tag{2}$$

where (2) is the 3-D update equation for calculating the acoustic pressure in free space at grid position $\{i\ j\ k\}$ at time index $n+1$. The update equations for boundary nodes are formulated by first defining a discrete first-order finite difference approximation of the appropriate continuous boundary definitions as shown in (3):

$$p^n_{i+1,j,k} = p^n_{i-1,j,k} + \frac{1}{\lambda \xi_w}\left(p^{n-1}_{i,j,k} + p^{n+1}_{i,j,k}\right) \tag{3}$$

where (3) is the discrete boundary model for a right hand boundary and parameter ξ_w is the specific wall impedance which is discussed in further detail in the remainder of the section. By way of example, the update equation for a right-hand surface boundary node is given in (4) and formed by substituting (3) into (2):

$$p^{n+1}_{i,j,k} = \Big[\lambda^2\left(2p^n_{i-1,j,k} + p^n_{i,j+1,k} + p^n_{i,j-1,k} + p^n_{i,j,k+1} + p^n_{i,j,k-1}\right) + \left(\frac{1}{\sqrt{3}\xi_w} - 1\right)p^{n-1}_{i,j,k}\Big] \Big/ \left(\frac{1}{\sqrt{3}\xi_w} + 1\right) \tag{4}$$

It is inferred from (4) that for every node position two acoustic pressure values must be stored, the current value and the previous value; the new value, $p^{n+1}_{i,j,k}$ simply overwrites the previous value, $p^{n-1}_{i,j,k}$. The specific wall impedance ξ_w is related to the plane wave reflection coefficient by (5) after [6]:

$$\hat{r}_{FD}(\theta_w, \phi_w) = \frac{\xi_w \cos\theta_w \cos\phi_w - 1}{\xi_w \cos\theta_w \cos\phi_w + 1} \tag{5}$$

where θ_w and ϕ_w are the angles of azimuth and elevation for a plane wave incident on the boundary surface. As these parameters are not explicitly available in the FDTD method ξ_w is approximated by always assuming normal incidence, $\theta_w = 0^\circ$ and $\phi_w = 0^\circ$ as in (6):

$$\xi_w = \frac{1 + r(0^\circ, 0^\circ)}{1 - r(0^\circ, 0^\circ)} \tag{6}$$

where r is the desired reflection coefficient of the boundary. The boundary update properties may also be adapted further to incorporate frequency dependent characteristics to represent any given surface type. In this case (4) is modified appropriately to include a digital impedance filter [14].

In the simplest implementation an omnidirectional sound source is approximated by setting non-zero initial conditions at each required source node. Directional sound sources can be implemented by using multiple excitation points whose individual radiated soundfields superimpose to reinforce and cancel over angular regions [31], [32]. The acoustic pressure value at each node at each time step is then calculated on a sequential node-by-node basis. The RIRs are synthesized by storing the pressure value at each time step at the nodes nearest to the desired receiver locations.

Dispersion error occurs because the discrete grid of nodes only supports sound propagation along certain paths. In the SRL formulation, propagation between nodes occurs only in the axial directions, as is implied by e.g., (4). The result of this restriction is that wavespeed is both frequency and angularly dependent. A detailed treatment of dispersion error as well as approaches to reducing its effects have been previously presented [14], [33]. The SRL scheme is chosen in this work due to its low memory requirements and the availability of the re-entrant corner definition that allows complex concave geometries to be modeled that will inherently include wave diffraction effects.

B. Beam Tracing

The beam tracing (BT) algorithm introduced in [34] is utilized in this work. The BT approach is an optimized implementation of the ISM for modeling convex/concave room geometries. The relative efficiency of BT is due to the pre-computation of a visibility structure called a *beam tree*. A beam tree is a hierarchical arrangement of *beams* and, beams are cone volumes with convex polygon bases through which a valid specular reflection path may lie. A list of valid reflection paths found using the beam tree form a reflection echogram. For each reflection the angle of incidence is known and used to weight the propagated energy using the reflection coefficient, $\hat{r}_{BT}$ in (7),

$$\hat{r}_{BT}(\theta_r, \phi_r) = \frac{\xi_w \cos\theta_r \cos\phi - 1}{\xi_w \cos\theta_r \cos\phi_r + 1} \tag{7}$$

where θ_r and ϕ_r are the angles of azimuth and elevation angles for a ray incident on a boundary surface. These parameters are explicitly available in the BT method and ξ_w is calculated as in (6) for the FDTD case. Note that this is demonstrated in Section V-A and this choice ensures that the FDTD and BT acoustic models are using the same analytical boundary model.

C. Acoustic Radiance Transfer

The theoretical foundation of the ART method was presented in the room acoustics rendering equation, relating the acoustic time-dependent energy, *radiance*, at every surface point in a room [5]. In practice, a finite number of surface *patches* are used to represent the infinite number of points and radiance is assumed to be invariant over a patch. Directionally dependent reflection properties of a patch are encapsulated by a *bi-directional reflectance distribution function* (BRDF). This being a definition of the portion of the energy that is reflected in each outgoing direction when excited from a given incoming directional slot, a finite angle. The portion of energy leaving one patch and arriving at another through the attributed BRDF, namely the *directional form factor*. For N patches and k directional slots, there will be a $Nk \times Nk$ form factor matrix describing all the relationships between the patches. Most of the matrix elements are zero because of the directional nature of the radiance and occlusion. In the first phase of ART, the radiance arriving at each patch from the sound source is calculated accounting for attenuation by distance, any occlusion and source directivity. After applying the BRDF for the reflection energy calculation, the outgoing radiance is stored for each directional slot on each patch. The energy is propagated and reflected and the outgoing energies are stored, until the propagated energy becomes negligible due to attenuation.

Finally, the radiance is collected from all visible patches to the receivers, and, similarly to the initial shooting, attenuation by distance, occlusion and directional properties of the receivers are taken into account. This phase may be repeated for several receiver positions without redoing the previous phases if the responses on the patches are stored in memory.

D. Combining FDTD, BT and ART Methods

The hybrid model presented here is founded on the FDTD method, and in theory this would be used to synthesize the entire impulse response as it provides a complete physical model based on a second-order solution to the 3-D wave equation. However, for room acoustic modeling, either system memory use or total simulation time will cause the FDTD approach to become impractical for high frequencies and a complete solution is therefore obtained through the combination of FDTD, BT and ART. The BT method is therefore employed in the high frequency region, where the FDTD method is no longer a realistic proposition, and where it approximates well the exact solution to the wave equation when the wavelength considered is much smaller than the room's geometric detail. Under certain conditions, such as in a shoebox room case where there are no diffracting edges, the BT method, as an optimized implementation of the ISM, becomes exact (this is evident in Section V-B). When the BT approach becomes impractical in terms of computation time due to high reflection orders the time-domain response continues to be calculated based on the statistically based ART method, which therefore completes the modeling of the high frequency region. Similarly, the computation is switched from BT to ART when the BT approach comes unsuitable due to diffuse reflections.

The approach presented here for combining the BT and ART results is similar to the one in [35]. To clarify, the BT is employed for the early part of the response up to a particular order and this limit is determined by the memory consumption of the beam-tree. It is desirable to use the BT model over the ART model as much as possible as it is more physically correct when modeling specularly reflected energy. Furthermore, the temporal mixing region between BT and ART depends on the geometry and its surfaces diffusion coefficients. Specifically, the mixing region is reflection path dependent, i.e., when the specular energy becomes negligible, meaning the diffuse energy is significant, the BT response no longer contributes to the final response even if the path has been calculated. The BT reflection paths are weighted by $\prod_i (1-\alpha_i)(1-w_i)$, where α_is and w_i correspond to the absorption and diffusion coefficients, respectively, at the reflection points along the path. The diffuse energy values are stored at each patch at each time step, and are used in the ART method as an initial state. At the last beam tree level, the remaining energy is also stored at the patches.

Beam traced reflection paths are evaluated and transformed into an echogram for each required receiver position. The arrival directions of the last segment of each reflection path is used to weight the reflection energy with a microphone directivity pattern, for spatially encoded RIRs [10], [16]. The ART portion of the receiver responses are obtained by gathering the radiance responses from all visible patches. Similar spatial encoding can be implemented by using the directions of the patch centroids as parameters for the directional weighting functions of the virtual microphones. Since the sampling frequency in the acoustic radiance transfer method is typically lower than the final audio sampling frequency, the gathered responses must also be upsampled. This involves jittering the reflection arrival times to avoid ringing at the ART sampling rate. The synthesized echogram combined from both the BT and ART models is converted from a list of reflection arrival times, with an associated energy quantity, to a pressure signal using the square root operation and has the sampling rate G_{f_s}.

Frequency dependent responses are obtained by performing BT and ART as described but using frequency band defined absorption, diffusion and BRDF coefficients at the boundaries. The responses from the BT and ART are then added together for each frequency band with appropriate band-pass filtering as shown in Fig. 2, thereby enabling the combination of these geometric methods with the FDTD technique.

E. Air Absorption

In the final post-calibration (See Section IV) stage of the hybrid model, the combined RIRs are processed to take account of air absorption effects. Based on an overlap-add convolution implementation of an analytical expression used to predict air absorption [36], the absorption weighting is changed as a function of the current time frame as this directly implies the total distance traveled.

IV. Impulse Response Calibration

A primary consideration in any hybrid approach is that all RIRs obtained across modeling domains are calibrated appropriately before being combined to give the correct relative energy ratio. Little work has been presented on how such RIRs might be optimally combined. In [27] FEM and geometric methods are normalized relative to the sound pressure calculated using each method for a receiver 1m from the source in the free-field. The resulting calibration is specific to the FEM model and therefore not applicable to the HAM presented in this article.

In [10] two methods are considered. The first is based on the ratio of the average magnitudes obtained over a defined frequency region for each RIR. Although straightforward, this is always an approximation as it does not take into account physical properties in the modeled space that may affect the RIRs obtained, which are otherwise assumed to be equivalent. The second method is an improvement, again based on a calibration of sound sources in each modeling domain at a distance of 1 m. This is trivial in a geometric model assuming there is a direct line of sight between source and receiver. The process is more involved for an FDTD model as the obtained RIR is typically not full bandwidth and the simulated wavefront, as with a real wavefront, is spread in time, an effect which is most apparent when there is a lack of high frequency energy (as is typically the case in an FDTD simulation). The proposed solution is to derive a calibration parameter based on the peak level measured from time-windowed, bandwidth matched RIRs obtained from both FDTD and geometric methods. The position and length of the window used are determined from the direct and first reflection arrival times calculated from the geometric model. Peak values are used, rather than total energy, as the window limits derived from the geometrical model may miss a diffracting edge reflection that arrives before the first reflection. Although this method was preferred in [10], it is dependent on successfully obtaining an accurate time window. This time window will in turn depend

on the given modeling scenario and the accuracy and equivalence of the boundary implementation that gives rise to the first reflection. A novel method which removes any such conditions is presented in the following and is a general solution that is independent of the room geometry.

A. Defining the Calibration Parameter

An appropriate calibration parameter can be determined as the square root of the ratio between the total energy received at a given point at an arbitrary reference radius, r, from the source and, the total energy of the source as in (8):

$$\eta = \sqrt{\frac{E_{ref}}{E}} \tag{8}$$

where η is the desired calibration gain parameter, E is the energy at r, and E_{ref} is the reference energy value taken as the total energy of a freefield point source in a lossless medium. The calibration will be performed using the same FDTD method as required in the hybrid model. For a source signal consisting of a unit pulse, $E_{ref} = 1$ and the total energy passing through the receiver point at the reference distance r is therefore:

$$E = \sum_{t=0}^{T} p[t]^2 \tag{9}$$

where T is the length of the received acoustic pressure signal p. In the context of the FDTD method used, calculating the calibration parameter using a unit pulse as the reference source is problematic if the freefield condition is to be strictly preserved. This is because in practice some energy is reflected from even the most highly absorbing boundary [14] that might be used to terminate the FDTD grid in order to simulate freefield conditions. As a consequence, undesired reflections can arrive at the receiving position before the direct sound source energy has attenuated to a negligible amplitude. This is alleviated by band-limiting the source signal which ensures that dispersion error is minimized and that the direct sound wave is approximately isotropic and has better defined onset and offset regions. The reference energy quantity, E_{ref} in (8) is defined as in (10):

$$E_{ref} = \sum_{t_s=0}^{T_s} s[t_s]^2 \tag{10}$$

where T_s and t_s are the total length of the source signal and time index respectively and s is the source signal. Note that Section IV-A1 discusses and determines in detail the choice of an acceptable band-limit required to make the source propagation dispersion free and approximately isotropic.

Under ideal circumstances the inter-nodal distance would be infinitely small so that the wave propagation, and related acoustic pressure, for an infinite number of nodes is calculated over a continuous region. Instead, for computation, the inter-nodal distance must be finite, and as such the calculated acoustic pressure at each node represents a finite region of d_x^3 in 3-D. Therefore the number of nodes representing the same volume of air will vary depending upon the chosen sampling rate W_{f_s} according to (1). Consequently, the number of nodes between source and receiver will also vary with W_{f_s}, and therefore the calibration parameter will depend upon W_{f_s}, the chosen FDTD stencil and the reference distance r. The source strength in the geometric model must be set so that at distance r the energy is E_{ref} for the calibration parameter to be valid. Once the calibration parameter is known for a given FDTD stencil (3D SRL in this case), sampling rate and reference distance, the linear nature of the FDTD system allows the gain adjustment to be made at simulation time by setting the amplitude of the unit pulse to η or, equivalently, by multiplying the output samples at every receiver position by η.

In the following section, the application of (8), (9) and (10) for computing η for a range of W_{f_s} is demonstrated, leading to a generalized definition of η, for an arbitrary r, based on a linear fitting to the results obtained. This therefore negates the need to perform further pre-simulations for such calibration of the SRL scheme in future work. The calibration parameter is demonstrated for settings the amplitude of an FDTD source signal to a desired absolute sound pressure level at a given distance. The use of the calibration parameter for correctly balancing the synthesized signals from FDTD and geometric models is demonstrated in the test cases presented in Sections V.

1) Obtaining the FDTD SRL Calibration Parameter: The formulation and validation of a generalized FDTD SRL calibration parameter is performed using a number of simulations for sampling rates $W_{f_s} = 13$–42 kHz in intervals of 1 kHz. The excitation signal is a band-limited source defined as the coefficients of an FIR low pass filter. In practice the excitation signal is a unit pulse and the linear nature of the FDTD model allows the FIR filtering to be applied on the synthesized receiver signals as a post process. An arc of receivers over 90° are positioned at $r = 1$ meter from the source, so that one end of the arc is directly above the source and the other creates a diagonal propagation path with respect to the horizontal plane of the FDTD grid. The receiver locations are defined across 91 ideal receiver locations at 1° intervals, which are then quantized to the nearest node position such that any receiver positions duplicated as a result are made unique. The number of receivers used in each simulation varies with W_{f_s} as the defined arc of receivers is better approximated at higher sampling rates.

Firstly it is necessary to justify the choice of the low-pass filter cut-off frequency. The direct sound is captured by the arc of receivers for each given W_{f_s}. Each receiver is then low-pass filtered with the normalized cut-off frequency ranging from 0.01 to 0.25 in increments of 0.01. The calibration parameter is computed for each source-receiver combination and across all sampling rates and cut-off frequencies and then used to adjust the gain of the response obtained at the corresponding receiver. The difference between the pass-band magnitude response of the reference source and the newly calibrated receiver response, for each source-receiver combination, is then quantified in terms of the maximum error in dB. In Fig. 3 the maximum error over the set of maximum errors is shown as a function of both sampling rate and cut-off frequencies. It can be observed that the maximum error is $\ll$3 dB JND when the cut-off frequency of the low-pass filter is chosen below approximately $0.15W_{f_s}$. Therefore, in what follows a normalized low-pass filter cut-off frequency of $0.01W_{f_s}$ has been used.

Based on this result, the energy E is computed for each receiver and the mean value used in (8). The value of η obtained is

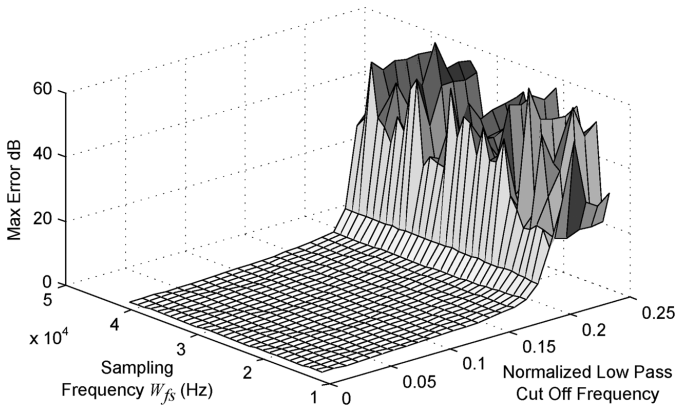


Fig. 3. The maximum dB error over the set of maximum dB errors between the pass-band magnitude response of the reference source and the calibrated receiver response, for each source-receiver combination, as a function of both sampling frequency W_{f_s} and normalized low-pass filter cut-off frequency. The maximum error is never greater than 3 dB for a cut-off value of less than approximately 0.15.

TABLE I
THE MEASURED VALUES OF η AND THOSE VALUES BASED ON PREDICTION, η_{SRL-3D}, WITH MEAN AND MAXIMUM ERROR VALUES CALCULATED BY COMPARING THE BAND-PASS MAGNITUDE RESPONSE OF THE REFERENCE SOURCE SIGNAL AND EACH CALIBRATED RECEIVER SIGNAL

	kHz							
	13	14	15	16	17	18	30	42
Measured η	69	75	81	87	93	94	159	225
Mean dB Error	0.36	0.25	0.19	0.09	0.06	0.15	0.02	0.02
Max dB Error	0.36	0.25	0.20	0.10	0.07	0.15	0.02	0.03
η_{SRL-3D}	67	72	78	83	89	94	159	225
Mean dB Error	0.05	0.07	0.16	0.25	0.32	0.10	0.04	0.03
Max dB Error	0.05	0.07	0.17	0.25	0.33	0.10	0.04	0.04

then multiplied with each receiver signal and the error between the reference source signal energy and calibrated receiver signal energy is given in Table I(a). The mean dB error describes the difference between the pass-band magnitude response of the reference source signal and the calibrated receiver signal based on this obtained value of η. Note that in Table I(a) these values are provided for η as calculated using (8), (9) and (10), based on the measurements obtained from the FDTD SRL simulation. For lower values of W_{f_s} the effect of spatial discretization on the desired receiver points becomes more critical and so by considering only measured values of η for $W_{f_s} \geq 18$ kHz it is possible to predict η based on a linear fit in (11). Note that as the calibration parameter is a pressure quantity it will vary by the factor $1/r$ which is accounted for by the multiplying factor r in (11).

$$\eta_{SRL-3D} = r(5.437 \times 10^{-3} W_{f_s} - 3.6347) \qquad (11)$$

where η_{SRL-3D} is the required calibration parameter in a 3-D FDTD SRL scheme for any given sampling rate W_{f_s} when the geometric model source is set as $E = 1$ at $r = 1$ m. The definition of (11) can be rationalized by first considering that the number of nodes defining a given distance from a source varies linearly with the sampling rate according to (1). Therefore, as the total energy at a receiver is generally dependent upon the number of nodes between it and the source, it is reasonable to assume that η also varies linearly with sample rate W_{f_s}. In Table I(b) the mean/max dB error again describes the difference between the pass-band magnitude response of the reference source signal and the calibrated receiver signal, this time based on η_{SRL-3D}, and can still be considered negligible. To further quantify the accuracy of the fit in (11), it is appropriate to calculate the maximum difference in dB between a signal calibrated with the ideally measured η and η_{SRL-3D}. For $W_{f_s} = 13$–42 kHz the maximum difference is $\max\,|20\log_{10}(\eta/\eta_{SRL-3D})| = 0.38$ dB. Furthermore, when $W_{f_s} = 18 - 42$ kHz it is 0.05 dB. This also indicates the necessity to limit the fitting region to $W_{f_s} \geq 18$ kHz. Note that these calibration parameters can be derived directly and explored in more detail via the Matlab script that accompanies this article [38].

2) Setting Arbitrary Source Strength: Using the general definition of the calibration parameter in (11) the absolute strength of a sound source sound pressure level may be set for an arbitrarily chosen reference distance r.

$$\nu = \eta(W_{f_s}, r) \cdot 10^{SPL/20}/S_{ref} \qquad (12)$$

where ν is set as the amplitude of the Dirac excitation in the FDTD model or alternatively multiplied with all receiver samples for a unit Dirac input. The former approach offers the flexibility to set multiple sources in the same simulation pass to exhibit different absolute source strengths at different distances. SPL is the desired sound pressure level at r and S_{ref} is a pressure quantity obtained from the anechoic source signal to be convolved.

V. TESTING AND VALIDATION OF HYBRID ACOUSTIC MODEL

It is important for any proposed hybrid modeling technique to ensure that the performance of the implemented boundary models is consistent. That is, setting appropriate boundary parameters for each model should produce comparable results. This section rigorously investigates the relationship between the observed FDTD and BT reflection coefficients, $\check{r}_{FD}$ and $\check{r}_{BT}$, compared to the desired response $\hat{r}_{FD}$ and $\hat{r}_{BT}$ defined in (5) and (7) respectively. In addition, the models are compared using a number of common room acoustic modeling scenarios and metrics and in each case the effective use of the calibration parameter derived in Section IV-A is demonstrated.

A. Reflection From a Single Surface

In the BT case the analytically obtained and observed reflection coefficients $\hat{r}_{BT}$ and $\check{r}_{BT}$ respectively are identical such that $\hat{r}_{BT}(\theta_r, \phi_r) = \check{r}_{BT}(\theta_r, \phi_r)$ for arbitrary r as the reflection model in (7) is implemented directly. However, as explained in Section III-A this is not so trivial in the case of the FDTD boundary (4).

The actual observed reflection coefficient, $\check{r}_{FD}$, is measured by applying the technique in [37] which proposed a measurement system consisting of a single microphone and loudspeaker mounted at a fixed distance (1.2 m or 1.5 m) from each other and the test surface. The microphone is positioned closest to the test surface and the direct sound is canceled by subtracting the freefield response of the loudspeaker at the microphone from the obtained measurement. The system can be varied according to angle of sound incidence. In this work, the technique was adapted for an acoustic model by taking advantage of acoustic reciprocity to reduce the reflection coefficient analysis to a

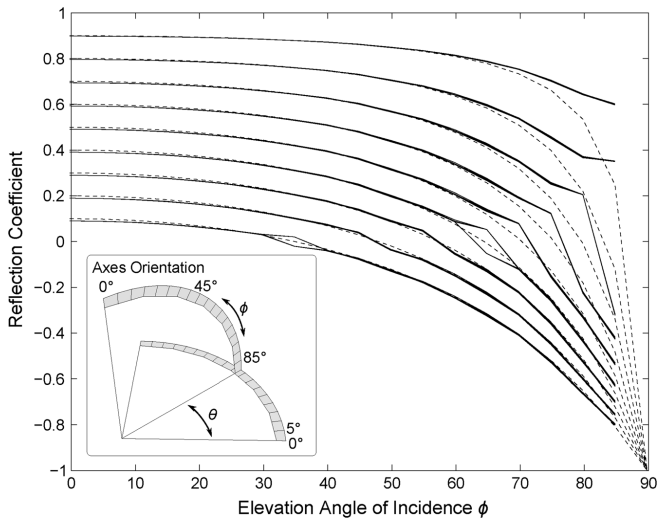


Fig. 4. A comparison of the actual observed broadband reflection coefficient $\check{r}_{FD}$ with the desired analytical case $\hat{r}_{FD}$ also implemented in the BT model $\hat{r}_{BT}$ for a range of r with absolute values from 0.1 to 0.9. Note that there is no variation in terms of the measurements obtained with azimuthal angle indicating that $\check{r}_{FD}$ only varies with the elevation angle of sound incidence.

single simulation pass for a complete dataset of incident angles. Three datasets, A, B and C are synthesized, each using a one quarter hemispherical array of 342 receivers and a single source. The centre of the quarter-hemisphere was taken as the image-source position for A, B and C in order to ensure a consistent specular path length of 66 nodes between each receiver and the source. The measured datasets were as follows: (A) freefield response to the source position; (B) freefield response to the image-source position; (C) a source located at $3 \cdot d_x$ from the boundary. The responses in A were subtracted from C to create a new dataset D, the reflection function dataset. The energy (taken as the sum of the squared samples) for each response in B and D were calculated and the square root of their ratio is the observed broadband reflection coefficient $\check{r}_{FD}$.

Fig. 4 compares the actual observed FDTD reflection coefficient as a function of angle of incidence/reflection in the FDTD model (solid lines) with that implied by the analytic reflection model (dotted line) by substituting (6) back into (5) for a range of r from 0.1 to 0.9. Note that $\phi_w = 0°$ is the normal to the reflecting surface being considered. For a given angle of elevation ϕ_w there is no variation in $\check{r}_{FD}$ with angle of azimuth θ_w and hence Fig. 4 shows the result varying with incident angle of elevation only. As can be seen, there is very good agreement between $\check{r}_{FD}$ and $\hat{r}_{BT}/\hat{r}_{FD}$. Divergence from the analytical case is noted as the reflection coefficient changes from negative to positive and also at grazing angles of incidence which is due to limitations in the measurement technique rather than a function of the FDTD boundary conditions.

B. Modal Analysis in a Cuboid Room

The ability of a room acoustic model to accurately simulate cyclic propagation paths leading to a frequency resonance is an important characteristic of any HAM. This section investigates the ISM (representing BT) and 3-D FDTD method and the boundary model used in Section V-A in the context of a modal analysis. In addition, validation of changing reflection coefficient values is further confirmed also using the same modal analysis test for a cuboid room of dimensions 5.56 m $\times$ 3.97 m $\times$ 2.81 m. Results are generated for a 3-D FDTD model with $W_{f_s} = 18$ kHz, using 5 sources, (S1–S5) and 5 receivers (R1–R5), given in Table II, resulting in 25 source-receiver combinations listed in Table II. Note that all 3-D FDTD receiver samples are multiplied with the calibration parameter from (11). Reflection coefficients r range across 21 values from $-0.999, -0.9, -0.8, \ldots, -0.1, 0.0, 0.1, \ldots, 0.9, 0.999$. This simulation setup is repeated using a standard image-source model, enabling modal responses to be obtained for both. The results are analysed up to 120 Hz, which ensures that maximum dispersion error in the FDTD simulation at a given frequency is never greater than 1%, and are presented in Fig. 5 for the S5-R4 combination. Fig. 5(a) demonstrates positive reflection coefficient values of 0.999, 0.9, 0.8, and Fig. 5(b) shows negative reflection coefficient values of $-0.999, -0.9, -0.8$. Overlaid on these plots are the analytical modes expected for a room given these dimensions.

Note first of all that the results presented for the FDTD model are based on an ideal transparent source Dirac excitation, and hence the ripple observed around DC is as expected given that there has been no pre- or post-filtering of the RIRs obtained. There is very good agreement between the analytical mode positions and the magnitude response of the RIRs and this holds for both the positive and negative reflection coefficient case. In the latter example, negative reflection coefficients can only occur under certain physical conditions, and this is evident in Fig. 5(b). Axial modes are not supported due to the requirement for a pressure maximum value at the boundary (a negative reflection coefficient implies a change in phase at the boundary and so a pressure maximum cannot be supported). Hence the analytical modes presented are only for the tangential and oblique cases and the observed magnitude response gives excellent agreement. It should also be noted that the magnitude responses have not been normalized in anyway for the purpose of the modal analysis, the close agreement between the ISM and 3-D FDTD is completely as a result of applying the proposed calibration parameter in Section IV.

C. Joined Cuboids Model

In this example a slightly more complex room geometry is tested through the introduction of two diffracting edges formed by partially superimposing two cuboids, as shown in Fig. 6. Source and receiver positions are defined such that there is no direct line-of-sight path between them. This scenario is simulated using FDTD with $W_{f_s} = 20$ kHz, and BT at 48 kHz, with a uniform reflection coefficient across all boundaries of 0.999. As an analytical solution is no longer trivial to derive for this example, the results obtained are also compared with a Finite Element Method equivalent based on [8].

Firstly, the influence of the diffracting edges is considered, and the limitations of the BT method to correctly account for the additional reflections that arise as a consequence are highlighted and presented in Fig. 7. The BT method correctly predicts the first reflection components arriving at the receiver after specular reflection with one of the room boundaries. In the FDTD case,

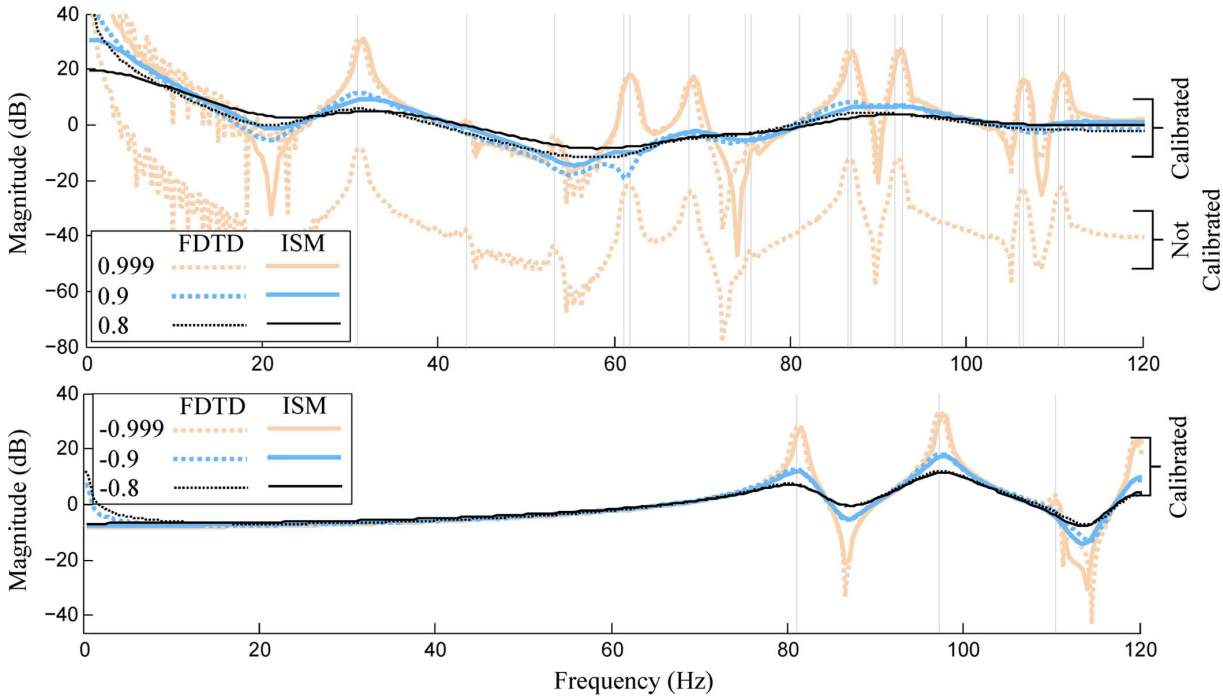
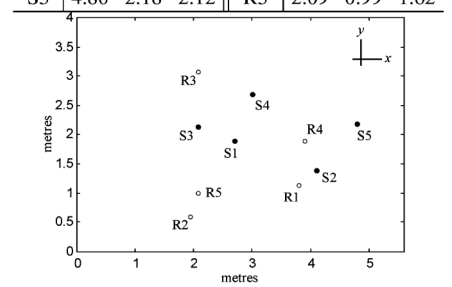


Fig. 5. Color online: The S5-R4 combination magnitude responses for calibrated FDTD (dotted) and image-source (solid) methods up to 120 Hz for a cuboid room of dimensions 5.56 m × 3.97 m × 2.81 m. (a) Positive reflection coefficient values of 0.999, 0.9, 0.8; (b) negative reflection coefficient values of $-0.999, -0.9, -0.8$. Analytical modal frequencies for a room of the same dimensions have been overlaid as vertical lines, with (b) plotting only the oblique cases. Note that without the presented calibration applied to the FDTD results, the absolute difference in magnitude is significant between the FDTD and ISM.

TABLE II
THE 5 RANDOM SOURCE AND 5 RECEIVER POSITIONS USED FOR THE CUBOID ANALYSIS. NOTE THAT SPATIAL LOCATIONS HAVE BEEN SELECTED RANDOMLY SUCH THAT NO SOURCE OR RECEIVER IS LESS THAN 0.5 M FROM A BOUNDARY

Src.	x	y	z	Rec.	x	y	z
S1	2.71	1.89	1.69	R1	3.81	1.13	1.92
S2	4.10	1.39	1.82	R2	1.95	0.60	2.28
S3	2.09	2.12	2.12	R3	2.09	3.08	0.96
S4	3.01	2.68	0.93	R4	3.91	1.89	1.69
S5	4.80	2.18	2.12	R5	2.09	0.99	1.62



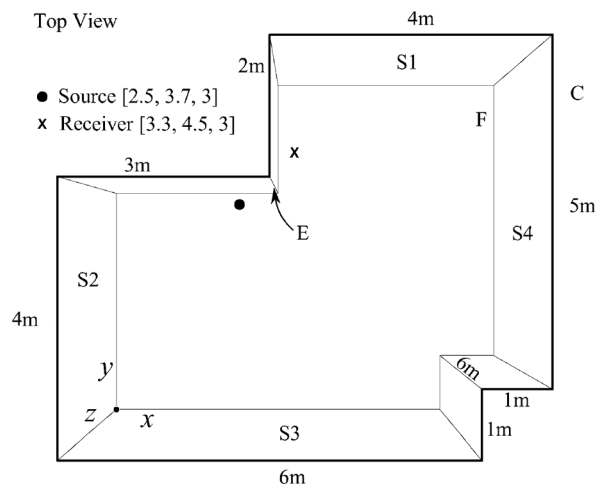


Fig. 6. Joined cuboids model designed to test the effect of diffracting edges. The source and receiver positions are defined such that the edge (E) obscures the direct line-of-sight. S1 to S4 are labels for the surfaces that enclose them, and F and C denote the floor and ceiling respectively and are referred to in Fig. 7.

the obtained RIR is first post-processed to remove DC and low-pass filtered at $0.15 \times W_{f_s}$ to account for dispersion error [39]. The result, also shown in Fig. 7, reveals significant components arriving before the early reflections detected by the BT method. These components can be accounted for by the two edges in the joined cuboids effectively adding in the direct sound and pre-first reflection paths via diffraction. Although more difficult to identify, there is also good agreement between the BT detected early reflections and those that occur later in the FDTD RIR.

This relatively simple case therefore reveals the limitations in accuracy of the BT method under certain geometrical conditions. The FDTD model, however, is still capable of giving accurate results. The modal responses for the FDTD and FEM cases are presented in Fig. 8. Although there is some difference in relative magnitude, the results are a very good match, especially given that the FEM model used adopts a simplified approach to calculating the full modal series [8]. Hence FDTD is valid for more complex spaces where the BT method falls down due to diffracting edges, or excessive computation time limits the reflection order that can be calculated leading to a truncated RIR. With this in mind the ART method provides a suitable alternative for late reflections and this method will be tested in the next case study.

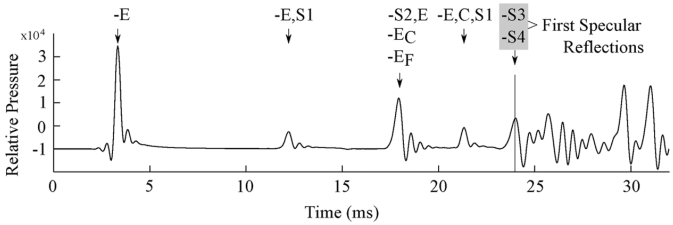


Fig. 7. The obtained RIR from the FDTD simulation of the joined cuboids model. The time of arrival of the first two specular reflections, as identified by BT, are marked and correspond well to the FDTD simulation. The labels on each peak indicate which surfaces or edges have formed the reflection. This may be explored in more detail with the accompanying FDTD GPU accelerated MATLAB script for this model [38].

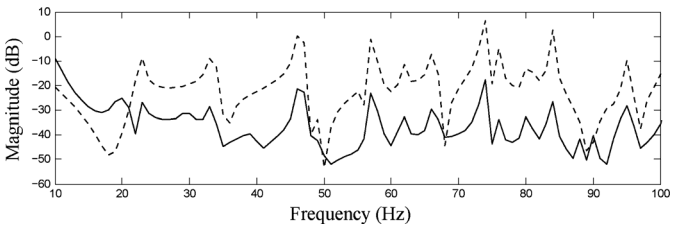


Fig. 8. The FEM (dashed) and FDTD (solid) based magnitude responses for identical source/receiver locations in the joined cuboids model. The introduction of the two diffracting edges creates inconsistent response predictions. Note that the relative difference in the overall response level is due the limited ability to adjust the source strength in the software accompanying [8]. This is not a limitation of the FEM or FDTD methods.

D. Late Reverberation Model

In this final test the cuboid room of Section V-B is again considered, but this time in the context of statistical objective room acoustic parameters that can be derived from RIR measurements according to [1], [2]. In addition to the image-source and 3-D FDTD datasets, the same cuboid is modeled again for the same 25 source-receiver combinations but only for the positive reflection coefficients $r = 0.1, 0.2, \ldots, 0.999$. This is performed using the proposed hybrid BT-ART in Section III-D and the 2-D FDTD approach for late reverberation proposed in [29].

In the first example, the reverberation time (RT60), for the cuboid room is derived using each of the 4 modeling methods. For all 4 modeling methods, for each source-receiver combination and reflection coefficient the RT60 in each octave band is calculated based on the T_{30} value obtained from the Schroeder decay curve. RT60 values are computed for each positive reflection coefficient based on Sabine and Norris-Eyring calculations. The results are presented in Fig. 9 for the 1 kHz octave band, demonstrating RT60 varying with reflection coefficient value.

The Sabine and Norris-Eyring predictions are computed by appropriately converting normal incidence pressure reflection coefficients, r used in the FDTD and BT models to their equivalent random incidence energy absorption coefficients, using the well known Paris formula [6]. The predictions are provided for reference, and as expected, these predictions differ most for lower reflection coefficients, where Sabine's formula is known to be the least accurate. There is very close agreement in T_{30} between the 3-D FDTD and the ISM results, and this is the case for both a single source-receiver combination and for the averaged T_{30} over all 25 combinations. As discussed in Section II, the 2-D FDTD method was proposed as an approach for synthesising the

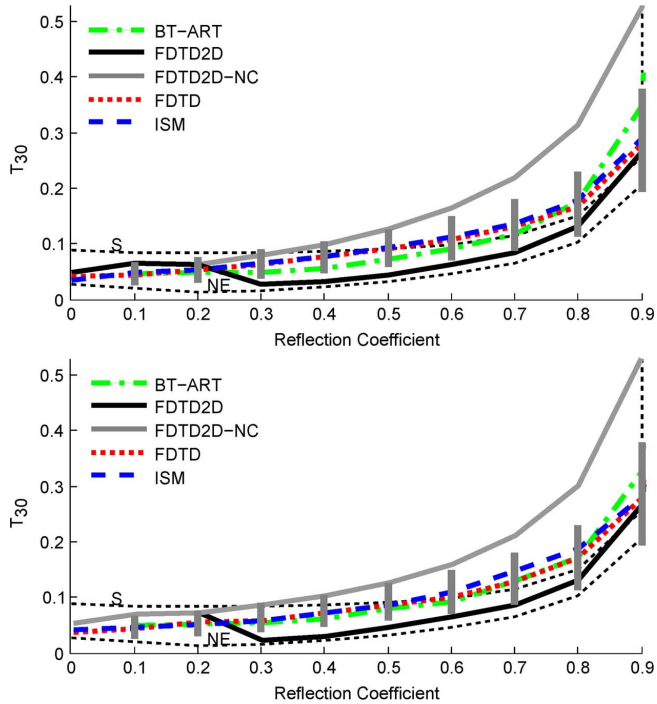


Fig. 9. A comparison of RIR derived T_{30} values for the 1 kHz octave band for a range of positive reflection coefficients for the cuboid model based on FDTD 2D no compensation (NC), 2D with compensation and 3D, as well as the image-source and ART methods, together with values derived from Sabine (S) and Norris-Eyring (NE) reverberation time calculations. The $\pm 30\%$ JND for T_{30} is indicated by the vertical bars. (a) T_{30} averaged across all source-receiver combinations (b) T_{30} values obtained for S1-R1.

late reverberation in a hybrid model [29]. In order to make comparisons with this previous work, T_{30} values derived form 2-D FDTD obtained RIRs are shown in Fig. 9, where the height component has simply been omitted from the random set of source/receiver positions used in all other methods. The required impulse response compensation, for better approximating a 3-D modeled impulse responses tail from a 2-D modeled impulse response [30], was applied and is indicated in Fig. 9. Generally, it can be seen that this compensation has significantly improved the T_{30} to better represent the Sabine and Norris-Eyring predictions.

Despite this, the 3-D FDTD and ISM results are more appropriate references/targets as these are proper solutions to the cuboid room scenario, as evidenced in Section V-A, rather than estimations as with the Sabine and Norris-Eyring formulae. This is an important point, because the aim is to validate the hybrid model by demonstrating that each individual modeling method can produce comparable responses for at least the most fundamental room geometry. This gives confidence in the results obtained when each method is employed separately in a given time-frequency region in the hybrid RIR synthesis, and for this reason their agreement with the Sabine and Norris-Eyring formulae is not so important. Therefore, while the compensated 2-D FDTD result generally agrees with Sabine and Norris-Eyring, it is clear that the ART method proposed for the high frequency late reverberation gives an improved match to the 3-D FDTD and ISM results.

Finally, in order to objectively quantify the observed differences in reverberation time, the minimum required change in

TABLE III
SELECTED ROOM ACOUSTIC PARAMETERS FOR THE CUBOID MODEL AT THREE REFLECTION COEFFICIENTS ACCOMPANIED BY THE JND VALUES FOR COMPARISON AS IN [41]

	Param.	ISM	3DFD	2DFD	BT-ART	JND
(a) 0.2	T_{30}	52.1	53.5	63.3	48.2	6/17
	EDT	11.2	11.4	20.0	12.2	0.8
	T_C	1.3	1.3	2.3	1.4	10.0
	C_{80}	94.2	67.0	61.3	243.0	1.0
	C_{50}	66.5	61.0	54.3	167.8	1.0
(b) 0.6	T_{30}	111.4	108.3	61.7	91.4	9/26
	EDT	65.5	66.1	37.8	93.0	2.6
	T_C	3.4	3.6	3.1	6.6	10.0
	C_{80}	45.4	45.5	80.2	54.7	1.0
	C_{50}	33.8	33.9	53.8	33.5	1.0
(c) 0.8	T_{30}	177.6	166.1	131.3	175.9	15/46
	EDT	125.4	131.2	127.2	187.4	6.3
	T_C	7.5	7.9	7.7	13.3	10.0
	C_{80}	31.5	32.3	37.6	26.9	1.0
	C_{50}	22.1	23.2	25.0	16.2	1.0

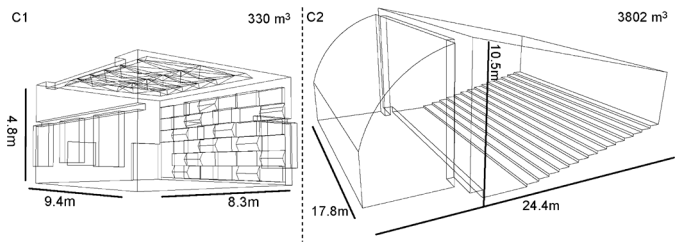


Fig. 10. Two room geometries, C1 the PTB studio and C2 a fictional hall, for which the validated hybrid acoustic model is applied to produce the auralizations at [38].

T_{30}, the JND, is marked in Fig. 9. The JND of T_{30} has been quoted previously by many researchers and ranges from 5% to 30% i.e., [40]–[42]. Note that the JND of acoustic parameters in general can vary with differing sound stimuli as observed using different music excerpts i.e., [43]. Therefore the $\pm 30\%$ values for each reflection coefficient are marked in Fig. 9. The mean T_{30} between the 3-D FDTD and ISM results are used as the reference. It is apparent that the 3-D FDTD, ISM and BT-ART curves are within the JND unlike the compensated 2-D FDTD method previously proposed in [29]. Furthermore, based on [1], [2], other room acoustic parameters can be derived and in Table III values for T_{30}, early decay time (EDT), clarity (C_{50} and C_{80}), and centre time (T_c) are presented for reflection coefficient values 0.2, 0.6 and 0.8. In addition, after i.e., [40], [41] the JND thresholds for each objective parameter are included for comparison purposes. As for T_{30}, the JND values for the other acoustic parameters should be compared using the mean of the 3-D FDTD and ISM values as the reference.

VI. HYBRID ROOM IMPULSE RESPONSE SYNTHESIS

Finally, based on the validations of Section V it is possible to synthesize a complete RIR suitable for auralization purposes based on the Hybrid Acoustic Model as outlined in Fig. 2. Two different complex room geometries were modeled, case 1 (C1) the PTB studio used in the third round robin on room acoustics computer simulation [40], and case 2 (C2) a fictional hall including a curved stage ceiling shown in Fig. 10.

C1 was modeled in the 1kHz band by both the FDTD and BT-ART parts separately for comparison with the real measured data with $W_{f_s} = 28$ kHz and $G_{f_s} = 48$ kHz. The boundary coefficients and source-receiver locations were those specified in the round robin data, noting that the random incidence absorption coefficients need to be converted to their normal incidence counterparts for use with the FDTD and BT algorithms. Table IV compares the measured and modeled T30 for the 1 kHz band over 6 source-receiver combinations.

TABLE IV
THE OCTAVE BAND REVERBERATION TIMES FOR C1. CALCULATED FROM FDTD AND BT-ART MODELED RESPONSES AND REAL MEASUREMENT DATA

C1,T_{30},1kHz	S1R1	S1R2	S1R3	S2R1	S2R2	S2R3
Measured	0.93	0.94	0.95	1.13	1.15	1,10
FDTD	0.96	0.91	0.97	0.98	1.00	0.95
BT-ART	1.02	1.06	1.00	1.06	1.12	1.07

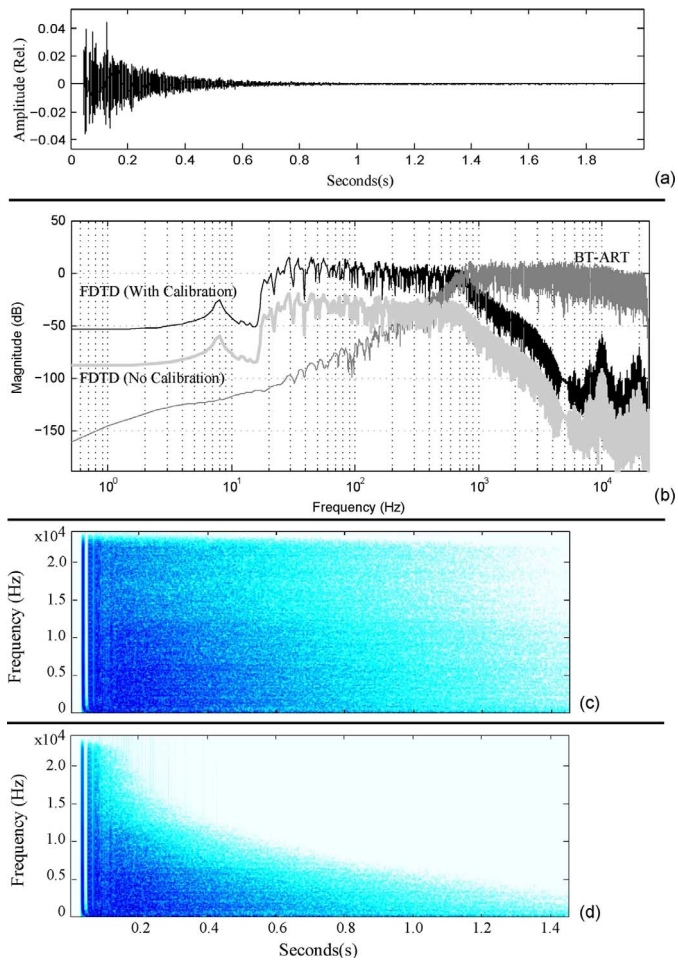


Fig. 11. Some typical aspects of the intermediate and final hybrid room impulse response. In (a) the final time domain hybrid impulse response and (b) the synthesized impulse responses after combining the octave bands for the FDTD with the proposed calibration in (11) applied (black) and BT-ART (dark gray). The FDTD with no calibration (light gray) is shown to further highlight the need for the calibration parameter. In (c) and (d) a comparison with and without the air absorption applied to each source in C2.

C2 was modeled with different absorption coefficients attributed to each boundary over octave bands from 32 Hz to 16 kHz with $W_{f_s} = 10$ kHz and $G_{f_s} = 48$ kHz. The geometric model was used for the 2 kHz octave band and above. One B-Format receiver captured the virtual equivalent of the 24 channel loudspeaker orchestra defined in [44]. The reciprocal setup was modeled to reduce the number of simulation passes from 24 to 4, one for each B-Format channel. For each octave

band the 24 synthesized impulse responses were convolved with the corresponding 24 anechoic recordings of the orchestra parts.

For both C1 and C2 the BT and ART crossover time region was defined by the longest possible reflection path that could be represented by the beam-tree and the output sampling rate $fs = 48$ kHz. In addition, a listing of the boundary materials and auralizations in mono, stereo and B-Format can be found at [38]. Fig. 11 shows the final hybrid RIR for C2 as well as some of the intermediate signals in the post-processing stages for comparison.

VII. Conclusion

This article has discussed the validation of a recently proposed hybrid acoustical model for synthesizing RIRs. Specific attention is given to matching the responses across the three modeling domains, namely Beam Tracing (BT), Acoustic Radiance Transfer (ART) and the Finite Difference Time Domain methods (FDTD).

The synthesis framework is organized into frequency bands, under ideal conditions the FDTD method would compute the entire RIR in all bands to ensure physical accuracy. In practice, for a given room geometry there is an upper sampling limit that, in turn, determines an upper usable frequency limit according to the well known problem of dispersion error. Upper frequency bands are modeled with a combination of geometric methods. BT is physically founded at high frequencies and is employed in the upper bands and ideally for the entire RIR in time. As the size of the beam-tree is limited by memory this determines the practical crossover time region to the statistically motivated ART method use in the late part in the geometric bands.

A new analytic function, with clear advantages over previous methods, is derived for correctly balancing the energy between the geometric and FDTD method based responses. Primarily this provides a convenient approach for calibrating the hybrid model but also for setting the absolute source strength of one or more sources in a 3D FDTD SRL acoustic model. The FDTD and BT methods exhibit excellent agreement in their single reflection, modal and acoustic parameter analysis when implemented with the same boundary model. This implies objective and subjective integrity in the high frequency contribution of the BT in the final RIR. The diffuse field assumption of the ART method gives good prediction of reverberation time, even though it is compared in the most challenging case of a reflective cuboid modeled with specular reflections with the FDTD and BT methods. Therefore it is reasonable to assume that the ART contribution would become more accurate in more complex room geometries where standing waves are not typically present at high frequencies. Furthermore, the accuracy of ART is compared and shown to be a considerable improvement over a compensated 2D FDTD method for the late reverberation used in some previous work.

Finally, the validated hybrid model is applied to the 3rd round robin PTB studio model and compares reverberation times against six measured responses using the FDTD and combined BT-ART responses at 1 kHz. Generally the agreement is very good, but further comparisons to real measured data along with listening tests is required to make more solid conclusions and this will be the subject of future work. Anechoic source material and a rendered-for-stereo auralization is available at [38] along with a list of the corresponding materials and surfaces. The subjective quality of the auralization is clearly dependent on the auralization method and parameters of the hall, though it is convincing and encouraging for future work.

References

[1] *Int. Org. for Standardization*, ISO 3382-1, 2009, Acoustics, Measurement of room acoustic parameters. Part 1: Performance spaces.

[2] *Int. Org. for Standardization*, ISO 3382-2, 2009, Acoustics, Measurement of room acoustic parameters. Part 2: Reverberation time in ordinary rooms.

[3] C. Schissler and D. Manocha, "GSound: Interactive sound propagation for games," in *Proc. Audio Eng. Soc. 41st Int. Conf., Audio for Games*, London, U.K., Feb. 2–4, 2011.

[4] A. Foteinou, D. T. Murphy, and A. Masinton, "Investigation of factors influencing acoustic characteristics in geometric acoustics based auralization," in *Proc. 13th Int. Conf. Digital Audio Effects (DAFx-10)*, Graz, Austria, Sept. 6–10, 2010, pp. 178–181.

[5] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja, "The room acoustic rendering equation," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1624–1635, 2007.

[6] H. Kuttruff, *Room Acoustics*, 5th ed. London, U.K.: Spon, 2009.

[7] D. Botteldooren, "Finite difference time domain simulation of low frequency room acoustic problems," *J. Acoust. Soc. Amer.*, vol. 98, no. 6, pp. 3302–3308, 1995.

[8] P. Macey, "Simple room acoustic analysis using a 2.5 dimensional approach," in *Proc. AES 130th Conv.* , London, U.K., May 13–16, 2011, Convention paper 8443.

[9] L. Savioja, A. Järvinen, K. Melkas, and K. Saarinen, "Determination of the low frequency behavior of an IEC listening room," in *Proc. Nordic Acoust. Meeting*, Jun. 1996, pp. 55–58.

[10] A. Southern, S. Siltanen, and L. Savioja, "Spatial room impulse responses with a hybrid modelling method," in *Proc. 130th AES Conv.*, London, U.K., May 13–16, 2011, Convention Paper 8385.

[11] S. Petrausch and R. Rabenstein, "Wave field simulation with the functional transformation method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '06)*, May 14–19, 2006, pp. 317–320.

[12] D. T. Murphy and M. Beeson, "The KW-boundary hybrid digital waveguide mesh for room acoustics applications," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 552–564, Feb. 2007.

[13] D. T. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley, "Acoustic modeling using the digital waveguide mesh," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 55–66, Mar. 2007.

[14] K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 1, pp. 34–46, Jan. 2011.

[15] K. Kowalczyk, M. van Walstijn, and D. T. Murphy, "A phase grating approach to modeling surface diffusion in FDTD room acoustics simulations," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 528–537, Mar. 2011.

[16] 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.*, vol. 20, no. 11, pp. 2420–2432, Nov. 2012, 2012.

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

[18] C. J. Webb and S. Bilbao, "Computing room acoustics with CUDA—3D FDTD schemes with boundary losses and viscosity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 22–27, 2011, pp. 317–320.

[19] J. J. Lòpez, D. Carnicero, N. Ferrando, and J. Escolano, "Parallelization of the finite-difference time-domain method for room acoustics modelling based on CUDA," *Math. Comput. Modell.*, Available online Dec. 14, 2011, .

[20] R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, and D. Manocha, "An efficient GPU-based time domain solver for the acoustic wave equation," *Applied Acoustics*, vol. 73, no. 2, pp. 83–94, Feb. 2012.

[21] M. Vorländer, "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm," *J. Acoust. Soc. Am.*, vol. 86, no. 1, pp. 172–178, 1989.

[22] J. Rindel, "Computer simulation techniques for acoustical design of rooms," *Acoust. Australia*, vol. 2, no. 3, pp. 81–86, 1995.

[23] R. A. Tenenbaum, T. S. Camilo, and J. C. B. Torres, "Hybrid method for numerical simulation of room acoustics with auralization: Part 1—Theoretical and numerical aspects," *J. Brazilian Soc. Mech. Sci. Eng.*, pp. 211–221, 2007.

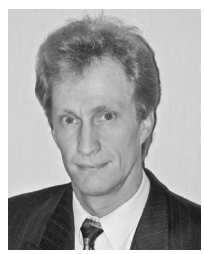
[24] E. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 267–277, 2008.
[25] E. Granier, M. Kleiner, and B.-I. L. Dalenbäck, "Experimental auralization of car audio installations," in *Proc. 130th AES Conv.*, Paris, France, Feb. 25–18, 1995, vol. 25, convention paper 3952.
[26] M. Bansal, S. Feistel, and W. Ahnert, "First approach to combine particle model algorithms with modal analysis using FEM," presented at the Proc. AES 118th Conv., Barcelona, Spain, May 28–31, 2005.
[27] M. Aretz, R. Nöthen, M. Vorländer, and D. Schröder, "Combined broadband impulse responses using FEM and hybrid ray-based methods," in *Proc. EAA Auralization Symp.*, Espoo, Finland, Jun. 15th–17th, 2009.
[28] I. Drumm, "A hybrid finite element/finite difference time domain technique for modelling the acoustics of surfaces within a medium," *Acta Acust. United With Acust.*, vol. 93, no. 5, pp. 804–809, 2007.
[29] D. T. Murphy, M. Beeson, S. Shelley, A. Southern, and A. Moore, "Hybrid room impulse response synthesis in digital waveguide mesh based room acoustics simulation," in *Proc. 11th Int. Conf. Digital Audio Effects (DAFx-08)*, Espoo, Finland, Sep. 1–4, 2008, pp. 129–136.
[30] J. J. Wells, D. T. Murphy, and M. Beeson, "Temporal matching of 2-D and 3-D wave-based acoustic modeling for efficient and realistic simulation of rooms," in *Proc. 126th AES Conv.*, Munich, Germany, May 7–10, 2009, Convention paper 7697.
[31] J. Escolano, J. J. Lòpez, and B. Pueo, "Directive sources in acoustic discrete-time domain simulations based on directivity diagrams," *J. Acoust. Soc. Amer.*, vol. 121, no. 6, pp. EL256–EL262, 2007.
[32] A. Southern and D. T. Murphy, "Low complexity directional sound sources for finite difference time domain room acoustic models," in *Proc. 126th AES Conv.*, Munich, Germany, May 7–10, 2009, Convention paper 7764.
[33] 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.*, vol. 8, no. 2, pp. 184–194, Feb. 2009.
[34] S. Laine, S. Siltanen, T. Lokki, and L. Savioja, "Accelerated beam tracing algorithm," *Appl. Acoust.*, vol. 70, no. 1, pp. 172–181, 2009.
[35] T. Lewers, "A combined beam tracing and radiation exchange computer model of room acoustics," *Appl. Acoust.*, vol. 38, pp. 161–178, 1993.
[36] H. E. Bass, "Atmospheric absorption of sound: Analytical expressions," *J. Acoust. Soc. Amer.*, vol. 52, no. 3B, pp. 821–825, 1972.
[37] E. Mommertz, "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," *Appl. Acoust.*, vol. 46, pp. 251–263, 1995.
[38] A. Southern, Dec. 2012, Hybrid model supporting material and Sound Examples. [Online]. Available: http://mediatech.aalto.fi/publications/acoustics/ham/index.html
[39] A. Southern, T. Lokki, L. Savioja, and D. T. Murphy, "The perceptual effects of dispersion error on room acoustic model auralization," in *Proc. Forum Acusticum 2011*, Aalborg, Denmark, Jun. 27–Jul. 1 2011.
[40] I. Bork, "Report on the 3rd round robin on room acoustical computer simulation—Part 1: Measurements," *Acta Acust. United With Acust.*, vol. 91, pp. 740–752, 2005.
[41] E. A. Lehmann and A. M. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1429–1439, Aug. 2010.
[42] Z. Meng, F. Zhao, and M. He, "The just noticeable difference of noise length and reverberation perception," in *Proc. Int. Symp. Commun. Inf. Technol.*, Oct. 18–Sep. 20 2006, pp. 418–421.
[43] T. J. Cox, W. J. Davies, and Y. W. Lam, "The sensitivity of listeners to early sound field changes in auditoria," *Acustica*, vol. 79, no. 1, pp. 27–41, 1993.
[44] J. Pätynen and T. Lokki, "Evaluation of concert hall auralization with virtual symphony orchestra," in *Proc. Int. Symp. Room Acoust. (ISRA'10)*, Melbourne, Australia, Aug. 29–31, 2010.
[45] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1421–1448, Jul. 2012.

Alex Southern is a researcher at the Department of Media Technology, Aalto University School of Science, Finland. He completed his Ph.D. in the Department of Electronics, University of York, UK, in 2011, where he was awarded an M.Sc. by Research in 2006. Before this he was awarded a first class B.Sc.(Hons.) in music technology and audio system design from the University of Derby, UK. His research interests include auralization and acoustic modeling. He has been a visiting researcher at the University of Aveiro, Portugal.

Samuel Siltanen received the degree of M.Sc. in computer science and engineering from the Helsinki University of Technology (TKK), Espoo, Finland, in 2005. He subsequently received the degree of Doctor of Science in Technology from Aalto University (former TKK). His doctoral thesis was on efficient and physically-based room acoustics modeling.

He worked at the TKK Laboratory of Telecommunications Software and Multimedia as a researcher from 2005 until the end of 2009. After that he continued the research in the Department of Media Technology at Aalto University School of Science. His research interests include room acoustics modeling and auralization.

Damian T. Murphy received the B.Sc. (Hons.) degree in mathematics, the M.Sc. degree in music technology, and the D.Phil. degree in music technology, all from the University of York, UK, in 1993, 1995, and 2000, respectively. He is a Reader in the AudioLab based in the Department of Electronics, University of York, where he has worked since 2000 and is a Visiting Lecturer in the Department of Speech, Music, and Hearing, KTH, Sweden and has held visiting researcher positions at a number of universities internationally. His research interests include acoustic modeling and spatial audio. He is a member of the Audio Engineering Society.



Lauri Savioja (M'00–SM'08) received the degrees of M.Sc., the Licentiate of Science, and the Doctor of Science in Technology, from the Helsinki University of Technology (TKK), Espoo, Finland, in 1991, 1995, and 1999, respectively. In all those degrees, he majored in computer science, and the topic of his doctoral thesis was room acoustic modeling.

He worked at the TKK Laboratory of Telecommunications Software and Multimedia as a researcher, lecturer, and professor from 1995 till the formation of the Aalto University where he currently works as a professor in the Department of Media Technology, School of Science. The academic year 2009–2010, he spent as a visiting researcher at NVIDIA Research. His research interests include room acoustics, virtual reality, and parallel computation.

Prof. Savioja is a fellow of the Audio Engineering Society (AES), a member of the Association for Computing Machinery (ACM), and a life member of the Acoustical Society of Finland. From 2010 he has been an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.