# ROOM IMPULSE RESPONSE INTERPOLATION FROM A SPARSE SET OF MEASUREMENTS USING A MODAL ARCHITECTURE

*Orchisama Das* [*]

CCRMA, Stanford University
Stanford, USA

*Paul Calamia, Sebastia V. Amengual Gari*

Facebook Reality Labs Research
Redmond, USA

## ABSTRACT

In augmented reality applications, where room geometries and material properties are not readily available, it is desirable to get a representation of the sound field in a room from a limited set of available room impulse response measurements. In this paper, we propose a novel method for 2D interpolation of room modes from a sparse set of RIR measurements that are non-uniformly sampled within a space. We first obtain the mode parameters of a measured room. Using the common-acoustical pole theory, the mode frequencies and decay rates are kept constant over space, and a unique set of mode amplitudes is obtained for each measurement location. Based on the general solution to the Helmholtz equation, these mode amplitudes are modeled as periodic functions of 2D spatial location. For low frequency room modes, the model parameters are found with sequential non-linear least squares. Results show accurate spatial interpolation of perceptually relevant low frequency modes in rooms with simple geometries having non-rigid walls.

***Index Terms—*** RIR Interpolation, Sound Field Reconstruction, Room Acoustics, Optimization

## 1. INTRODUCTION

In room acoustics, a long-standing problem is the interpolation of Room Impulse Responses (RIRs). This is equivalent to reconstructing the sound field in a room from a limited set of measurements. In augmented reality applications, there might not be sufficient or accurate information to reliably conduct simulations due to missing geometries, inaccurate or unknown materials, etc. In such cases, a viable solution is to measure the RIR at specific positions in the room, thus obtaining a sparse representation of sound field of the room. The goal of this paper is to utilize this sparse dataset of RIRs to characterize the entire sound field of a room for real-time interpolation and extrapolation of RIRs.

Several attempts have been made in the literature for RIR interpolation and sound field reconstruction, such as dynamic time warping [1, 2], parametric approaches [3, 4], compressive sensing [5, 6, 7], spherical harmonics [8], physics-based methods [9, 10, 11] and more recently, neural networks [12]. In this paper, we extend the common-acoustical pole and residue model proposed by Haneda et al. in [13]. The common acoustical poles correspond to the resonance frequencies (or modes) of a room, while the residues are simple periodic functions of the source and receiver positions for rectangular rooms (also the general solution to the homogeneous Helmholtz equation).

We use the modal decomposition of RIRs as a basis for the interpolation. First, we find a common set of mode frequencies and decay rates using subband ESPRIT [14] on an average of time-aligned RIRs measured in a room at different locations. Then, we estimate the mode amplitudes at each location with linear least squares. The mode amplitudes (equivalent to the residue) are periodic functions of the source and listener positions, and can be parameterized by a spatial frequency (wave number) and a complex amplitude. For low-frequency room modes, we find the wave numbers and complex amplitudes of these functions with sequential non-linear least squares optimization. Once these parameters are estimated offline, the interpolated RIR can be synthesized in real-time by using a bank of parallel second order filters [15].

The rest of this paper is organized as follows. In Section 2, we discuss room modes and their pressure distribution as a function of the spatial location. In Section 3, we give the details of our proposed method; Section 3.1 describes the modal estimation algorithm and Section 3.2 describes the proposed optimization procedure for interpolation of low-frequency room modes. In Section 4, we show the results of our proposed method by interpolating the low-frequency magnitude response of rooms simulated with the finite difference time-domain method (FDTD). Finally, we conclude the paper in Section 5 and delineate scope for future work.

## 2. ROOM MODES

The sound field in a room is given by the 3D wave equation

$$\frac{\partial^2 p}{\partial t^2} = c^2 \left( \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right) \tag{1}$$

[*] Orchisama Das performed the work while at Facebook Reality Labs Research.

$$\begin{bmatrix} h_1(0) & \cdots & h_L(0) \\ h_1(1) & \cdots & h_L(1) \\ \vdots & \ddots & \vdots \\ h_1(T-1) & \cdots & h_L(T-1) \end{bmatrix} = \begin{bmatrix} e^{(j\omega_1-\alpha_1)0} & \cdots & e^{(j\omega_M-\alpha_M)0} \\ e^{(j\omega_1-\alpha_1)1} & \cdots & e^{(j\omega_M-\alpha_M)1} \\ \vdots & \ddots & \vdots \\ e^{(j\omega_1-\alpha_1)(T-1)} & \cdots & e^{(j\omega_M-\alpha_M)(T-1)} \end{bmatrix} \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{L1} \\ \gamma_{12} & \cdots & \gamma_{L2} \\ \vdots & \ddots & \vdots \\ \gamma_{1M} & \cdots & \gamma_{LM} \end{bmatrix} \tag{7}$$

where $p$ is the acoustic pressure, $c$ is the speed of sound in the medium, and $x, y, z$ are Cartesian coordinates. The solutions to this equation are standing waves, or room modes. The room impulse response can be characterized by a sum over $M$ modes, whose complex amplitudes, $\gamma$, are functions of space, whereas frequencies and dampings, $\omega$ and $\alpha$ respectively, determine the temporal response,

$$h(x, y, z, t) = \sum_{m=1}^{M} \gamma_m(x, y, z) \exp\left[(j\omega_m - \alpha_m)t\right]. \tag{2}$$

The complex mode amplitudes are the solution to the homogeneous Helmholtz equation,

$$\nabla^2 p + k^2 p = 0; k = \frac{2\pi f}{c} \tag{3}$$

where $\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$ is the Laplacian operator, $k$ is the 3D wave number and $f$ is the frequency. For a fixed source and a moving listener (or vice versa) in a rectangular (shoebox) room with rigid walls, the solution is well-known [16]. For a mode at a particular frequency $\omega_m$,

$$\begin{aligned} p_m(x, y, z) &= p(x)\, p(y)\, p(z) \\ &= C_m \cos(k_{x_m} x) \cos(k_{y_m} y) \cos(k_{z_m} z). \end{aligned} \tag{4}$$

From the boundary conditions (zero pressure gradient at the walls), we can derive the wave numbers in each direction as

$$\begin{aligned} k_{\mu_m} &= \frac{n_m \pi}{l_\mu}; \mu \in (x, y, z) \\ k &= \sqrt{k_x^2 + k_y^2 + k_z^2} \end{aligned} \tag{5}$$

where $n_m \in \mathbb{Z}^+$ and $l_\mu$ is the length of the room in the $\mu$ direction, and $k_{\mu_m}$ is the wave number associated with the $m$th mode in the $\mu$ direction.

In a rectangular room with non-rigid walls, damping is introduced and the general solution changes to

$$\begin{aligned} p_m(\mu) &= C_{\mu_m} \exp(j k_{\mu_m} u) + D_{\mu_m} \exp(-j k_{\mu_m} \mu) \\ k_{\mu_m} &= \frac{n_m \pi}{l_\mu} + j\delta_\mu. \end{aligned} \tag{6}$$

The wave numbers, determined by the boundary conditions, are no longer real but have an imaginary component determined by the wall absorption, $\delta_\mu$. Typically, we do not have information about $l_\mu$ or $\delta_\mu$. The goal of this paper is to find the wave numbers, $k_{\mu_m}$, and the constants, $C_{\mu_m}, D_{\mu_m}$, from an observed set of RIR measurements when either the source or the listener is fixed and the other moves in a 2D plane that contains the measurement points.

## 3. PROPOSED METHOD

### 3.1. Modal estimation

The RIRs measured at different locations in a specific room are time-aligned and averaged since they share a common set of poles. The common mode parameters - frequencies and decay rates, are estimated with subband ESPRIT [14]. The mean RIR is filtered into 12 non-uniform overlapping frequency bands based on a Bark scale [17], and decimated by a factor of $r = 8$, so that nearby modes in the low frequencies can be resolved. The mode frequencies and dampings are the generalized eigenvalues of the Hankel matrix formed by the impulse response and its shifted copy. Repeated poles are discarded, and the effect of filtering and decimation is undone by shifting the frequencies up the spectrum, and raising the decay rates to their $r$th roots.

In Eq. (7), a matrix of RIR measurements, $\boldsymbol{H} \in \mathbb{R}^{T \times L}$, with sampled time along its columns and sampled locations along its rows, is written as a product of a Vandermonde matrix, $\boldsymbol{V}$, of decaying complex sinusoids and a matrix of complex mode amplitudes, $\boldsymbol{\Gamma}$, sampled at different locations. Once the mode frequencies and dampings are estimated with ESPRIT, the mode amplitudes at each location can be estimated using linear least squares. Here, † stands for the Moore-Penrose matrix inverse.

$$\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Gamma}, \quad \boldsymbol{\Gamma} \approx \boldsymbol{V}^\dagger \boldsymbol{H} \tag{8}$$

### 3.2. Non-linear optimization for low-frequency modes

At low frequencies, the room modes are well separated. Beyond the Schroeder frequency, however, the resonant peaks overlap and it is not possible to isolate the effect of individual modes [16]. According to the Nyquist limit, if the minimum distance between any two measured points is $d$ m, then the maximum mode frequency that can be correctly interpolated without aliasing is $f_u = \frac{c}{2d}$ Hz. The advantage of optimization lies in overcoming this constraint. We want to spatially interpolate low frequency modes given an arbitrarily sampled, sparse set of RIR measurements from a room. Once the modal parameters are estimated, our aim is to fit parametric functions of the form of Eq. (6) to the complex mode amplitudes as a function of spatial location. Each mode amplitude at a specified location $(x, y)$ on a 2D plane can be written as

$$\begin{aligned} \hat{\gamma}_m(x, y) = &C_{1m} e^{-j(k_{x_m} x + k_{y_m} y)} + D_{1m} e^{j(k_{x_m} x + k_{y_m} y)} + \\ &C_{2m} e^{-j(k_{x_m} x - k_{y_m} y)} + D_{2m} e^{-j(k_{x_m} x - k_{y_m} y)} \end{aligned} \tag{9}$$

The constants $C_1, D_1, C_2, D_2$ and wave numbers $k_x, k_y$ are unknown for each mode. For a set of measurements at $L$ locations - $(x_1, y_1), (x_2, y_2), \cdots, (x_L, y_L)$,

$$\begin{bmatrix} \hat{\gamma}_m(x_1, y_1) \\ \hat{\gamma}_m(x_2, y_2) \\ \vdots \\ \hat{\gamma}_m(x_L, y_L) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{m_1} \\ \boldsymbol{u}_{m_2} \\ \vdots \\ \boldsymbol{u}_{m_L} \end{bmatrix} \begin{bmatrix} C_{1m} \\ D_{1m} \\ C_{2m} \\ D_{2m} \end{bmatrix} \tag{10}$$

$$\boldsymbol{u}_{m_l} = [e^{-j(k_{x_m} x_l + k_{y_m} y_l)} \quad e^{j(k_{x_m} x_l + k_{y_m} y_l)} \quad e^{-j(k_{x_m} x_l - k_{y_m} y_l)} \quad e^{j(k_{x_m} x_l - k_{y_m} y_l)}]$$

$$\hat{\boldsymbol{\gamma}}_m = \boldsymbol{U}_m(k_{x_m}, k_{y_m}) \boldsymbol{c}_m.$$

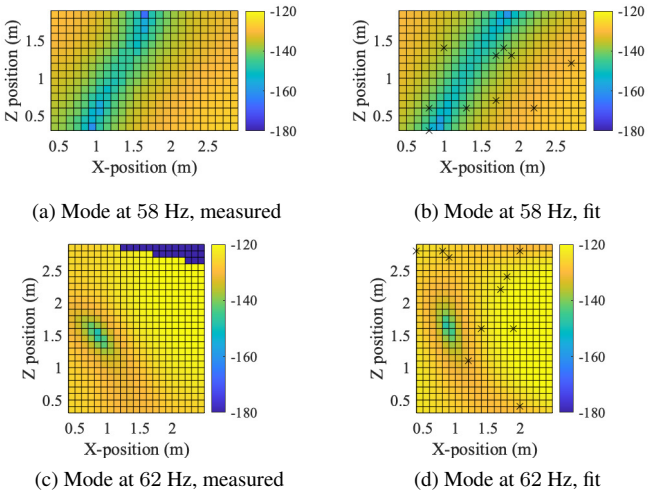
To find the optimal parameters, we use a sequential optimization scheme [18]. For each mode, we first update the constants using linear least squares, and then update the wave numbers with non-linear least squares. The details of the algorithm are given in Algorithm 1. $\oslash$ is the element-wise division of two matrices and $M_c$ is the number of low-frequency room modes. The difference in measured and modeled mode amplitudes is specified in decibels.

**Algorithm 1** Sequential optimization

**Require:** $0 \leq k_{x_m}, k_{y_m} \leq \frac{\omega_m + j\alpha_m}{c} \; \forall \; m$
  **for** $m = 1 \cdots M_c$ **do**
    Initialize $k_{x_m} = k_{y_m} = \frac{\omega_m + j\alpha_m}{\sqrt{2}c}$
    **repeat**
      $\boldsymbol{c}_{m_i} = \boldsymbol{U}^{*\dagger}_{m_{i-1}} \boldsymbol{\gamma}_m$
      $\hat{\boldsymbol{\gamma}}_{m_i} = \boldsymbol{U}^{*}_{m_{i-1}} \boldsymbol{c}_{m_i}$
      $J(k_{x_{m_i}}, k_{y_{m_i}}) = ||20 \log_{10}(\boldsymbol{\gamma}_m \oslash \hat{\boldsymbol{\gamma}}_{m_i})||_2^2$
      $k^*_{x_{m_i}}, k^*_{y_{m_i}} = \arg\min_{k_{x_m}, k_{y_m}} J$
    **until** convergence
  **end for**

## 4. EXPERIMENTS AND RESULTS

To test the efficacy of our proposed method, we simulated RIRs with a wave-based FDTD solver [19]. We present results for two simulated rooms - a shoebox room of dimensions $3 \times 2 \times 3$ m$^3$ with different frequency-independent admittances ($K_d$) on the walls - front and back wall $K_d = 0.9$, left and right wall $K_d = 0.8$, floor and ceiling $K_d = 0.99$, and another non-rectangular room with no parallel walls, of dimension 3 m on the longest edge in each direction, made of the same materials and having tilted walls at an angle of $9.4°$ in the $x, z$ directions. We used an omni-directional point source and placed virtual microphones in a rectangular grid on the $xy$ plane at a height of 1.7 m (nominal standing height). The grid resolution was 0.2 m, yielding a total of 442 measured RIRs for the shoebox room, and 594 RIRs for the room with tilted walls. The sampling rate was 240 kHz for the simulator[1], and 0.5 s of RIR was calculated. We resampled the RIRs to 48 kHz.




(a) Mode at 58 Hz, measured (b) Mode at 58 Hz, fit

(c) Mode at 62 Hz, measured (d) Mode at 62 Hz, fit

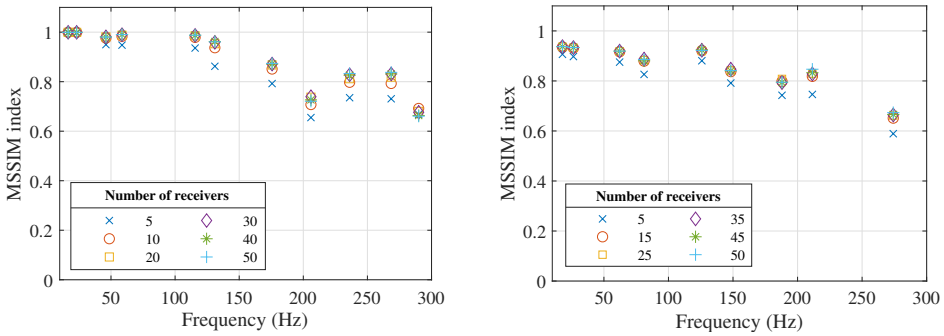
**Fig. 1**: Optimization fits using 10 receivers at locations $\times$. Shoebox room (top), room with tilted walls (bottom).

Mode shape fits for a low frequency are shown in Fig. 1 for both rooms. 10 microphones (locations marked in crosses) were used for the fit. The color intensity indicates sound pressure magnitude in decibels. The dark blue grids in Fig. 1c indicate points outside the room where the measured pressure is zero. To replicate the sparsity condition, we varied the number of microphones from 5 to 50. We ran $N_{tr} = 100$ trials for each set, placing the microphones in different configurations in each trial and evaluated the following metrics to show the effect of the number of measurement points on our results.

- **Mean Structural Similarity Index Measure (MSSIM)** - SSIM indicates the structural similarity between two images [20], and has values in the range $[0, 1]$. It has been used in [12] as an evaluation metric. A high SSIM index indicates that the measured and optimized mode shapes match closely. We average the SSIM over 100 trials to get the mean SSIM (MSSIM).
- **Absolute Mean Spectral Difference Error (AMSDE)** - The absolute difference between the frequency responses (averaged over all measurement points and all configurations) of the measured and modeled RIR, expressed in decibels. An AMSDE of 0 dB indicates perfect reconstruction. $H$ and $\hat{H}$ denote the measured and fit frequency responses respectively.

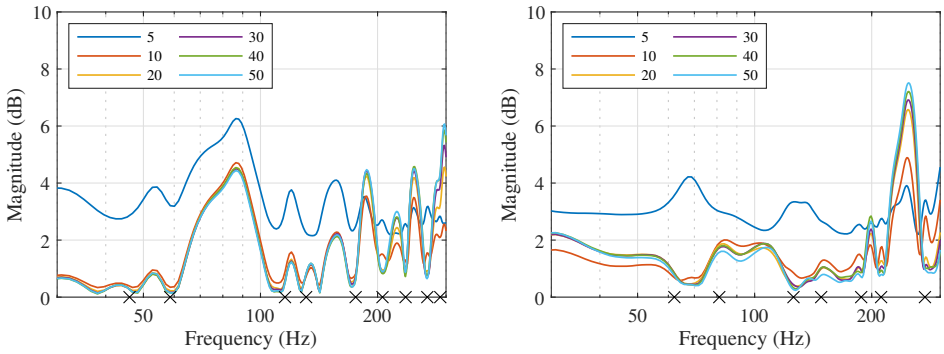
$$\text{AMSDE}(\omega) = \frac{1}{N_{tr}} \sum_{n=1}^{N_{tr}} \left| 20 \log_{10} \left( \frac{\sum_{l=1}^{L} H_{l,n}(\omega)}{\sum_{l=1}^{L} \hat{H}_{l,n}(\omega)} \right) \right| \tag{11}$$

[1]The sampling rate for FDTD simulations is typically large to reduce dispersion errors.




(a) Shoebox (b) Tilted Walls

**Fig. 2**: Mean SSIM index




(a) Shoebox (b) Tilted Walls

**Fig. 3**: Absolute Mean Spectral Difference Error (dB) between measured and interpolated responses

The variation in SSIM index with number of microphones is shown in Fig. 2. With only 5 microphones, the SSIM is noticeably worse. Beyond that, increasing the number of microphones does not yield a significant improvement. It is to be noted that the SSIM is only calculated at the modal frequencies. There is an overall decrease in SSIM with increase in frequency, similar to what is observed in [12]. With increasing frequency, the mode shapes become more complex, and need to be characterized by multiple periodic functions.

The AMSDE, as shown in Fig. 3, shows a similar trend, with 5 microphones performing worse than others. These results are promising given the small number of data points, and show that as few as 10 measurement points are adequate to capture the low frequency sound field of a room. The minima often appear at the modal frequencies, marked in black crosses, where the fit is nearly perfect. As we go further away from the mode frequencies, the error reaches a maximum.

## 5. CONCLUSION AND FUTURE WORK

In this paper we have proposed a novel method for interpolation of low-frequency room modes based on the general solution to the Helmholtz equation. The spatial distribution of mode amplitudes has been modeled as a periodic function characterized by a wave number in each direction, and an associated complex amplitude. We have proposed a sequential optimization scheme to estimate these parameters. Finally, we have tested the method on two rooms - a simple shoebox made of different materials, and another non-rectangular room with tilted walls. Two metrics have been evaluated as objective measures - the SSIM index and absolute mean spectral difference error between measured and modeled signals.

Our analysis with FDTD simulations of rooms has shown that the proposed method is capable of accurate RIR interpolation in the lower frequencies with a very small number of randomly distributed microphones. One advantage of this method compared to machine-learning based approaches [12] is that only a handful of parameters need to be stored for characterizing the low-frequency spatial response of a room - mode frequencies, dampings, two wave numbers and four constants for each mode. The RIR can be interpolated very efficiently in real-time once these parameters have been estimated offline. However, further questions need to be addressed, such as, comparison with existing approaches, more robust analysis with rooms of different sizes and geometries, and perceptual evaluation. It will also be useful to extend the model beyond low frequencies. We leave these questions open for a more detailed future work.

## 6. REFERENCES


[1] Claire Masterson, Gavin Kearney, and Frank Boland, "Acoustic impulse response interpolation for multichannel systems using dynamic time warping," in *AES Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009.

[2] Victor Garcia-Gomez and Jose J Lopez, "Binaural room impulse responses interpolation for multimedia real-time applications," in *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[3] Oliver Thiergart, Giovanni Del Galdo, Maja Taseska, and Emanuël AP Habets, "Geometry-based spatial sound acquisition using distributed microphone arrays," *IEEE Transactions on Audio, Speech, and Language processing*, vol. 21, no. 12, pp. 2583–2594, 2013.

[4] Mirco Pezzoli, Federico Borra, Fabio Antonacci, Stefano Tubaro, and Augusto Sarti, "A parametric approach to virtual miking for sources of arbitrary directivity," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2333–2348, 2020.

[5] Rémi Mignot, Laurent Daudet, and Francois Ollivier, "Room reverberation reconstruction: Interpolation of the early part using compressed sensing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2301–2312, 2013.

[6] Samuel A Verburg and Efren Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3770–3779, 2018.

[7] Elias Zea, "Compressed sensing of impulse responses in rooms of unknown properties and contents," *Journal of Sound and Vibration*, vol. 459, pp. 114871, 2019.

[8] Federico Borra, Israel Dejene Gebru, and Dejan Markovic, "Soundfield reconstruction in reverberant environments using higher-order microphones and impulse response measurements," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 281–285.

[9] Niccolo Antonello, Enzo De Sena, Marc Moonen, Patrick A Naylor, and Toon van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 1929–1941, 2017.

[10] Thach Pham Vu and Hervé Lissek, "Low frequency sound field reconstruction in a non-rectangular room using a small number of microphones," *Acta Acustica*, vol. 4, no. 2, pp. 5, 2020.

[11] Shoichi Koyama and Laurent Daudet, "Sparse representation of a spatial sound field in a reverberant environment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 172–184, 2019.

[12] Francesc Lluis, Pablo Martinez-Nuevo, Martin Bo Moller, and Sven Ewan Shepstone, "Sound field reconstruction in rooms: Inpainting meets super-resolution," *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. 649–659, 2020.

[13] Yoichi Haneda, Yutaka Kaneda, and Nobuhiko Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 709–717, 1999.

[14] Corey Kereliuk, Woody Herman, Russel Wedelich, and Daniel J Gillespie, "Modal analysis of room impulse responses using subband ESPRIT," in *Proceedings of the International Conference on Digital Audio Effects*, 2018.

[15] Jonathan S Abel, Sean Coffin, and Kyle Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Audio Engineering Society Convention 137*. Audio Engineering Society, 2014.

[16] Heinrich Kuttruff, *Room Acoustics*, CRC Press, 2016.

[17] Julius O Smith and Jonathan S Abel, "Bark and ERB bilinear transforms," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, 1999.

[18] PL Ainsleigh and JD George, "Modeling exponential signals in a dispersive multipath environment," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 1992, vol. 5, pp. 457–460.

[19] Jukka Saarelma, Jonathan Califa, and Ravish Mehra, "Challenges of distributed real-time finite-difference time-domain room acoustic simulation for auralization," in *AES International Conference on Spatial Reproduction-Aesthetics and Science*. Audio Engineering Society, 2018.

[20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.