

Auralization of a Hybrid Sound Field using a Wave-Stress Tensor Based Model

Aidan Meacham, Roland Badeau, Jean-Dominique Polack

▶ **To cite this version:**

Aidan Meacham, Roland Badeau, Jean-Dominique Polack. Auralization of a Hybrid Sound Field using a Wave-Stress Tensor Based Model. Forum Acusticum, Dec 2020, Lyon, France. pp.523-529, 10.48465/fa.2020.0833 . hal-03235295

HAL Id: hal-03235295
https://hal.science/hal-03235295

Submitted on 26 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AURALIZATION OF A HYBRID SOUND FIELD USING AN ENERGY-STRESS TENSOR BASED MODEL

Aidan Meacham[1] **Roland Badeau**[2] **Jean-Dominique Polack**[1]

[1] Sorbonne Université, Institut d'Alembert, CNRS UMR 7190, Paris, France
[2] LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
aidan.meacham@sorbonne-universite.fr

ABSTRACT

A hybrid approach to room impulse response synthesis and auralization is developed in the context of an energy-stress tensor based model of stochastic reverberation. This method for efficiently computing spatially varying energy envelopes has been demonstrated to represent the sound field in a sufficiently-diffusing 1-dimensional hallway above 250 Hz. To synthesize a realistic impulse response from the computed decay curves, the direct path, early reflections, and low frequency portion of the sound field must be calculated separately and then combined with the stochastic field to form a hybrid scheme. In this work, we propose one strategy for generating the stochastic field from the aforementioned energy envelopes and suggest the use of a typical pressure-velocity wave-based scheme and the image source method to generate the other necessary sound field components. Because of the efficiency of the energy-stress tensor based method and the reduced demands on the secondary simulation technique, such a hybridization presents a promising architecture for future real-time auralization in large spaces that may be difficult to model using only a single method.

1. INTRODUCTION

Combining multiple synthesis or simulation techniques in room acoustics is a common way to exploit the advantages of each method while also avoiding their pitfalls. For example, while wave-based simulation techniques accurately represent acoustic phenomena like modal behavior by construction, their computational complexity often grows with the cube of the highest frequency that one wishes to model, making them untenable for near-real-time audio tasks. Conversely, geometric approaches to acoustics are efficient at modeling the early part of sound propagation across the entire frequency band, but often require non-physical hacks to approximate more complex phenomena like edge diffraction. Finally, there are a number of methods focused on the so-called "late reverberation" that can generally be categorized as either energy-based, of which diffusion approaches are the most common, or tunable reverberation schemes, such as feedback delay networks. In both cases, it can be difficult to properly model the direct path and early reflections, and calibration of a particular space can be challenging, but the result is often an extremely efficient way to synthesize the tail part of sound propagation.

One alternative approach to late reverberation is an energy-stress tensor based method (or EST, for short) which is capable of modeling the sound field in sufficiently diffusing spaces [1]. It has been demonstrated that this theoretical approach is valid in primarily unidimensional spaces above the 250 Hz octave band, and is quite efficient due to the large cell sizes and low sample rates that are required to generate spatially and temporally accurate energy envelopes for the late reverberation. Up until this point, however, the integration of such an approach with an auralization method has not been realized. In order to determine how it may perform in a more general real-time acoustics tool, the hallway that has been characterized and replicated numerically in [2] makes for a useful test case.

It is clear that it would be advantageous to use each method in the region it is best suited for: wave-based for low frequencies, geometric for the early reflections, and the EST scheme for the late part. In fact, many modern room impulse response synthesis tools use approaches like this to accelerate computation times and are an area of active research in the academic community [3–6]. Often referred to as hybrid models, these approaches make a trade off between approximation, complexity, and computational speed, since it is often simpler to simulate a scenario entirely with one method, thus avoiding the need to intelligently pick cutoff frequencies or times, or to calibrate the different methodologies such that there is a reasonable expectation that they are modeling the same source or room characteristics. However, while recent advances in computational acoustics on graphics processing units (GPUs) has accelerated the speed at which wave-based or raytraced results can be generated, particularly with the use of clusters [7], on typical consumer hardware, the ability to synthesize full-band impulse responses in real time (and thus allowing for moving sources or receivers) is still the domain of hybrid approaches.

Therefore, we propose one possible method for rapidly generating full-band hybrid impulse responses in much the same way as in previous real-time acoustics engines, but introducing the computation of the late reverberation with the EST model to better take into account room geometry and motivate further development of the approach.

2. BACKGROUND

In this paper, we will mainly focus on the auralization of the late, high-frequency portion of the impulse response and the formation of the hybrid scheme, as the fundamentals of wave-based and geometric methods have been treated extensively in the literature. "Late reverberation" is the temporal region of a response with a sufficient density of reflections such that it can be represented as a stochastic process, rather than the distinct events of the direct sound and early reflections [8]. Similarly, at low frequencies, reverberation is typically dominated by behavior from a few isolated modes, but as the number of modes increases with frequency, they become sufficiently dense such that they can also be represented stochastically. The time and frequency cutoffs for these two criteria are commonly called the transition time and Schroeder's frequency, respectively. We will refer to this region of the time-frequency domain as the "stochastic reverberation," which is the main focus of the proposed hybrid model.

Because of this stochastic structure, this region has traditionally been modeled with reverberation algorithms that are significantly more efficient than wave-based or geometric schemes. One simple approach is directly filtering noise with the goal of approximating measured reverberation times in a frequency-dependent manner. More recent approaches in this vein use feedback delay networks (FDNs), a collection of delay lines that mix initial sound energy over time through feedback to produce natural sounding decay [9]. With a particular choice of absorption coefficients and delay line lengths, certain room characteristics can be reflected in the response of such a network, including (to some extent) the arrival times of early reflections or frequency dependent behavior. One downside of such schemes is that they often require complicated iterative tuning to match a specific chamber, since only some of their parameters are physically motivated.

One alternative approach that takes geometry into account more directly is the class of energy-based methods. Unlike most algorithms that model pressure and velocity waves, this category generally models the flow of sound energy in spaces with diffusing boundaries, and can be considered an extension of classical statistical reverberation. Because of their efficiency, they are often used to model very large spaces, especially when there are coupled volume effects present [10, 11]. One approach is the diffusion equation method (DEM), where sound energy diffuses throughout a space, much as in a heat diffusion problem [12–14]. Previously, models of this type have been auralized [15], however, little work has been done to unite them carefully with other methods in a hybrid model.

In contrast to DEM, the EST method supports a wave equation, which makes it possible to recover sound fields that build up in time depending on room geometry. Conveniently, though, since the end result is still a time series of energy envelopes, we can apply the energy envelopes to filtered noise in much the same fashion as previously studied with the DEM. In this paper, we implement both of the refinements suggested by Luizard et al. [15], taking advantage of multiple simulations for different frequency bands as well as the including the direct path and early reflections. The purpose of auralization is not only practical, in the sense that results from the EST method are bound to be used in applications where auralization is desired, but also pedagogical, as it provides a way to rapidly check that the theoretical framework as well as the numerical implementation are working as expected. Furthermore, it provides a framework for instant auditory feedback on future work on the model, particularly with respect to the development of a model for three dimensional spaces or checking that phenomena mathematically representable with the EST approach are actually audible.

In the interest of space, we will only present a brief overview of the theory here, and we will eschew the numerical treatment entirely, referring those interested readers to the publications in question [1, 16].

The main strategy in the EST model is the development of conservation laws for energy density E and sound intensity $\mathbf{J}$ [17]. In contrast, typically, wave-based acoustic simulations use an energy balance based on Gauss's theorem in terms of a pressure field p and a velocity field $\mathbf{v}$. Using these laws, one can derive a system of coupled equations relating the energy density, the sound intensity, and the wave-stress symmetric tensor $\underline{\underline{E}}$ [18]. In the case where one length dominates a space and the cross-sectional area is relatively constant (as is the case in a long hallway), dimensional reduction can eliminate the off-diagonal terms in the wave-stress tensor $\underline{\underline{E}}$ such that it resembles the scalar pressure field p, and a similar connection can be drawn between the vector velocity field $\mathbf{v}$ and $\mathbf{J}$. Using this formalism, one can follow the development of numerical models for common acoustic models with a few small modifications to the boundary conditions rather than requiring a domain-specific finite difference solution.

The main advantage of the EST method is that the propagation of diffuse energy allows for very large spatial discretization and low sample rates. Since we presume that the stochastic decay rate of the diffuse high frequency energy is not changing very rapidly in time and space, we can average over large cells and step through time at a relatively slow rate.

3. ROOM GEOMETRY AND MEASUREMENTS

For this work, we used the same hallway and impulse response dataset detailed in the frequency validity study [2]. A plan of the corridor can be seen in Figure 1.

The narrow portion of the hallway had an overall length of 45 m, with a width of 1.59 m, a suspended metal grating ceiling at 2.375 m, and a hard ceiling at 3.26 m, with a decrease to 2.8 m for metal support beams every 1.5 m along its length. The width increased to 2.39 m and the height decreased to 2.2 m in the recesses, some of which contained glass display cases or small pieces of furniture. The floor was linoleum and the walls were wooden panel and masonry. The heating, ventilation, and air conditioning systems were audible near vents but did not present an obstacle for the calculation of acoustical indices. All doors

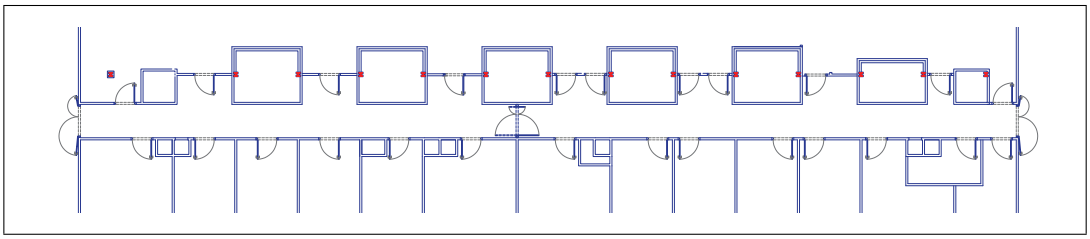


Figure 1. Floorplan for the corridor under consideration.

entering the hallway were closed, and the doors in the center of the hallway were fully open.

Impulse responses were collected using a *SoundField* ST250 microphone and an *Outline* GRS omnidirectional speaker with a *MOTU* Traveler sound card. The source was positioned 1 m away from the right end of the hall in Figure 1, 1.5 m above the ground, and centered between the two walls.

Beginning 1 m from the source, measurements were collected with the microphone's X-axis aligned along the length of the hallway. A spacing of 1 m was used out to 10 m, which corresponds exactly to the discretization distance in the numerical simulations. Then, recordings were made every 2 m until the end of the hallway for a total of 26 sampling locations. Recordings were made using the the swept sine method as implemented in the *Adobe Audition* plugin *Aurora* [19, 20]. The sweep length was 20 seconds at a sample rate of 44100 Hz, and the source level was adjusted digitally to maximize the signal-to-noise ratio without clipping as the microphone was moved further and further from the source. These gains were recorded in order to recover the true measured energy level for each measurement location. The sweep responses were then post-processed by convolution with the inverse sweep to recover impulse responses for each location. The collected impulse responses were time-aligned to the moment playback began so as to preserve the time-of-flight to the receiver, then trimmed to two seconds in length.

4. HYBRID MODEL

To facilitate direct comparison with the measured results, the synthesized impulse responses were also two seconds long at the same sample rate. For clarity, this does not imply that each of the methods used in the hybrid were run at the auralization sample rate, but in order to optimize for computation time, each was run at the lowest possible sample rate required for twice the number of samples in a second. These results were then upsampled (with the necessary filtering) in order to match the length of the measured impulse responses.

4.1 Stochastic Reverberation

The simulation of the stochastic reverberation for the hybrid method in this work comes from the finite volume time domain (FVTD) formulation of the EST model developed previously [16]. In short, the model uses a typical two-step scheme to solve for the energy density and sound intensity fields in a collection of cells. Since we are interested in modeling these fields in multiple frequency bands, we run the simulation multiple times independently with frequency-dependent modified absorption and scattering coefficients.

$$
\begin{aligned}
E_j^+ = &\left[E_j\left(1-\frac{cT}{V}\sum_{l=1}^{N_b}\gamma_{jl}S_l\frac{A}{8}\right)\right.\\
&\left.+\frac{cT}{V}\mathcal{P}_j-\frac{cT}{V}\sum_{k=1}^{N}\beta_{jk}SJ_{jk}\right]\\
&\Bigg/\left[1+\frac{cT}{V}\sum_{l=1}^{N_b}\gamma_{jl}S_l\frac{A}{8}\right],\\
J_{jk} = &\left[J_{jk}^-\left(1-\sum_{l=1}^{N_b}\zeta_{jkl}\frac{S_l}{V}cT\frac{D}{8}\right)\right.\\
&\left.+cTQ_{jk}+\frac{cT}{h}(E_j-E_k)\right]\\
&\Bigg/\left[1+\sum_{l=1}^{N_b}\zeta_{jkl}\frac{S_l}{V}cT\frac{D}{8}\right].
\end{aligned}
\tag{1}
$$

The update equation for this scheme is given in Eqn. (1), where E and J are the discretized average energy densities and sound intensities, j, k, and l are indices for a cell, another cell, and a boundary respectively, $\pm$ exponents indicate the next and previous discrete time steps for time series variables, c is the speed of sound, T is the chosen temporal sampling step, h is the inter-cell distance (with respect to centroids), S and S_l are the areas of boundaries between a cell and a neighboring cell or a boundary, V is the volume of a cell, A and D are the so-called modified absorption and scattering coefficients, γ, β, and ζ are indicator functions that select for relationships between particular cells and boundaries, and $\mathcal{P}$ and Q are source terms used to implement time-varying source behavior. In this case, h, S, S_l, V, A, and D are constant, but in general, could also be indexed on a cell-by-cell or boundary-by-boundary basis. Additionally, we omit the frequency index on A, D, E, J, Q, or $\mathcal{P}$ for clarity.

In this case, since we are interested in assembling an impulse response rather than processing audio directly, it is sufficient to observe the response of the system to an initial stimulus. It is equivalent to do so using the source terms directly, or to set the source terms to zero and use the initial conditions of the energy density field to excite a response in the space. As before, we use measurements

made in the real hallway to inform the parameters used for the EST model for each frequency band. In order to match the measurements during the model calibration, the virtual source is also placed 1 m from the end of the hallway. Then, a brute force tuning search over the model parameter space is performed to see which combinations of modified absorption and scattering coefficients give similar acoustical indices in each frequency band as the real recordings. The two target indices were spatial decays, or the slope of the decrease in sound energy as a function of distance from the source, and temporal decays, essentially T_{60}, both of which must fall within 10% (or better) of the observed value. The corresponding parameters in each frequency band are noted in Tab. 1 and were subsequently used in the hybrid model. More detail on the parameter fitting procedure can be found in [2].

After the model has been tuned for a given frequency band, it is straightforward to change the initial conditions in order to move the source location or add other sources. Given the large cell sizes and low sample rate, energy density envelopes can be rapidly generated when this occurs. Finally, the results are resampled to the audio sample rate that the different methods will be composited at.

In order to auralize the energy density envelopes, white noise of the length desired at the final sample rate is generated and split into octave bands using a filterbank. Then, the envelope associated with each octave band is multiplied with the bandlimited noise to create a signal with the same time of arrival and decay rate as in the measured hall. Finally, all of the resulting bandlimited responses are added in the time domain to create a single stochastic reverberation signal.

In the case where the source is presumed to be stationary, the results at every cell in the hallway can be cached and interpolated given a particular receiver location. It is only when the source distribution changes that the simulation and computation needs to be brought up to date.

4.2 Low-Frequency Reverberation

Modal behavior and diffraction that cannot be modeled by the EST method is handled by a FVTD simulation up through the octave band centered at 500 Hz [21]. Since we don't need to perform any boundary adjustments, the method is equivalent to a typical finite difference approach. In order to ensure that the dispersion error is minimized, the simulation sample rate f_{FVTD} (and by convention the element size) is chosen such that the maximum desired frequency occurs at $0.15 \times f_{\text{FVTD}}$, or about 7x oversampling [22]. We assigned boundary conditions in the model using the example materials given in the original work that best corresponded to each hallway surface. Since in this case the highest desired frequency is about 700 Hz, the sample rate was set at 4900 Hz. The resulting pressure field is then low-passed, spatially downsampled as needed to reduce the total memory footprint, and temporally resampled to match the final auralization sample rate.

Similarly to the stochastic reverberation, the simulation for this contribution only needs to be rerun in the case where the source distribution changes, but otherwise, can be cached in order to interpolate for a particular receiver location.

4.3 Direct Path and Early Reflections

The earliest arriving sound energy is computed with the image source method directly at the final sample rate [23]. There is some freedom to choose the maximum order of sources that will be included in the computation, constrained by the requirement of sufficiently representing the early reflections on one hand and without expending computation time unnecessarily on the multitude of reflections that will be modeled stochastically on the other.

For this method, the contribution must be recomputed any time the source or receiver position changes.

5. CALIBRATION

5.1 Hybrid Model

To calibrate the level between the separate simulations, we use an energy equalization in a frequency band where crossover occurs, using a distance of 1 m from the source as reference [24, 25]. Since all three methods have a valid portion of their frequency response in the octave band centered at 500 Hz, we can set an arbitrary level and find a gain for each simulation type that matches its energy to the target energy in that region. Then, when combining the methods by superposition in the time domain, the FVTD part is used directly as it is the most accurate, and the other methods are high-passed in order to avoid duplication of energy in the band of overlap.

Hybrid strategies are often delineated by frequency, and the point where two methods meet is chosen in order to optimize for speed of simulation at a given quality trade off. For other hybrid models, the topic of crossover frequency is well studied [26, 27]. In this case, however, because the EST method is not applicable at all below the Schroeder frequency, the cutoff frequency between the FVTD portion and the higher frequency models is determined in advance. Fortunately, it is not very high, and therefore the low-frequency FVTD portion can be simulated without too much difficulty. With a GPU, real-time simulation of even small concert halls has long been feasible at the frequency ranges under discussion [28]. Furthermore, given the speed at which the image source and stochastic reverberation portions can be computed, we believe that an optimized version of this approach could approach real-time performance with proper tuning.

5.2 Measurements & Hybrid

To calibrate levels between the measured and simulated impulse responses, a reference distance of 1 m is chosen, and the energy of the two signals are matched at that point. The resulting gain is then applied regardless of where the measurement or simulation is sampled so as to preserve the relative decay as a function of distance from the source.

Frequency [Hz]	250	500	1000	2000	4000	8000
T_{60} [S]	0.38	0.39	0.43	0.35	0.33	0.30
Spatial Decay [dB/m]	-0.88	-0.94	-0.66	-0.54	-0.66	-0.84
Mod. Abs. Coef. A	0.0516	0.0512	0.0473	0.0551	0.0577	0.0671
Mod. Sca. Coef. D	0.5359	0.5920	0.3818	0.2483	0.3210	0.4114

Table 1. Measured characteristics and resulting simulation parameters.

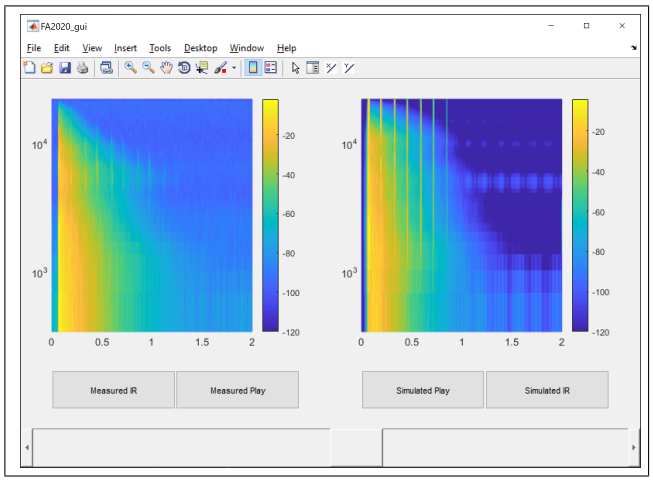


Figure 2. The interface used to compare measured and simulated impulse responses.

6. EVALUATION

The overall behavior of the system was inspected using a graphical interface that allowed visual examination of spectrograms (taken with the same parameters and with the same color scale for spectral power), direct auralization of the resulting impulse responses, as well as convolution of the IRs with source material. A slider ranging from 1 m to 42 m (the maximum measured observation distance from the source) was included, and moving the slider triggered an update of the receiver position, causing any relevant interpolation or re-simulation. With this scheme, it is straightforward to assess the changes in time of arrival, overall spectral envelope, and informally, the perceptual differences between the simulated and measured auditory presentations throughout the hallway as a function of distance.

In Fig. 2, the spectrogram of the measured impulse response is presented on the left, and the spectrogram of the simulated impulse response is presented on the right, corresponding to a receiver distance of around 23 m from the source.

7. RESULTS

Brief informal listening sessions revealed some of the advantages and shortcomings of the hybrid model.

The overall decay times and absolute levels at any given position are fairly similar between the simulated and measured impulse responses, and the arrival times of the most notable reflections are accurately rendered. In the cases where there are differences in decay times, they appear to be mostly due to the stochastic reverberation model's octave-band calibration using the temporal and spatial information from the measurements, which in some cases, was not sufficiently narrow. The FVTD and ISM portions also do not reproduce the measured decay as a function of distance perfectly, resulting in some spatial variation in the prominence of particular portions of the hybrid as a result of the constant level calibration gain. This is likely a result of the lack of tuning for the simulation parameters in both cases, as well as not taking viscothermal losses into account for the high frequency portion of the ISM, which is particularly prominent in the example simulation spectrogram in Fig. 2 above the 8 kHz octave band.

As a result, near the source, the stochastic portion is somewhat overrepresented, resulting in impulse responses that sound too diffuse relative to the strong direct path and early reflections present in the measurements. Furthermore, along the entire length of the hallway, the simulated impulse responses seem to have a slight tonal coloration shifted toward the higher frequencies, possibly resulting from the lack of tuning for the stochastic model above the band centered at 8 kHz.

One of the most apparent phenomena in the recorded impulse responses is the presence of a strong echo near the far end of the hallway from the source as a result of proximity to the reflecting surface. In the simulated impulse responses, this effect appears courtesy of the ISM portion. While the overall envelope of the EST method reflects the corridor-length delay behavior, the fine-structure detail that gives that area of the hallway its characteristic sound is decidedly not a part of the diffuse field, and therefore it cannot be expected that it would be precisely reproduced by the large elements and low sample rate. The second and third recurrences in time of this phenomenon are apparent in the overall effect of the measured impulse responses, if not directly audible as a flutter depending on position. Using a low order for the ISM in an attempt to avoid a proliferation of image sources eliminates these salient features from the simulated version entirely, so it is important to consider the tradeoff between computational efficiency and the relevance of particular specular phenomena.

Overall, while the speed and general characteristics of the simulation were promising, additional improvements must be made in order to provide a satisfactory facsimile of the original hallway.

8. FUTURE WORK

When the source distribution is stationary, the results of the stochastic reverberation contribution can be evaluated everywhere and interpolated on the fly to improve speed. However, if the source distribution changes frequently, it is most efficient to only upsample and do the computations at the points nearest the receiver. Thus, to balance between the two behaviors, it should be possible to begin the computation with only the necessary points, and if the source distribution ceases to change, to then fill in the rest of the possible receiver positions as able.

While the origin of the tonal differences between the simulated and measured impulse responses is unclear, the use of a third-octave filterbank in the design of the stochastic response as well as changing the crossover frequency may help in diagnosing where the issue is located. Additionally, moving to a 3-dimensional model may help model the differences between the narrow portions and alcoves along the hallway, which may have an effect on the overall frequency response at a given position.

In terms of salient spatial features that are not captured by the stochastic reverberation (such as the echo mentioned earlier), it is difficult to predict in the general case how to efficiently generate these effects under the current simulation strategy. In the specific case of the hallway, it is of course possible to generate a higher order of longitudinal image sources, which would give a reasonable approximation of the behavior in the full-band response; however, that is not the case in an arbitrary space where it is less obvious which image sources may be perceptually relevant without calculating higher orders in the first place. While the ISM approach in this case was relatively simple, the wide range of improvements that have been proposed in the literature are fair game for implementation, including frequency-dependent boundary conditions or viscothermal losses.

Recent advances in late reverberation techniques account for directionality in the simulated impulse responses [29]. Simulation of anisotropy is also possible with the EST method, and may present a useful framework for considering non-homogeneous diffusion coefficients as well.

Finally, the prediction of boundary conditions for the EST from material and geometry measurements remains an important avenue of research to eliminate the calibration step and enable simulation of unseen spaces.

9. CONCLUSIONS

In this paper, we have detailed ongoing work on a hybrid auralization scheme utilizing the EST based model for the stochastic reverberation combined with ISM and FVTD models to recreate and listen to the sound field in a long hallway. Using filtered noise and temporal envelopes on an octave-band by octave-band basis, we formed a portion of the impulse response valid above Schroeder's frequency and after the transition time. The gains used for the each method were calibrated using an octave band where all simulations were valid, then for the summation, the ISM and EST method were high-passed above the highest frequency present in the FVTD model, which was used as-is.

An interface to explore the acoustic space and compare the simulated and measured impulse responses was developed and used to derive some initial impressions of the auralization scheme. Informal results from this hybridization were promising in terms of computational speed and reproduction of the calibrated acoustical parameters, however, additional work remains in order to better match other salient spatial cues, and theoretical challenges remain regarding the temporal transition between the high-frequency methods. Nonetheless, the structure presented is an encouraging first step toward geometry-aware stochastic reverberation simulation strategies for real-time applications.

10. REFERENCES

[1] H. Dujourdy, B. Pialot, T. Toulemonde, and J.-D. Polack, "An Energetic Wave Equation for Modelling Diffuse Sound Fields – Application to Corridors," *Acta Acustica united with Acustica*, vol. 103, pp. 480–491, May 2017.

[2] A. Meacham, R. Badeau, and J.-D. Polack, "Lower Bound on Frequency Validity of Energy-Stress Tensor Based Diffuse Sound Field Model," in *Proceedings of the 23rd International Congress on Acoustics*, (Aachen), Sept. 2019.

[3] D. Murphy, S. Shelley, M. Beeson, A. Moore, and A. Southern, "Hybrid room impulse response synthesis in digital waveguide mesh based room acoustics simulations," in *DAFx*, 2008.

[4] D. Schröder and M. Vorländer, "RAVEN: A real-time framework for the auralization of interactive virtual environments," in *Proceedings of Forum Acusticum*, pp. 1541–1546, Jan. 2011.

[5] S. Oxnard and D. Murphy, "Room impulse response synthesis based on a 2D multi-plane FDTD hybrid acoustic model," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, NY, USA), pp. 1–4, IEEE, Oct. 2013.

[6] D. Poirier-Quinot, B. Katz, and M. Noisternig, "EVERTims: Open source framework for real-time auralization in architectural acoustics and virtual reality," in *DAFx-17*, (Edinburgh), Sept. 2017.

[7] H. Lai and B. Hamilton, "Computer Modeling of Barrel-Vaulted Sanctuary Exhibiting Flutter Echo with Comparison to Measurements," *Acoustics*, vol. 2, pp. 87–109, Mar. 2020.

[8] R. Badeau, "Common mathematical framework for stochastic reverberation models," *The Journal of the Acoustical Society of America*, vol. 145, pp. 2733–2745, Apr. 2019.

[9] J.-M. Jot and A. Chaigne, "Digital Delay Networks for Designing Artificial Reverberators," in *Audio Engineering Society Convention 90*, Audio Engineering Society, Feb. 1991.

[10] Z. Sü Gül, N. Xiang, and M. Çalışkan, "Investigations on sound energy decays and flows in a monumental mosque," *The Journal of the Acoustical Society of America*, vol. 140, pp. 344–355, July 2016.

[11] Z. Sü Gül, E. Odabaş, N. Xiang, and M. Çalışkan, "Diffusion equation modeling for sound energy flow analysis in multi domain structures," *The Journal of the Acoustical Society of America*, vol. 145, pp. 2703–2717, Apr. 2019.

[12] A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the use of a diffusion model for acoustically coupled rooms," *The Journal of the Acoustical Society of America*, vol. 120, pp. 2043–2054, Oct. 2006.

[13] Y. Jing and N. Xiang, "A modified diffusion equation for room-acoustic predication," *The Journal of the Acoustical Society of America*, vol. 121, pp. 3284–3287, June 2007.

[14] N. Xiang, Y. Jing, and A. C. Bockman, "Investigation of acoustically coupled enclosures using a diffusion-equation model," *The Journal of the Acoustical Society of America*, vol. 126, pp. 1187–1198, Sept. 2009.

[15] P. Luizard, J.-D. Polack, and B. F. G. Katz, "Auralization of coupled spaces based on a diffusion equation model," in *Proc. Sound and Music Computing Conference (SMAC/SMC)*, (Stockholm, Sweden), pp. 616–621, 2013.

[16] A. Meacham, R. Badeau, and J.-D. Polack, "Implementation of Sources in an Energy-Stress Tensor Based Diffuse Sound Field Model," in *Proceedings of the International Symposium on Room Acoustics*, (Amsterdam), Sept. 2019.

[17] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*. Mc Graw-Hill Book Company, 1953.

[18] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. Princeton University Press, 1968.

[19] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *Audio Engineering Society Convention 108*, Audio Engineering Society, Feb. 2000.

[20] A. Farina, "Advancements in Impulse Response Measurements by Sine Sweeps," in *Audio Engineering Society Convention 122*, Audio Engineering Society, May 2007.

[21] S. Bilbao, B. Hamilton, J. Botts, and L. Savioja, "Finite Volume Time Domain Room Acoustics Simulation under General Impedance Boundary Conditions," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 24, pp. 161–173, Jan. 2016.

[22] A. Southern, T. Lokki, L. Savioja, and D. Murphy, "The Perceptual Effects of Dispersion Error on Room Acoustic Model Auralization," in *Proceedings of Forum Acusticum*, (Aalborg), pp. 1553–1558, Jan. 2011.

[23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, Apr. 1979.

[24] A. Southern, S. Siltanen, and L. Savioja, "Spatial Room Impulse Responses with a Hybrid Modeling Method," in *Audio Engineering Society Convention 130*, Audio Engineering Society, May 2011.

[25] A. Southern, S. Siltanen, D. T. Murphy, and L. Savioja, "Room Impulse Response Synthesis and Validation Using a Hybrid Acoustic Model," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, pp. 1940–1952, Sept. 2013.

[26] S. Oxnard, *Efficient Hybrid Virtual Room Acoustic Modelling*. PhD thesis, University of York, Sept. 2016.

[27] C. Rougier, *Influence of Crossover Frequency on a Hybrid Acoustic Model for Room Impulse Response Synthesis*. Masters thesis, University of York, 2018.

[28] L. Savioja, "Real-time 3D finite-difference time-domain simulation of low- and mid-frequency room acoustics," in *DAFx*, 2010.

[29] B. Alary, A. Politis, S. Schlecht, and V. Välimäki, "Directional Feedback Delay Network," *JAES*, vol. 67, pp. 752–762, Oct. 2019.