A COMMON-SLOPES LATE REVERBERATION MODEL BASED ON ACOUSTIC RADIANCE TRANSFER

Matteo Scerbo,[1] Sebastian J. Schlecht,[2] Randall Ali,[1] Lauri Savioja,[2] and Enzo De Sena[1]

[1]Institute of Sound Recording
University of Surrey
Guildford, UK
m.scerbo, r.ali, e.desena@surrey.ac.uk

[2]Acoustics Lab
Aalto University
Espoo, Finland
sebastian.schlecht, lauri.savioja@aalto.fi

ABSTRACT

In rooms with complex geometry and uneven distribution of energy losses, late reverberation depends on the positions of sound sources and listeners. More precisely, the decay of energy is characterised by a sum of exponential curves with position-dependent amplitudes and position-independent decay rates (hence the name *common slopes*). The amplitude of different energy decay components is a particularly important perceptual aspect that requires efficient modeling in applications such as virtual reality and video games. Acoustic Radiance Transfer (ART) is a room acoustics model focused on late reverberation, which uses a pre-computed acoustic transfer matrix based on the room geometry and materials, and allows interactive changes to source and listener positions. In this work, we present an efficient common-slopes approximation of the ART model. Our technique extracts common slopes from ART using modal decomposition, retaining only the non-oscillating energy modes. Leveraging the structure of ART, changes to the positions of sound sources and listeners only require minimal processing. Experimental results show that even very few slopes are sufficient to capture the positional dependency of late reverberation, reducing model complexity substantially.

1. INTRODUCTION

Room acoustics modeling, i.e. the prediction of reverberation in enclosed spaces, has a wide variety of applications [1, 2, 3]. In virtual reality, video games, and other interactive applications, it is desirable to model reverberation in real time, based on changing parameters such as the positions of sound sources and listeners [4, 5]. One of the main perceptual attributes of reverberation is the rate of decay of acoustic energy in the room [6]. This aspect is particularly crucial in environments with highly uneven energy absorption distributions (e.g. coupled volumes, semi-open spaces), where several decay rates may be observed depending on the positions of the sonic agents [7, 8]. The common-slopes model [9] is particularly apt for applications that require an interactive update rate as well as agent-adaptive late reverberation [10]. Said model relies on the separation of parameters that depend on the static room geometry (the energy decay rates, hence the name common slopes) and those that depend on mobile agents (the amplitude of each slope) [11]. Previous works [11, 9] evaluated said parameters from a grid of point-to-point Room Impulse Responses (RIRs). Our proposed method seeks to evaluate them from the scene geometry, through the modal decomposition of a Time-Domain Acoustic Radiance Transfer (TD-ART) model.

Acoustic Radiance Transfer is a Geometrical Acoustics (GA) model that predicts the energy response of a room based on the geometric and material properties of the environment [12]. What makes it appealing for the purpose of common-slopes modeling is that its fundamental parameters are in fact independent of source and listener positions. Moreover, TD-ART may be interpreted as a state-space model in which each state variable describes the acoustic energy flowing through a particular point in space, in a particular direction. In any state-space model with a physical interpretation, such that each state variable is related to a position in space, the eigenvectors of the state transition matrix are related to modes' spatial distribution, while its eigenvalues are related to modes' evolution over time [13, 14]. In this paper we show that the eigendecomposition of TD-ART yields the energy decay rates of the environment in the form of eigenvalues, and the slope amplitudes' positional dependence in the form of eigenvectors. Through the eigenvectors, the common slopes parameters can be evaluated for arbitrary positions at runtime. The RIRs can then be synthesized using a sum of artificial reverberators [10].

Our target is the complexity reduction of ART. This endeavour was previously undertaken by Bai *et al.* [15]. Their proposed method analyses the physical aspects of the ART model from the perspective of acoustics, whereas the approach in this paper concerns the analysis of the model in the framework of system theory. Another noteworthy difference is that the method in [15] is designed and validated for rooms with spatially uniform energy decay, i.e. no dependence on source and listener positions, and a single decay rate dominating the RIR. Our method is designed to capture multiple slopes, and their positional dependency.

Section 2 reviews the fundamentals of energy decay rates, the common-slopes model, TD-ART, and the modal decomposition of state-space models. In Section 3, we describe our proposed method for late reverberation modeling. Experimental results are illustrated and discussed in Section 4. Section 5 closes this paper with some concluding remarks and thoughts on future work.

2. BACKGROUND

We first recall the relationships of impulse responses, energy responses and decay, and the common-slopes model, in Section 2.1.

Sebastian J. Schlecht is with the Dept. of Information and Communications Engineering and also at the Media Lab, Dept. of Art and Media, Aalto University, Espoo, Finland. Lauri Savioja is with the Dept. of Computer Science, Aalto University.

This work was supported by the Engineering and Physical Sciences Research Council under the SCalable Room Acoustics Modelling Grant EP/V002554/1.

Section 2.2 presents the TD-ART model, detailing the physical significance of its components. Section 2.3 reviews recent modal decomposition techniques related to models of the form of TD-ART.

2.1. Common-slopes model of late reverberation

The late reverberation energy decay $\bar{\epsilon}(\boldsymbol{\phi}, n)$ in a frequency band can be retrieved as the short-time average, denoted with $\langle\cdot\rangle$, of the band-passed energy response $\epsilon(\boldsymbol{\phi}, n)$, i.e., squared pressure response $h(\boldsymbol{\phi}, n)$ [6]:

$$\bar{\epsilon}(\boldsymbol{\phi}, n) = \langle \epsilon(\boldsymbol{\phi}, n) \rangle = \langle h(\boldsymbol{\phi}, n)^2 \rangle \,, \tag{1}$$

where $\boldsymbol{\phi}$ is a vector containing information on the source and listener positions and orientations, and n is time in samples. The energy response $\epsilon(\boldsymbol{\phi}, n)$, also known as *echogram* [6], is the primary output of most GA models – including ART [12], as discussed in the next section. The pressure-domain RIR can be retrieved from it through "noise shaping", meaning the square-rooted band energy $\sqrt{\epsilon(\boldsymbol{\phi}, n)}$ is used to modulate the amplitude of a band-passed stochastic signal [1].

The common-slopes model [9] approximates the energy decay using a superposition of N exponential slopes

$$\bar{\epsilon}(\boldsymbol{\phi}, n) \approx \sum_{i=1}^{N} \boldsymbol{\chi}_i^2(\boldsymbol{\phi})\, \delta_i^n \,, \tag{2}$$

where $0 < \delta_i < 1$ is the energy decay rate per sample, and $\boldsymbol{\chi}_i^2$ is the amplitude of the i^{th} slope depending on the positions $\boldsymbol{\phi}$. The key aspect of the model is that decay rates, δ_i, are position-independent and instead only depend on the room geometry and materials. In contrast, the slope amplitudes depend on the positions, but not on time. In an ideal diffuse room, i.e., with a late reverberation that is homogeneous and isotropic, the energy decay is a single slope, i.e., $N = 1$ with no positional change, i.e., $\boldsymbol{\chi}_i^2$ is constant [16]. For realistic scenarios, multiple slopes can occur concurrently. In [9], the common-slopes model has been validated to approximate the late reverberation in a three-room scenario accurately with only three slopes.

The method described in this paper is concerned with the approximation of $\bar{\epsilon}(\boldsymbol{\phi}, n)$ for one frequency band, but it may easily be extended to several bands. In an interactive application, $\boldsymbol{\phi}$ can change rapidly due to source and listener movements such that efficient position adaption is key. In the following, we drop the position dependency $(\boldsymbol{\phi})$ for brevity.

2.2. Acoustic Radiance Transfer

Acoustic Radiance Transfer is a GA model, i.e. it models sound propagation in the form of acoustic rays rather than acoustic waves. More precisely, it acts as a numerical solver for the Room Acoustics Rendering Equation (RARE), which is itself the foundation of GA [12]. The analytical RARE characterizes the acoustic energy (or "acoustic radiance") leaving each point of a room's boundary, in the direction of each other surface point, taking the form of a surface integral. By subdividing the room's boundary into a finite number of surface patches, and assuming that the transferred radiance is uniform from all points on one patch to all points of another, ART discretises the RARE integral in space. The paths connecting pairs of surface patches are referred to as "discrete directions". A visualization of two discrete directions is shown in Fig. 1. Note that each discrete direction is related to exactly two

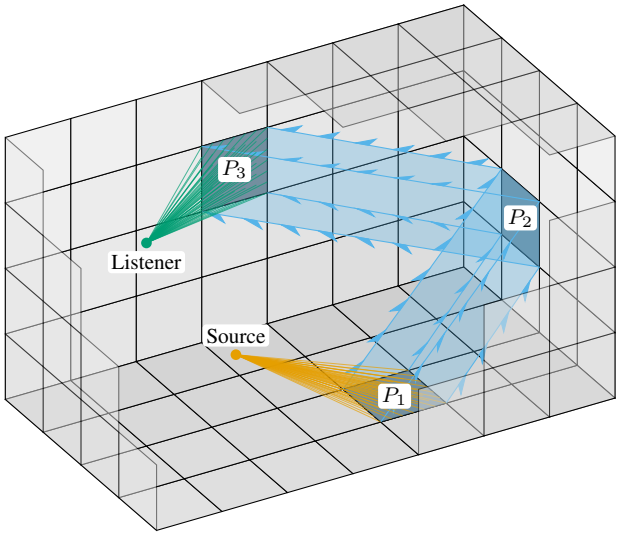


Figure 1: Illustration of discrete directions in ART. Two discrete directions are shown – one connecting the pair of surface patches P_1 and P_2, the other P_2 and P_3. The figure also shows a sound source injecting energy (acoustic rays) into the former discrete direction, and a listener detecting energy from the latter.

surface patches, and – as the name "direction" implies – the discrete direction going from patch P_1 to P_2 is distinct from the one going from P_2 to P_1.

Additionally, while the analytical RARE is a function of continuous time, describing it as a function of discrete time[1] gives rise to Time-Domain Acoustic Radiance Transfer (TD-ART) [17]. In the z domain, a TD-ART model with L discrete directions takes the form

$$\boldsymbol{s}(z) = \boldsymbol{A}\boldsymbol{T}_{\text{a}}(z)\boldsymbol{s}(z) + \boldsymbol{b}(z)x(z) \,, \tag{3a}$$

$$y(z) = \boldsymbol{c}^{\top}(z)\boldsymbol{s}(z) + d(z)x(z) \,, \tag{3b}$$

where $\boldsymbol{s}(z) \in \mathbb{C}^L$ is the time-dependent energy being transferred over each discrete direction; $\boldsymbol{T}_{\text{a}}(z) \in \mathbb{C}^{L\times L}$ is a diagonal matrix containing the propagation delays $\boldsymbol{\tau}_{\text{a}} \in \mathbb{N}^L$, each of which models the time it takes for energy to traverse the corresponding discrete direction; $\boldsymbol{A} \in \mathbb{R}^{L\times L}$ is the reflection matrix, governing the reflection of energy from each direction to the others; $\boldsymbol{b}(z) \in \mathbb{C}^L$, $\boldsymbol{c}(z) \in \mathbb{C}^L$, $d(z) \in \mathbb{C}$ are the input, output, and direct filters, which apply scaling and delay as defined in the following; $x(z) \in \mathbb{C}$, $y(z) \in \mathbb{C}$ are the input and output signals, respectively. The complete model is illustrated in Fig. 2.

The transfer function of (3) is given by [18]

$$E(z) = \boldsymbol{c}^{\top}(z)\left[\boldsymbol{I} - \boldsymbol{A}\boldsymbol{T}_{\text{a}}(z)\right]^{-1}\boldsymbol{b}(z) + d(z) \tag{4a}$$

$$= \boldsymbol{c}^{\top}(z)\,\boldsymbol{T}_{\text{a}}(z)^{-1}\left[\boldsymbol{T}_{\text{a}}(z)^{-1} - \boldsymbol{A}\right]^{-1}\boldsymbol{b}(z) + d(z) \,. \tag{4b}$$

While this transfer function strongly resembles the Feedback Delay Network (FDN) form, note that (4b) differs slightly from the form in [19, 20, 21], due to the presence of $\boldsymbol{T}_{\text{a}}(z)^{-1}$ after $\boldsymbol{c}^{\top}(z)$.

[1] The discretized time steps only need to be small enough to capture energy variations, i.e. the sample rate may be much lower than would be required for auralization.

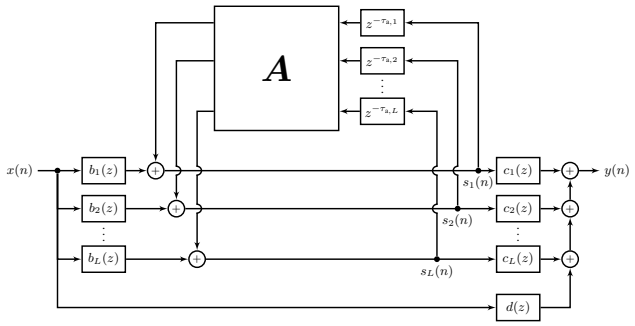


Figure 2: Block diagram of a TD-ART model, as expressed in (3).

The evaluation of the parameters in (3), and the operation of the TD-ART model, is divided in three steps:

1. The energy reflection between each pair of discrete directions is computed, forming the matrix $\boldsymbol{A}$. This matrix holds all of the information on the surface patches' absorption and scattering coefficients, as well as their form factors[2]. The set of propagation delays $\boldsymbol{\tau}_\mathrm{a}$ is also estimated, based on the distance covered by each discrete direction.
2. Rays are traced from the sound source position to evaluate how much energy is injected into each discrete direction and at what time, providing the input scaling weights $\boldsymbol{b}\in\mathbb{R}^L$ and input delays $\boldsymbol{\tau}_\mathrm{b}\in\mathbb{N}^L$. Rays are also traced from the listener position, providing the output scaling weights $\boldsymbol{c}\in\mathbb{R}^L$ and output delays $\boldsymbol{\tau}_\mathrm{c}\in\mathbb{N}^L$. If there is a direct line of sight between the source and listener, the direct component's energy and delay ($d\in\mathbb{R}$, $\tau_\mathrm{d}\in\mathbb{N}$) are also evaluated.
3. The input operators $\boldsymbol{b}$ and $\boldsymbol{\tau}_\mathrm{b}$ are used to initialize the acoustic energy carried by each discrete direction at different points in time, $\boldsymbol{s}(n)\in\mathbb{R}^L$. This is then iteratively propagated ($\boldsymbol{\tau}_\mathrm{a}$) and reflected ($\boldsymbol{A}$) until the desired time range is covered or the desired level of diffusion is reached. Finally, $\boldsymbol{s}(n)$ is scaled by $\boldsymbol{c}$ and delayed by $\boldsymbol{\tau}_\mathrm{c}$ to obtain the energy contributions at the listener position. The sum of these contributions and the direct component gives the energy response over time.

Note that all of the time delay operations may make use of fractional delays, as well as time spreading [17] in the form of Finite Impulse Response (FIR) filters. The method proposed in this paper extends to both of these scenarios, as it is applied in the z domain.

One of the main advantages of ART is that once $\boldsymbol{A}$ and $\boldsymbol{\tau}_\mathrm{a}$ have been computed, simulations can be run efficiently for different source and listener positions. This includes situations with multiple sources and/or receivers in the same scene. Frequency-domain ART models have, in fact, been shown to run in real time [23, 24]. Another advantage is that ART is particularly suitable for late reberberation rendering, at the expense of directionally approximative early reflections [12]. The Ray Tracing Method suffers from late-response uncertainty and "vanishing" energy [1], and the Image Source Method from exponential complexity as the desired reflection order increases [6, 1], but the accuracy and cost of ART do not vary over the response's duration. Both of these aspects – a position-independent foundation and an aptness for late reverberation – make ART particularly attractive for the prospect of common-slopes analysis.

[2]a.k.a. shape factors or view factors, these are the factors which governs radiance transfer [22, p. 38]. They can be interpreted as the patches' relative "view" of each other.

2.3. Modal decomposition

Any Linear Time-Invariant (LTI) system may be represented as a state-space model. Alternatively, the transfer function of a discrete-time LTI system can be characterized in the z domain as [18]

$$E(z) = d(z) + \sum_{i=1}^{M} \frac{\rho_i}{1-\lambda_i z^{-1}}\,, \tag{5}$$

where M is the number of modes, $\lambda_i, \rho_i \in \mathbb{C}$ are the pole and residue of the i^{th} mode respectively, and $d(z)$ is the system's *direct* (i.e. non-recursive) component. The corresponding time-domain impulse response $\epsilon(n)$ can be equivalently described as a sum of complex oscillators, or *modes* [6]. Neglecting the direct component for simplicity,

$$\epsilon(n) = \sum_{i=1}^{M} \rho_i \lambda_i^n = \sum_{i=1}^{M} |\rho_i||\lambda_i|^n e^{j(\angle\rho_i + n\angle\lambda_i)}\,, \tag{6}$$

where $j=\sqrt{-1}$, and $\angle(\cdot), |\cdot|$ respectively indicate the phase and magnitude of a complex number.

In this paper, we apply the modal decomposition to the delay state space form similar to (3), as described in [20, 21]. While discussing the pole estimation in detail is outside the scope of this paper [25], the estimation of residues is presented in the following, as it will prove relevant in Section 3. For the given system, the residues are defined as [20]

$$\rho_i = \frac{\boldsymbol{c}^\top(\lambda_i)\boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\,\mathrm{adj}\left(\boldsymbol{P}(\lambda_i)\right)\boldsymbol{b}(\lambda_i)}{\mathrm{tr}\left(\mathrm{adj}\left(\boldsymbol{P}(\lambda_i)\right)\boldsymbol{P}'(\lambda_i)\right)}\,, \tag{7}$$

where $\boldsymbol{P}(z) = \left[\boldsymbol{T}_\mathrm{a}(z)^{-1} - \boldsymbol{A}\right]$, $\mathrm{tr}(\boldsymbol{P})$ is the trace of $\boldsymbol{P}$, $\mathrm{adj}(\boldsymbol{P})$ is the adjugate of $\boldsymbol{P}$, and $\boldsymbol{P}'(z) = \frac{\mathrm{d}\boldsymbol{P}(z)}{\mathrm{d}z}$. Since $\boldsymbol{A}$ is not a function of z, we can say

$$\boldsymbol{P}'(z) = \left[\boldsymbol{T}_\mathrm{a}(z)^{-1}\right]'\,. \tag{8}$$

Note that $\boldsymbol{T}_\mathrm{a}(z)$ is diagonal, therefore $\boldsymbol{P}'(z)$ is diagonal as well. If no time spreading is present in the recursion, the diagonal elements of $\boldsymbol{T}_\mathrm{a}(z)^{-1}$ are $\left[z^{\tau_{\mathrm{a},1}}, z^{\tau_{\mathrm{a},2}}, \ldots, z^{\tau_{\mathrm{a},L}}\right]$, and those of $\boldsymbol{P}'(z)$ are $\left[\tau_{\mathrm{a},1} z^{\tau_{\mathrm{a},1}-1}, \tau_{\mathrm{a},2} z^{\tau_{\mathrm{a},2}-1}, \ldots, \tau_{\mathrm{a},L} z^{\tau_{\mathrm{a},L}-1}\right]$.

The decomposition method proposed in [21] relies on the observation that, due to the nature of poles, $\det\left(\boldsymbol{P}(\lambda_i)\right) = 0$, and $\mathrm{adj}\left(\boldsymbol{P}(\lambda_i)\right)$ can therefore be expressed as the outer product

$$\mathrm{adj}\left(\boldsymbol{P}(\lambda_i)\right) = \boldsymbol{v}_i \boldsymbol{w}_i^H\,, \tag{9}$$

where $\boldsymbol{w}_i^H$ is the Hermitian transpose of $\boldsymbol{w}_i$. For the same reason, we have $\boldsymbol{P}(\lambda_i)\boldsymbol{v}_i = \boldsymbol{0}$ and $\boldsymbol{w}_i^H \boldsymbol{P}(\lambda_i) = \boldsymbol{0}^\top$, meaning that the vectors $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$ can be found by solving the generalized eigenvalue problems

$$\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\boldsymbol{v}_i\,, \tag{10a}$$

$$\boldsymbol{w}_i^H \boldsymbol{A} = \boldsymbol{w}_i^H \boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\,, \tag{10b}$$

where they respectively act as left and right eigenvectors.

Combining (7)–(9), and leveraging the diagonal nature of $\boldsymbol{P}'(z)$, one can find that

$$\rho_i = \frac{\boldsymbol{c}^\top(\lambda_i)\boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\boldsymbol{v}_i\boldsymbol{w}_i^H\boldsymbol{b}(\lambda_i)}{\boldsymbol{w}_i^H\left[\boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\right]'\boldsymbol{v}_i}\,. \tag{11}$$

The modal decomposition of (3) can thus be summarized as [21]:

1. Estimation of poles λ_i using the Ehrlich-Aberth Iteration (EAI) method[3] [25];
2. Estimation of left and right eigenvectors $\boldsymbol{v}_i, \boldsymbol{w}_i$ through a generalized eigenvalue problem;
3. Estimation of residues ρ_i based on $\boldsymbol{v}_i, \boldsymbol{w}_i$ using (11).

3. PROPOSED METHOD

The following sections present our proposed method for model complexity reduction. The most prominent aspect of the reduction is the fact that only modes with real, positive poles are selected, as explained in Section 3.1. Section 3.2 highlights which elements of the model need to be computed at runtime for interactive applications, and discusses the physical significance of the decomposition. Finally, Section 3.3 discusses some considerations that can facilitate the pole-finding process.

3.1. Pole selection

Let us consider the energy response's short-time average, as expressed in (1), in light of its modal decomposition in (6). Recall that all complex poles of a real system appear in conjugate pairs [18], such that the imaginary parts of their modes cancel out, leaving only the oscillating real part. Real, negative poles also lead to oscillating modes, as λ_i^n changes sign at each time sample n. Removing oscillatory components is in fact the purpose of the short-time averaging, which is meant to characterize the "steady" decay of energy over time [6]. If the averaging window is longer than all oscillation periods, the only modes contributing to the energy decay are those from real, positive poles, a subset of $\boldsymbol{\lambda} \in \mathbb{C}^M$ which we will denote as $\overline{\boldsymbol{\lambda}} \in \mathbb{R}^N_{>0}$. Thus, the energy decay is

$$\overline{\epsilon}(n) = \langle \epsilon(n) \rangle \approx \sum_{i=1}^{N} \overline{\rho}_i \overline{\lambda}_i^n\,, \tag{12}$$

where $\overline{\rho}_i$ denotes the residues related to the subset of poles $\overline{\lambda}_i$. Comparing (2) and (12), it becomes apparent that the real, positive poles $\overline{\lambda}_i$ are identical to the slope decay rates δ_i, while the corresponding residues $\overline{\rho}_i$ take the role of $\boldsymbol{\chi}_i^2$:

$$\boldsymbol{\chi}_i^2 = \overline{\rho}_i\,, \tag{13a}$$

$$\delta_i = \overline{\lambda}_i\,. \tag{13b}$$

The equivalence of these quantities also extends to their physical significance. Recall that the decay rates δ_i are only dependent on the room geometry and materials [9], with the amplitudes $\boldsymbol{\chi}_i^2$ scaling each term based on the combined positions and orientations of sound sources and listeners, $\boldsymbol{\phi}$. As discussed in Section 2.2, the ART parameters $\boldsymbol{A}$ and $\boldsymbol{T}_\mathrm{a}(z)$ are only dependent on the room geometry and materials – therefore, the same can be said about the system poles λ_i. Conversely, the residues carry information on $\boldsymbol{\phi}$, as they are dependent on ART's "injection" and "detection" operators $\boldsymbol{b}$ and $\boldsymbol{c}$. This aspect is explored in the following section.

[3] Alternatively, via eigendecomposition of the state transition matrix.

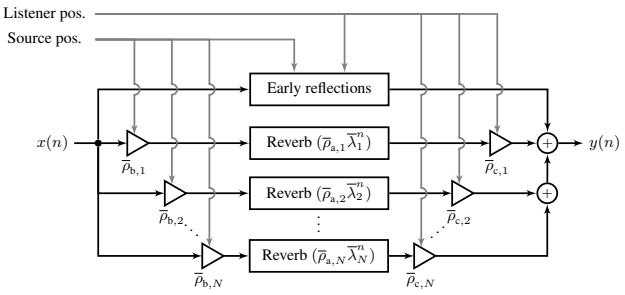


Figure 3: Block diagram of a common-slopes reverberator [9] with parameters set using the proposed method. Gray lines indicate positional dependency data, while black lines indicate audio signals. Each slope is controlled by a real, positive pole $\overline{\lambda}_i$ and the relative residue $\overline{\rho}_i = \overline{\rho}_{\mathrm{a},i}\overline{\rho}_{\mathrm{b},i}\overline{\rho}_{\mathrm{c},i}$, with the terms $\overline{\rho}_{\mathrm{b},i}$ and $\overline{\rho}_{\mathrm{c},i}$ being controlled by the source and listener position, respectively. Note that the audio signals in this figure, including the input and output signals, are defined in the energy domain – not pressure.

3.2. Residue estimation

The residue formulation in (11) can be expressed as the product of three scalar terms:

$$\rho_i = \rho_{\mathrm{a},i}\,\rho_{\mathrm{b},i}\,\rho_{\mathrm{c},i}\,, \tag{14}$$

where

$$\rho_{\mathrm{a},i}^{-1} = \boldsymbol{w}_i^H\left[\boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\right]'\boldsymbol{v}_i\,, \tag{15a}$$

$$\rho_{\mathrm{b},i} = \boldsymbol{w}_i^H\boldsymbol{b}(\lambda_i)\,, \tag{15b}$$

$$\rho_{\mathrm{c},i} = \boldsymbol{c}^\top(\lambda_i)\boldsymbol{T}_\mathrm{a}(\lambda_i)^{-1}\boldsymbol{v}_i\,. \tag{15c}$$

Note that $\rho_{\mathrm{a},i}^{-1}$, known in the literature as the *undriven residue* [20], does not depend on the source or listener positions. Of the other two terms, $\rho_{\mathrm{b},i}$ only depends on the source position, and $\rho_{\mathrm{c},i}$ only on the listener position. If an interactive model is required, these scalar terms can be applied separately, as shown in Fig. 3. Doing so, $\rho_{\mathrm{b},i}$ and $\rho_{\mathrm{c},i}$ will only need to be updated if and when[4] the respective entity moves. This separation of the components based on their role in the acoustic scene is possible thanks to the combination of ART operators' significance from the physical modeling perspective, and the equivalence in (13a) from the system theory perspective.

It is worth noting the role of left and right eigenvectors in (15). As anticipated in Section 1, the eigenvectors of a state-space model are related to modes' spatial distribution when the model's state variables are tied to space. Indeed, that is the case in ART, as the state signals $\boldsymbol{s}(n)$ model the acoustic energy at specific locations, and in specific directions (the "discrete directions" fundamental to ART). By their definitions, $\boldsymbol{b}$ and $\boldsymbol{c}$ encode respectively how much energy the sound source "injects" in each discrete direction, and how much energy the listener "detects" from each discrete direction. In light of this, (15b) may be interpreted as showing that the left eigenvectors $\boldsymbol{w}_i$ encode each mode's susceptibility to energy contributions from each discrete direction, and (15c) as showing that the right eigenvectors $\boldsymbol{v}_i$ encode each mode's contribution to each discrete direction.

[4] If the sonic agents move relatively slowly, $\rho_{\mathrm{b},i}$ and $\rho_{\mathrm{c},i}$ may be updated sporadically, i.e. at a rate lower than the reverberators' sample rate.

3.3. Pole estimation

As mentioned in Section 2.3, the more technical details of pole-finding are not discussed in this paper. Nevertheless, in this section, we discuss some aspects of our proposed method – and of ART models – that offer considerable advantages for the estimation of the desired poles.

The first consideration to make is that the number N of sought poles $\overline{\boldsymbol{\lambda}}$ is much lower than the total number M of poles $\boldsymbol{\lambda}$, i.e. the system order. As such, a full decomposition is not required. Moreover, as shown in [9] and corroborated by results in the following section, a very small number N of slopes is usually sufficient to achieve accurate energy decay.

Another important consideration regards the size and sparsity of $\boldsymbol{A}$. If the surface is divided into N_P surface patches, the number of discrete directions is at most[5] $L = N_\mathrm{P}^2$. However, the only nonzero elements of $\boldsymbol{A}$ are those relating pairs of discrete directions with a shared endpoint – in other words, the energy transferred from patch P_1 to P_2 may only be reflected into discrete directions that start from P_2. Since the number of discrete directions starting at any given patch can be no higher than N_P (one outgoing direction for each possible destination), the number of nonzero elements of $\boldsymbol{A}$ can be no higher than $L \cdot N_\mathrm{P} = N_\mathrm{P}^3$, whereas the size of $\boldsymbol{A}$ is $N_\mathrm{P}^2 \times N_\mathrm{P}^2$.

If the pole search is conducted via eigen-decomposition of the state-transition matrix [21], the highly sparse nature of the matrix can be leveraged to find the desired subset of poles with low memory requirements [26]. In this case, the right eigenvectors are found simultaneously [21], and (10b) can be used to find their left counterparts. Alternatively, the EAI approach [25, 20] can be modified to focus on real, positive poles, and both left and right eigenvectors can be found with (10). The discussion and comparison of different approaches are left for future work.

3.4. Implementation

The proposed method's implementation is summarized as follows.

1. Evaluation of the TD-ART recursive operators $\boldsymbol{A}$ and $\boldsymbol{\tau}_\mathrm{a}$ (performed offline);
2. Modal decomposition of the TD-ART model, providing the most prominent real, positive poles $\overline{\lambda}_i$, the relative eigenvectors v_i, w_i, and the position-independent residue terms $\overline{\rho}_{\mathrm{a},i}$ (performed offline);
3. Ray-tracing from the source and listener positions[6], to evaluate the TD-ART position-dependent operators $\boldsymbol{b}(z)$, $\boldsymbol{c}(z)$, and $d(z)$ (performed at runtime, sporadically);
4. Computation of the position-dependent residue terms $\overline{\rho}_{\mathrm{b},i}$, $\overline{\rho}_{\mathrm{c},i}$ as in (15) (performed at runtime, sporadically);
5. A set of time-invariant resonators is driven using the dynamic input-output weights $\overline{\rho}_{\mathrm{b},i}$, $\overline{\rho}_{\mathrm{c},i}$ (runtime).

Note that, in situations with multiple sources and/or receivers, a single set of resonators can be shared by all agents.

[5] If a pair of patches has no mutual visibility, there can be no energy transfer, and the discrete direction is not included in the model.

[6] Only one tracing step, i.e. one reflection order, is required for each agent. These tracing operations may make use of directional transfer functions, if desired.

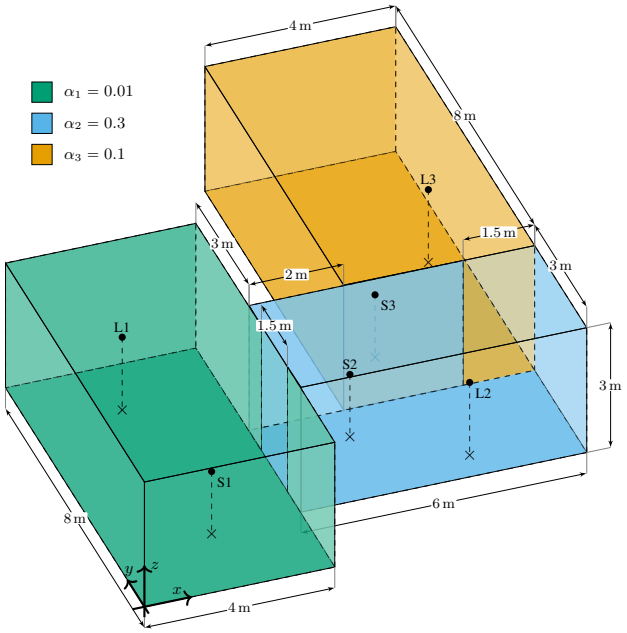


Figure 4: The environment used in the presented tests. Source and listener positions are also reported in Table 1.

4. RESULTS

The proposed method was tested by modeling an environment consisting of three coupled rooms with different absorption coefficients, $\alpha_1 = 0.01$, $\alpha_2 = 0.3$, and $\alpha_3 = 0.1$, as illustrated in Fig. 4. These are frequency-independent coefficients, and a broadband analysis was carried out, leaving the test of multi-band applications for future work. Nine configurations were compared, with three source positions and three listener positions, reported in Table 1 and shown in Fig. 4. The surface was divided into 70 patches with a maximum edge length of 3 m, producing a TD-ART model with 2068 discrete directions (having excluded those with no visibility), with a sample rate of 500 Hz. Air absorption was not modeled in any of the responses, and neither was edge diffraction. Time spreading [17] was applied to the input and output operators, but not to the propagation delays.

Fig. 5 shows the Energy Decay Curves (EDCs) of the proposed common-slopes model with different numbers of slopes for three of the tested source-listener configurations. In all configurations, four slopes appear to be sufficient to achieve a very close match with TD-ART. Expectedly, the common-slopes model with only one slope does not capture the early behaviour correctly. The abrupt decay it displays at the start of Fig. 5c is due to the line-of-sight component, which is not present in the other two S-L configurations shown here.

Fig. 6 shows how the residue values change for different positions of the source and listener, for the three slopes with the longest T_{60}. The source position is fixed in each subplot, while the listener position is moved forming a grid with 0.2 m spacing, at a height of 1.75 m. As might be expected, the slope with the longest reverberation time has highest amplitude when both the source and the listener are in the most reverberant room (room 1). The same can be said of the slope with the second-longest reverberation time being tied to the second-most reverberant room (room 3). The third

Table 1: *Source and listener positions used in the presented tests.*

	S1	S2	S3	L1	L2	L3
x (m)	2.1	5.8	7.2	1.9	7.8	9.3
y (m)	1.9	4.1	6.5	6.5	2.7	9.2
z (m)	1.5	1.5	1.5	1.75	1.75	1.75

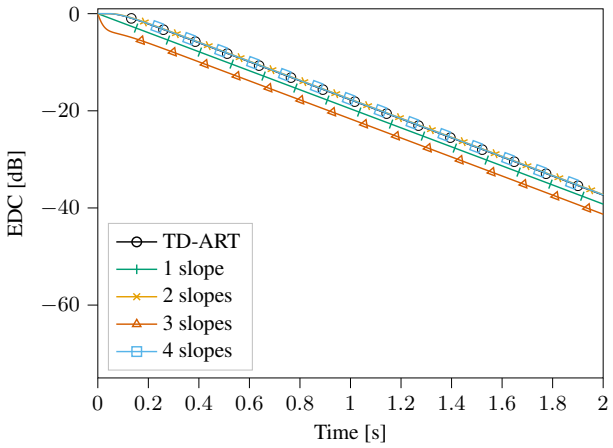
most reverberant slope seems to be only loosely tied to room 2 – this is not too surprising, given the central location of room 2, which makes it at least partially visible from most positions in the other two.

In addition to the above, some observations relative to the combinations of source and listener positions can be made. For example, compare the sub-figures 6g and 6h. As already stated, it can be seen that the amplitude of the most reverberant mode – Fig. 6g – is largest when the listener is in room 1. However, said amplitude is very low when compared to that of the second-most reverberant mode – Fig. 6h – even when the listener is in room 1, and especially so at the positions from which room 3 is visible through the doorway. A similar observation can be made comparing Figs. 6d and 6f.

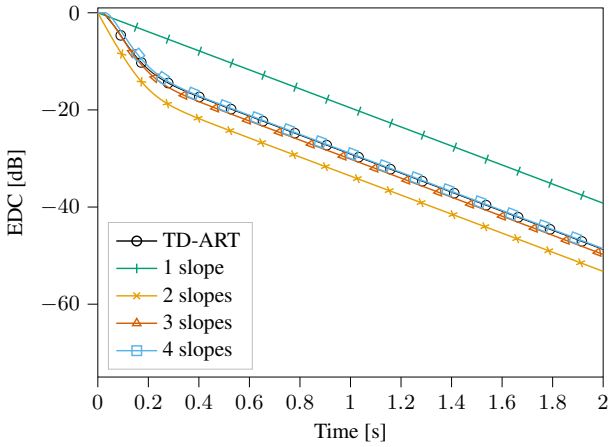
5. CONCLUSIONS

In this paper, we presented a method to retrieve both the decay rate and amplitude parameters for a common-slopes reverberator from the modal decomposition of an TD-ART model. By leveraging the physical significance of the model's eigenvectors, it is possible to compute the position-dependent parameters of the model at interactive speeds, for arbitrary source and listener positions. In particular, we showed that the position-dependent amplitude of slopes can be separated into three components: one that depends only on the source position, one that depends only on the listener position, and one that depends on neither; these terms can be updated independently based on simulation requirements. Comparison of EDCs showed that both the slopes' decay rates and their positional dependency are correctly captured by the proposed model, achieving a good match even with a very small number of slopes.

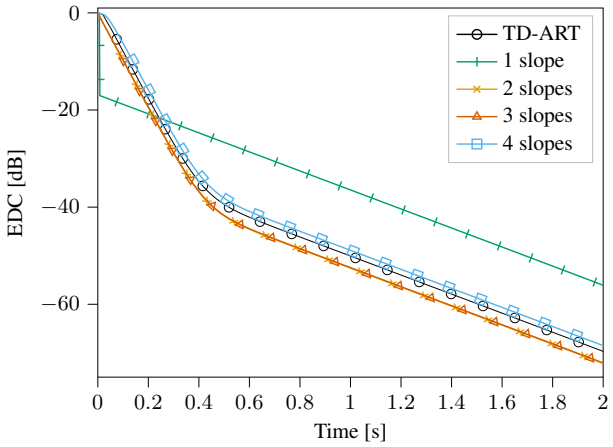
Future work will investigate important computational aspects associated to the modal decomposition, including leveraging the matrix sparsity to improve efficiency, as well as potential issues with numerical accuracy when very small poles are evaluated. Another potential avenue for further research concerns the potential of clustering similar modes by leveraging their physical significance, i.e. grouping slopes with similar decay rates only if their positional dependencies are also similar.



(a) S3-L1 configuration

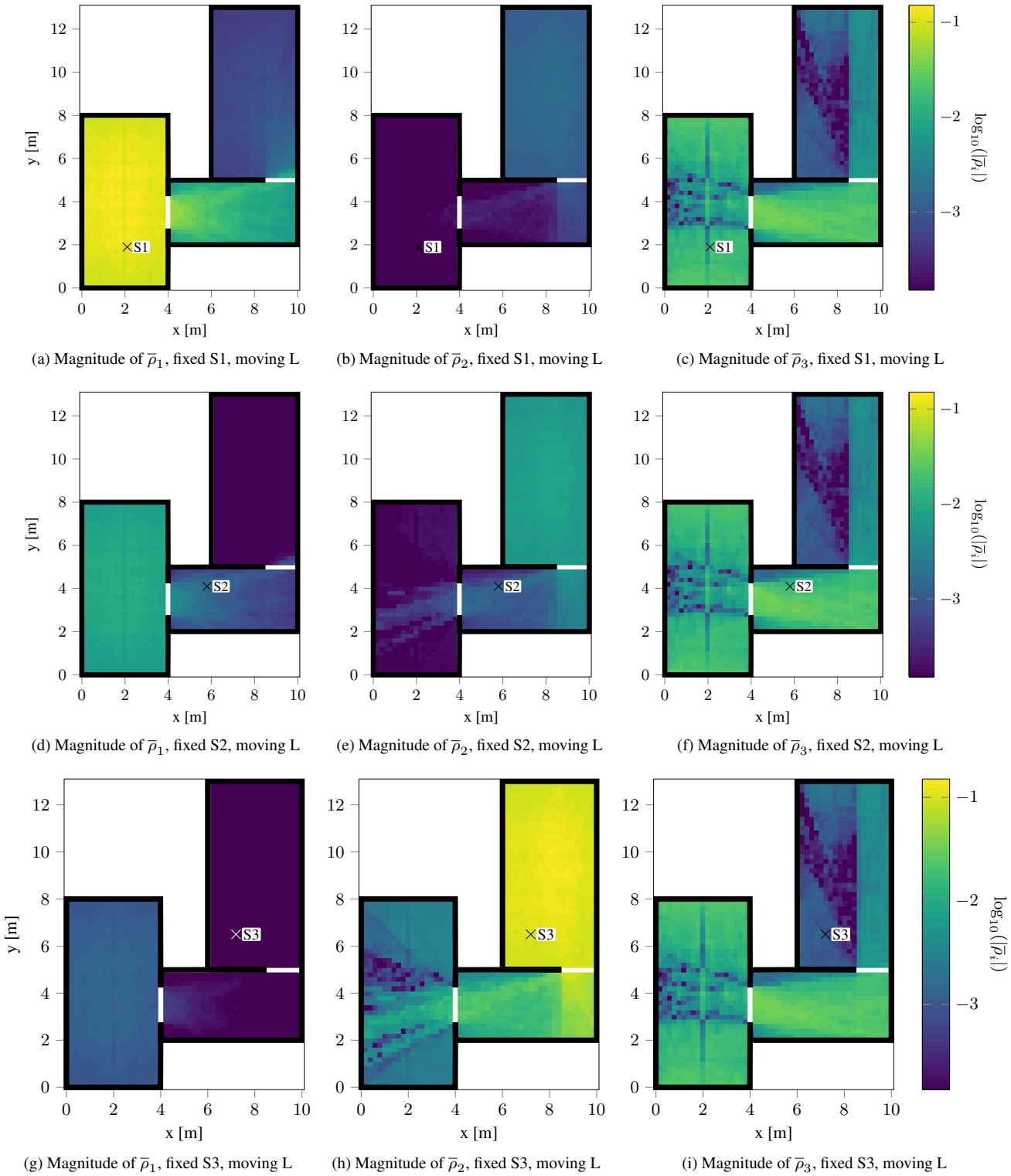


(b) S3-L2 configuration



(c) S3-L3 configuration

Figure 5: EDCs for three different source-listener configurations: S3-L1, S3-L2, and S3-L3 (see Table 1 and Fig. 4). Each plot compares the EDC of the proposed model with different numbers of slopes against TD-ART.



(a) Magnitude of $\overline{\rho}_1$, fixed S1, moving L (b) Magnitude of $\overline{\rho}_2$, fixed S1, moving L (c) Magnitude of $\overline{\rho}_3$, fixed S1, moving L

(d) Magnitude of $\overline{\rho}_1$, fixed S2, moving L (e) Magnitude of $\overline{\rho}_2$, fixed S2, moving L (f) Magnitude of $\overline{\rho}_3$, fixed S2, moving L

(g) Magnitude of $\overline{\rho}_1$, fixed S3, moving L (h) Magnitude of $\overline{\rho}_2$, fixed S3, moving L (i) Magnitude of $\overline{\rho}_3$, fixed S3, moving L

Figure 6: Magnitude of the residues $\overline{\rho}_i$ related to the three poles of largest magnitude. These plots were generated by placing the source at one of the positions in Table 1 and evaluating the residue for different positions of the listener. Subfigures (a), (d), and (g) show the residues related to the pole $\overline{\lambda}_1 = 0.991$ ($T_{60} = 3.06\,\text{s}$); (b), (e), and (h) those related to $\overline{\lambda}_2 = 0.956$ ($T_{60} = 0.62\,\text{s}$); (c), (f), and (i) those related to $\overline{\lambda}_3 = 0.89$ ($T_{60} = 0.24\,\text{s}$).

6. REFERENCES

[1] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer, 2008.

[2] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[3] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "More than 50 years of artificial reverberation," in *Proc. of the 60th Inter. Conf. of the Audio Eng. Soc.*, 2016.

[4] T. Potter, Z. Cvetković, and E. De Sena, "On the relative importance of visual and spatial audio rendering on VR immersion," *Front. in Sig. Proc.*, vol. 2, 2022.

[5] J.-M. Jot, R. Audfray, M. Hertensteiner, and B. Schmidt, "Rendering spatial sound for interoperable experiences in the audio metaverse," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–15.

[6] H. Kuttruff, *Room Acoustics, Fifth Edition*, CRC Press, 2009.

[7] N. Xiang, J. Escolano, J. M. Navarro, and Y. Jing, "Investigation on the effect of aperture sizes and receiver positions in coupled rooms," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3975–3985, 2013.

[8] C. Kirsch, T. Wendt, S. Van De Par, H. Hu, and S. D. Ewert, "Computationally-Efficient Simulation of Late Reverberation for Inhomogeneous Boundary Conditions and Coupled Rooms," *J. Audio Eng. Soc.*, vol. 71, no. 4, pp. 186–201, 2023.

[9] G. Götz, S. J. Schlecht, and V. Pulkki, "Common-slope modeling of late reverberation," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, 2023.

[10] G. Götz, T. Kerimovs, S. J. Schlecht, and Pulkki, V., "Dynamic late reverberation rendering using the common-slope model," in *6th Int. Conf. on Audio for Games of the Audio Eng. Soc.*, 2024.

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

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

[13] J. O. Smith, *Physical Audio Signal Processing*, `http://ccrma.stanford.edu/~jos/pasp/`, accessed (10/03/2024), online book.

[14] J. Botts and L. Savioja, "Spectral and pseudospectral properties of finite difference models used in audio and room acoustics," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 22, no. 9, pp. 1403–1412, 2014.

[15] H. Bai, G. Richard, and L. Daudet, "Late reverberation synthesis: From radiance transfer to feedback delay networks," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 23, no. 12, pp. 2260–2271, 2015.

[16] J.-D. Polack, "Modifying chambers to play billiards: the foundations of reverberation theory," *Acta Acustica united with Acustica*, vol. 76, no. 6, pp. 256 – 272, 1992.

[17] S. Siltanen, T. Lokki, and L. Savioja, "Efficient acoustic radiance transfer method with time-dependent reflections," in *Proc. of Meetings on Acoustics*, 2011, vol. 12.

[18] J. O. Smith, *Introduction to Digital Filters with Audio Applications*, `http://ccrma.stanford.edu/~jos/filters/`, accessed (10/03/2024), online book.

[19] D. Rocchesso, "Maximally diffusive yet efficient feedback delay networks for artificial reverberation," *IEEE Signal Processing Lett.*, vol. 4, no. 9, pp. 252 – 255, 1997.

[20] S. J. Schlecht and E. A. P. Habets, "Modal decomposition of feedback delay networks," *IEEE Trans. Signal Processing*, vol. 67, no. 20, pp. 5340–5351, 2019.

[21] S. J. Schlecht, M. Scerbo, E. De Sena, and V. Välimäki, "Modal excitation in feedback delay networks," *Submitted to IEEE Signal Processing Lett.*, Accessed: Mar. 30, 2024. [Online]. Available: `https://doi.org/10.36227/techrxiv.171173272.21484682/v1`.

[22] J. M. Palmer and B. G. Grant, *Art of Radiometry*, Berlin: SPIE, 2010.

[23] S. Siltanen, T. Lokki, and L. Savioja, "Frequency domain acoustic radiance transfer for real-time auralization," *Acta Acustica united with Acustica*, vol. 95, no. 1, pp. 106–117, 2009.

[24] E. Stavrakis, N. Tsingos, and P. Calamia, "Topological sound propagation with reverberation graphs," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 921–932, 2008.

[25] D. A. Bini and V. Noferini, "Solving polynomial eigenvalue problems by means of the Ehrlich–Aberth method," *Linear Algebra and its Applications*, vol. 439, no. 4, pp. 1130–1149, 2013.

[26] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*, SIAM, 1998.