

Common-Slope Modeling of Late Reverberation

Georg Götz , Sebastian J. Schlecht , *Senior Member, IEEE*, and Ville Pulkki 

Abstract—The decaying sound field in rooms is typically described by energy decay functions (EDFs). Late reverberation can deviate considerably from the ideal diffuse field, for example, in multiple connected rooms or non-uniform absorption material distributions. This paper proposes the common-slope model of late reverberation. The model describes spatial and directional late reverberation as linear combinations of exponential decays called common slopes. Its fundamental idea is that common slopes have decay times that are invariant across space and direction, while their amplitudes vary across both. We explore different approaches for determining the common slopes for large EDF sets describing different source-receiver configurations of the same environment. Among the presented approaches, the k-means clustering of decay times is the most general. Our evaluation shows that the common-slope model introduces only a small error between the modeled and the true EDF, while being considerably more compact than the traditional multi-exponential model. The amplitude variations of the common slopes yield interpretable room acoustic analyses. The common-slope model has potential applications in all fields relying on late reverberation models, such as source separation, dereverberation, echo cancellation, and parametric spatial audio rendering.

Index Terms—Anisotropy, inhomogeneity, late reverberation model, multi-exponential decay, sound energy decay.

I. INTRODUCTION

THE sound field in a room can be described with room impulse responses (RIRs), which represent the time domain transfer functions for specific source-receiver configurations. Due to reflections from the room boundaries, the sound field is a superposition of multiple sound waves with distinct amplitudes, phases, and propagation directions [1]. Consequently, the sound pressure and energy decay over time vary within the room, and their spatial variations (inhomogeneity) can be described if RIRs are available for every possible source and receiver position. Additionally, directional variations (anisotropy) must be taken into account when considering directional sources and receivers.

Manuscript received 20 June 2022; revised 2 April 2023 and 8 June 2023; accepted 7 September 2023. Date of publication 20 September 2023; date of current version 20 October 2023. This work was supported in part by the Aalto University Doctoral School of Electrical Engineering and in part by the Academy of Finland under Grant 317341. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jon Barker. (Corresponding author: Georg Götz.)

Georg Götz and Ville Pulkki are with the Aalto Acoustics Lab, Department of Information and Communications Engineering, Aalto University, 00076 Aalto, Finland (e-mail: georg.gotz@aalto.fi; ville.pulkki@aalto.fi).

Sebastian J. Schlecht is with the Aalto Acoustics Lab, Department of Information and Communications Engineering, Aalto University, 00076 Aalto, Finland, and also with the Media Lab Department of Art and Media, Aalto University, 00250 Espoo, Finland (e-mail: sebastian.schlecht@aalto.fi).

Digital Object Identifier 10.1109/TASLP.2023.3317572

RIRs are often parameterized to obtain analytical and potentially compact representations of the source-receiver transfer functions. To this end, various parametric sound field models have been established, such as the description as a stochastic random process [2], [3], [4] or the decomposition into plane waves [5], [6], [7], spherical waves [7], [8], spherical harmonics [9], or room modes [10], [11], [12]. Such models are valuable for estimating the RIR of spaces using computational tools, or for describing the source-receiver transfer function in audio signal processing applications like source separation, dereverberation, echo cancellation, or spatial audio rendering.

RIRs are usually separated into direct sound, early reflections, and late reverberation [13]. Late reverberation is characterized by its large echo density [1] and can consequently be modeled as exponentially decaying noise with one or more decay rates [3], [14], [15]. Furthermore, late reverberation is usually assumed to be diffuse and isotropic, thus requiring the energy density to be uniform throughout the room and over all directions [1], [2]. Other stochastic late reverberation properties in diffuse and isotropic sound fields are derived by Badeau in the space, time, and frequency domain [2].

Various studies have empirically investigated inhomogeneous and anisotropic sound fields, and the following gives a short summary of those that deal with late reverberation variations. Previous investigations have often quantified late reverberation variations in terms of ISO 3382 parameters [16], [17] like reverberation time, early decay time, or clarity [18], [19], [20], [21], [22], [23], [24], [25]. Some studies have also explored variations of multi-exponential decay parameters [26], [27] or entire energy decay functions [28]. Overall, previous studies investigated both inhomogeneous [18], [19], [20], [23], [24], [25], [26] and anisotropic [21], [22], [29] late reverberation, and found effects of various magnitudes, highlighting that inhomogeneity and anisotropy should be considered when establishing a comprehensive late reverberation model.

Coupled rooms were the subject of various studies, some of which have established models for predicting their energy decay. For instance, Cremer and Müller [30, pp. 261 ff.] derived a model via the analysis of power balances in a diffuse sound field. Their model can predict the energy decay in each room, but does not account for position-dependent effects. A recent room acoustic simulator uses Cremer and Müller's model to render late reverberation in coupled rooms [31]. Luizard et al. [32] extended Cremer and Müller's work and proposed a parametric solution of the diffusion equation for modeling the inhomogeneous reverberation in coupled rooms. Their model features two exponential decays with spatially invariant decay rates, whose amplitudes are adapted according to the source-receiver and aperture-receiver distance.

Energy decay functions (EDFs) obtained from the Schroeder backwards integration procedure [33] are suitable descriptors of the decaying sound field when investigating late reverberation. Usually, modern sound decay analyses utilize a well-established model consisting of multiple exponentials and a noise term [15]. This model can analyze energy decays from various environments, including coupled rooms or geometries with considerably non-uniform absorption material distributions [15]. Due to its generality, the model has many degrees of freedom, namely two for each exponential (one amplitude and one decay time) and one for the noise amplitude. Consequently, the model may easily overfit when it is carelessly used to analyze large EDF sets consisting of spatially or directionally distributed EDFs.

Therefore, this article introduces the common-slope model of late reverberation. Its fundamental idea is that large EDF sets, whose EDFs correspond to different source-receiver configurations within the same environment, can be described with one common set of decay times. Consequently, all spatial and directional EDF variations are described solely in terms of exponential and noise amplitudes. The common-slope model of late reverberation significantly reduces the degrees of freedom in spatial and directional energy decay analysis, while introducing only a small error between the modeled and the true EDF.

The common-slope model is conceptually similar to the previously described parametric model by Luizard et al. [32] because it uses multiple decaying exponentials with fixed¹ decay times but variable amplitudes. In contrast to the model by Luizard et al., our model is descriptive and not predictive. It constitutes a concise parametric framework, allowing us to model inhomogeneous and anisotropic late reverberation. Our proposed model is also inspired by the common-acoustical-pole and residue (CAPR) model proposed by Haneda et al. [10], which can be used to interpolate and extrapolate room transfer functions by exploiting that room mode decay times are independent of the source-receiver configuration. More recently, the CAPR has been extended to leverage frequency-band-wise processing [34].

The common-slope model extends, and generalizes the previous two models regarding two aspects. Firstly, it requires fewer parameters than the CAPR model, because it is based on common EDF slopes as opposed to common modes. This property implicitly assumes a dense modal overlap, thus limiting our model's advantage over the CAPR model to frequencies above the Schroeder frequency. Secondly, our model extends Luizard et al.'s model, because it is applicable to general room geometries with inhomogeneous and anisotropic late reverberation.

The remainder of this article is structured as follows. Section II summarizes the background and some acoustic fundamentals. Section III derives the common-slope decay model and outlines different approaches for determining the common decay times. Section IV describes the two datasets that are used throughout this article. In Section V, the common-slope model is evaluated

on two datasets. Finally, Section VI discusses the results, and Section VII concludes the article.

II. BACKGROUND

A room impulse response (RIR) describes the combined effects on sound waves traveling from a sound source at position $\mathbf{x}_s = (x_s, y_s, z_s)$ to a receiver at position $\mathbf{x}_r = (x_r, y_r, z_r)$. For directional sound sources and receivers, it is also important to consider the propagation path's direction of departure (DOD) from the source and direction of arrival (DOA) at the receiver [35]. They are denoted by $\mathbf{\Omega}_s = (\phi_s, \theta_s)$ and $\mathbf{\Omega}_r = (\phi_r, \theta_r)$, respectively, where ϕ is the azimuth angle and θ is the elevation angle. For conciseness of notation, we combine source and receiver position and the propagation directions into the *source-receiver configuration* $\mathbf{x} = (\mathbf{x}_s, \mathbf{x}_r, \mathbf{\Omega}_s, \mathbf{\Omega}_r)$ throughout the remainder of this article.

RIRs can be described as a superposition of modes [1, pp. 82–88]. For a large number of modes, M , such a modal decomposition of the RIR is given as [1]

$$h(\mathbf{x}, t) = \sum_{m=1}^M \chi_m(\mathbf{x}) \tau_m(t), \quad (1)$$

where $\chi_m(\mathbf{x})$ describes how the mode amplitudes and phases vary for different source-receiver configurations, and $\tau_m(t)$ models the temporal mode shape as decaying sinusoids

$$\tau_m(t) = \cos\left(\frac{\omega_m t}{f_s}\right) \exp\left(-\frac{\delta_m t}{f_s}\right), \quad (2)$$

with the mode frequency ω_m and decay rate δ_m . Throughout this article we assume discrete time, i.e., t is the discrete-time sample index and f_s is the sampling frequency.

Equation (1) exhibits an important property that is central to this work. The temporal terms $\tau_m(t)$ do not depend on the source-receiver configuration \mathbf{x} . More precisely, they only depend on the room geometry and wall properties [1]. It is well understood that inhomogeneity can be fully described in terms of mode amplitude variations, thus not affecting the mode decay times [10]. Furthermore, the modal formulation can take anisotropy into account as well, which becomes important when considering directional sound sources or receivers.

For example, the source directivity affects how strongly individual room modes are excited. This phenomenon becomes evident when decomposing a source with arbitrary directivity into monopoles with different amplitudes and phases [36], [37]. Each monopole excites the same room modes, but with different amplitudes. As the sound field evoked by the directional source is a superposition of all monopole sound fields, the corresponding mode amplitudes are the sum of the individual monopole mode amplitudes. Consequently, the mode amplitudes will vary when the source directivity changes or a directional source rotates. However, the mode decay times are not affected by the source directivity, thus making it possible to assign a set of common mode decay times for variable source directivities and orientations.

¹Throughout this article, we use the term “fixed” for spatially and directionally invariant.

Similarly, directional receivers affect how strongly certain modes are measured. For example, a directive microphone focuses the measurement on modes that coincide with the microphone's sensitivity, while suppressing modes from other directions. Consequently, the mode amplitudes change with varying receiver directivity and orientation. The mode decay times are unaffected by the amplification or suppression of modes due to the receiver directivity. Therefore, it is possible to choose a set of common mode decay times for describing sound energy decays with variable receiver directivities and orientations.

To summarize, the temporal terms $\tau_m(t)$ are fixed for a given room geometry and wall properties, whereas mode amplitudes $\chi_m(\mathbf{x})$ describe how the RIR varies for different source-receiver configurations. We will refer to this property from now on as *common-decay property (CDP)*. The CDP was previously exploited by Haneda et al. [10], who used it in their common-acoustical-pole and residue (CAPR) model to interpolate between RIRs.

III. COMMON-SLOPE DECAY ANALYSIS

The first part of this section introduces the *common-slope model* for sound energy decay analysis. It is based on modeling sound energy decays of multiple source-receiver configurations with a set of common decay times. In other words, the common-slope model applies the CDP to sound energy decay analysis. The second part of this section describes different approaches for determining the common decay times.

A. The Common-Slope Model

The Schroeder backwards integration procedure [33] is commonly used to describe the sound energy decay in rooms and obtain smooth energy decay functions (EDFs), which are also called energy decay curves (EDCs) when presented graphically. EDFs are usually calculated in frequency bands, e.g., octave or fractional octave bands [16], [17]. Throughout this article, we work with EDFs in octave bands. Following the Schroeder backwards integration procedure [33], an EDF can also be calculated directly from a (bandlimited) RIR as

$$d(\mathbf{x}, t) = \sum_{l=t}^L h^2(\mathbf{x}, l), \quad (3)$$

where L is the number of samples in the EDF. EDFs are often normalized with the total EDF energy $E(\mathbf{x}) = \sum_{l=1}^L h^2(\mathbf{x}, l)$. However, as this article deals with inhomogeneity and anisotropy, the normalization will not be applied here.

By squaring (1) and time-averaging over the cosine terms, one obtains a modal expression for the sound energy decay [1], [38]:

$$\overline{d(\mathbf{x}, t)} = \sum_{m=1}^M \chi_m^2(\mathbf{x}) \exp\left(-\frac{2\delta_m t}{f_s}\right). \quad (4)$$

If the individual mode decay rates δ_m do not differ considerably in the investigated frequency band, they may be approximated by their average value $\bar{\delta}$ [1]. The single-exponential decay model $\hat{d}_1(\mathbf{x}, t)$ of an RIR measurement at position \mathbf{x} is then

given as

$$\hat{d}_1(\mathbf{x}, t) = A_{1,\mathbf{x}} \exp\left(-\frac{\bar{\delta} t}{f_s}\right), \quad (5)$$

where

$$\bar{\delta} = \frac{1}{M} \sum_{m=1}^M 2\delta_m \quad \text{and} \quad A_{1,\mathbf{x}} = \sum_{m=1}^M \chi_m^2(\mathbf{x}). \quad (6)$$

The reverberation time RT_{60} , which describes how much time elapses until the sound energy in an enclosure has decreased by 60 dB [16], [17], can be obtained via

$$\text{RT}_{60} = -\frac{\ln(10^{-6})}{\bar{\delta}} = \frac{13.8}{\bar{\delta}}, \quad (7)$$

where $\ln(\cdot)$ denotes the natural logarithm. It can be seen from (5) that the exponential amplitudes $A_{1,\mathbf{x}}$ are the only part of the decay model that depend on the source-receiver configuration \mathbf{x} .

There are however many scenarios, where the δ_m within a frequency band are dissimilar. For example, previous studies showed that coupled rooms or geometries with non-uniform absorption material distributions exhibit sound decays with multiple decay rates, and how the decay rates and amplitudes can be predicted from diffuse field assumptions and power balance equations using knowledge about the room geometry, absorption distribution, coupling factor, and the decay rates of the uncoupled rooms [30], [32], [39], [40], [41]. From a modal perspective, the different decay rates in coupled rooms can be explained by the localization of modes, which means that some modes appear primarily in either one of the coupled rooms, while contributing less to the sound field of the other [42]. Similarly, in scenarios with non-uniform absorption distributions, modes can be grouped into non-grazing and grazing incidence, i.e., modes which involve the surfaces with differing absorption properties and surfaces which do not [40], [41]. Generally, the modes in (1) can be grouped into multiple distinct mode groups, yielding

$$h(\mathbf{x}, t) = \sum_{k=1}^{\kappa} \sum_{m \in \mathcal{M}_k} \chi_m(\mathbf{x}) \tau_m(t), \quad (8)$$

where \mathcal{M}_k denotes a set containing all mode indices of the k th mode group, and κ is the number of mode groups. By using (4) and (5) on every mode group and summing all contributions, a multi-exponential decay model can be established.

Decay models consisting of multiple exponentials and a noise term are frequently used to model sound decays in coupled rooms and geometries with non-uniform absorption material distributions [15], [27], [43]. In this article, we use the previously elaborated CDP to extend a well-established multi-exponential decay model [15] to account for different source-receiver configurations. We propose the *common-slope model*, which is given by

$$d_{\kappa}(\mathbf{x}, t) = N_{0,\mathbf{x}} \Psi_0(t) + \sum_{k=1}^{\kappa} A_{k,\mathbf{x}} [\Psi_k(t) - \Psi_k(L)], \quad (9)$$

with the decay kernel

$$\Psi_k(t) = \begin{cases} L - t, & \text{if } k = 0 \\ \exp\left(\frac{-13.8}{f_s T_k} t\right), & \text{if } k > 0 \end{cases} \quad (10)$$

In the model, T_k and $A_{k,\mathbf{x}}$ are the decay times and amplitudes of the k th mode group, respectively, $N_{0,\mathbf{x}}$ is the amplitude of the noise term, and the constant $-13.8 = \ln(10^{-6})$ ensures that the sound energy has decayed by 60 dB after T_k seconds. The decay times T_k are obtained from the decay rates analogously to (7). The number of mode groups κ is also called model order. The constant second term in the square brackets of (9) accounts for the finite upper limit of integration during the Schroeder backwards integration and can be neglected for large L [44].

Equation (9) can be written using matrix notation

$$\mathbf{d}_{\kappa,\mathbf{x}} = \Psi \mathbf{a}_{\mathbf{x}}, \quad (11)$$

where the modeled decay function $\mathbf{d}_{\kappa,\mathbf{x}}$, decay kernels Ψ , and amplitude values $\mathbf{a}_{\mathbf{x}}$ are given as

$$\mathbf{d}_{\kappa,\mathbf{x}} = [d_{\kappa}(\mathbf{x}, 1), d_{\kappa}(\mathbf{x}, 2), \dots, d_{\kappa}(\mathbf{x}, L)]^T, \quad (12a)$$

$$\mathbf{a}_{\mathbf{x}} = [N_{0,\mathbf{x}}, A_{1,\mathbf{x}}, A_{2,\mathbf{x}}, \dots, A_{\kappa,\mathbf{x}}]^T, \quad (12b)$$

$$\Psi = \begin{bmatrix} \Psi_0(1) & \Psi_1(1) & \dots & \Psi_{\kappa}(1) \\ \Psi_0(2) & \Psi_1(2) & \dots & \Psi_{\kappa}(2) \\ \dots & \dots & \dots & \dots \\ \Psi_0(L) & \Psi_1(L) & \dots & \Psi_{\kappa}(L) \end{bmatrix}, \quad (12c)$$

with $(\cdot)^T$ denoting the matrix transpose, $\mathbf{d}_{\kappa,\mathbf{x}} \in \mathbb{R}^{L \times 1}$, $\mathbf{a}_{\mathbf{x}} \in \mathbb{R}^{(\kappa+1) \times 1}$, and $\Psi \in \mathbb{R}^{L \times (\kappa+1)}$.

Equations (9) and (11) use the previously elaborated CDP because the mode group decay times T_k , which are equivalent to the decay times of the individual EDF slopes, do not vary with the source-receiver configuration \mathbf{x} . Therefore, we will call them *common decay times* and *common slopes* throughout the remainder of this article. In contrast, the decay and noise amplitudes $A_{k,\mathbf{x}}$ and $N_{0,\mathbf{x}}$ vary with the source-receiver configuration. More precisely, the $A_{k,\mathbf{x}}$ values model the mode amplitude variations, which were previously described by $\chi_m(\mathbf{x})$ in (1) and (8). The spatial dependence of the $N_{0,\mathbf{x}}$ values accounts for localized noise sources.

After determining the common decay times T_k , the remaining model parameters $A_{k,\mathbf{x}}$ and $N_{0,\mathbf{x}}$ need to be estimated for all source-receiver configurations. With fixed decay times, this endeavor simplifies to a constrained linear least-squares problem

$$\mathbf{a}_{\mathbf{x}} = \Psi^\dagger \mathbf{d}_{\mathbf{x}}, \quad (13)$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse and $\mathbf{d}_{\mathbf{x}}$ the measured EDF in vector notation analogous to (12a). To obtain meaningful solutions, the problem has to be constrained, such that $A_{k,\mathbf{x}} \geq 0$ and $N_{0,\mathbf{x}} \geq 0$.

B. Determination of Common Decay Times

The mode decay rates and amplitudes follow certain distributions and their relationship to EDFs was established by Kuttruff [1], [38]. While mode amplitude distributions have been analytically described for certain well-established scenarios like

shoebox rooms [1], less is known about the mode decay times. Given an arbitrary RIR set, the entire modal decomposition of the sound field is usually not available. Consequently, the determination of common decay times is in practice not as straightforward as described in the previous section.

The following section therefore investigates four approaches for determining common decay times in general settings. All approaches use an RIR set as their input and determine a common-slope set as their output. In the most general case, an RIR set may consist of RIRs with spatial variations (i.e., different source and receiver positions) and directional variations (i.e., different source directivities, source orientations, receiver directivities, and receiver orientations). In this article, the analyses are based on octave band processing, resulting in one set of common decay times per octave band, but the analyses can also be carried out in frequency bands with other bandwidths. However, for very low frequencies, where room modes are sparsely distributed in the frequency domain, it may be necessary to model individual modes analogously to the CAPR model by Haneda et al. [10].

Fig. 1 summarizes the common-slope analysis. Please note that the figure only features the last of the four approaches for determining common decay times because it is the most general, robust and therefore preferable approach.

1) *Average EDFs*: The first approach for obtaining common decay times is based on averaging EDFs of different source-receiver configurations

$$\hat{d}(t) = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} d(\mathbf{x}_n, t), \quad (14)$$

where \mathcal{N} is the number of source-receiver configurations. This averaging step is analogous to the spatial averaging for reverberation time measurements outlined by ISO 3382-2 [17]. The resulting average EDF $\hat{d}(t)$ can be analyzed with standard decay analysis approaches [45], [46], directly yielding the common decay times T_k .

However, the standard only covers averaging for single-exponential sound energy decays, and extending it to multi-exponential decays is not straightforward due to the need for manual selection of EDFs to average. In a coupled room scenario, the faster decay is often masked by the slower decay when measured in the more reverberant room, making it harder to observe both slopes. Fig. 2 illustrates this phenomenon on coupled room measurements from the Room Transition dataset [47], [48].

Averaging problems may occur if there are more single-slope measurements from the more reverberant room than double-slope measurements from the less reverberant room. This results in the faster decay being lost during averaging. While this issue can be resolved by only averaging EDFs from the less reverberant room, it is unclear how well this approach generalizes to more complex coupled geometries.

2) *Average $T_{k,\mathbf{x}}$ Values*: The second approach for obtaining common decay times is based on spatial or directional averaging of decay time values. In a first step, the sound energy decays of all RIRs are analyzed with a standard decay analysis method [45], [46]. This step yields an EDF parameterization in terms of

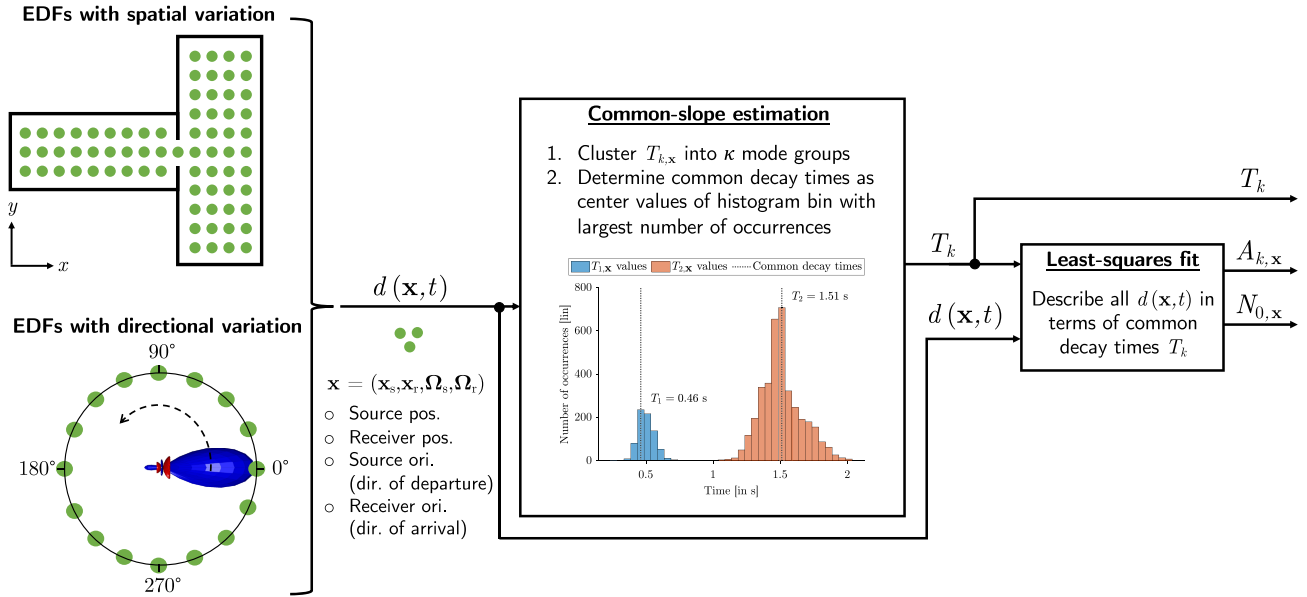


Fig. 1. Common-slope modeling of energy decay functions (EDFs) with spatial (i.e., source or receiver position varies, upper left part of the figure) and/or directional (i.e., source or receiver orientation varies, lower left part of the figure) variations. The green dots indicate different source-receiver configurations \mathbf{x} . Although not explicitly depicted in the figure, it is important to understand that source-receiver configurations describe the entire source-receiver configuration, i.e., the position and orientation of both. The common-slope analysis of an acoustic environment involves three steps. Firstly, the EDFs $d(\mathbf{x}, t)$ of all available source-receiver configurations are analyzed with a standard decay analysis algorithm [45], [46] to obtain configuration-dependent $T_{k,\mathbf{x}}$ values. Secondly, the $T_{k,\mathbf{x}}$ values are clustered into κ mode groups to obtain the common decay times T_k . Lastly, the common-slope amplitudes $A_{k,\mathbf{x}}$ of the common-slope model [c.f. (9) and (10)] are determined via a least-squares fit to the EDFs [c.f. (13)].

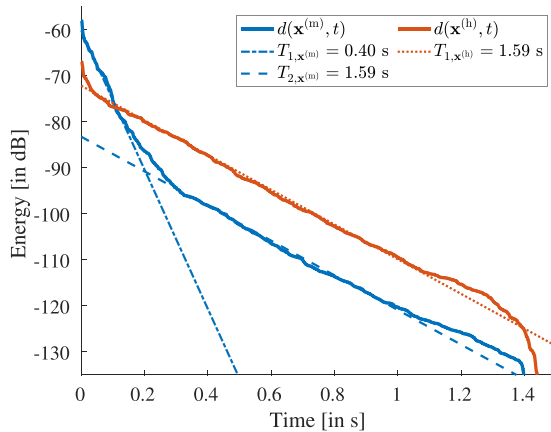


Fig. 2. Energy decay curves (EDCs) measured in a coupled room geometry consisting of a less reverberant meeting room (source-receiver configuration $\mathbf{x}^{(m)}$) and a more reverberant hallway (source-receiver configuration $\mathbf{x}^{(h)}$). In the hallway, only the slower decay time is observable. For both measurements, the sound source was located in the meeting room. The measurements are part of the Room Transition dataset [47], [48], with receiver positions at $x_r^{(m)} = 0$ cm and $x_r^{(h)} = 500$ cm, respectively [c.f. Fig. 4]. The EDCs were calculated in the 1 kHz octave band and the decay times were obtained with the DecayFitNet [45]. Both EDCs are based on measured RIRs with background noise, thus featuring the characteristic noise floor bump at their ends.

$T_{k,\mathbf{x}}$, $A_{k,\mathbf{x}}$, and $N_{0,\mathbf{x}}$, where the decay times $T_{k,\mathbf{x}}$ still exhibit inhomogeneity and anisotropy, as indicated by the subscript. In a second step, the decay time values $T_{k,\mathbf{x}}$ are averaged over the

source-receiver configuration,

$$T_k = \frac{1}{N} \sum_{n=1}^N T_{k,\mathbf{x}^{(n)}}, \quad (15)$$

resulting in the common decay times T_k . This approach is analogous to the reverberation time measurement procedure outlined by ISO 3382-1 and ISO 3382-2, in which reverberation time values are averaged over multiple source and receiver positions [16], [17].

For the coupled room example above, this approach features the same drawback as the previous one if single-slope decays from the more reverberant room are blindly analyzed with the same double-slope decay model that is used for the double-slope EDFs from the less reverberant room. The fitting algorithms will either return the same decay time twice or return an incorrect second decay time with near-zero decay amplitude. While this issue can be resolved by manually selecting and grouping decay time values, it is not feasible for large-scale automatic analyses.

3) *Choose T_k Values With Minimum Fitting Error on all EDFs:* Another approach for determining common decay times is to estimate $T_{k,\mathbf{x}}$ values for all source-receiver configurations and choose the combination that best fits all EDFs in the set. The $T_{k,\mathbf{x}}$ values can be estimated using a standard decay analysis method [45], [46], and candidate fits can be obtained with a constrained least-squares fit [c.f. (13)]. Finally, the goodness of fit can be evaluated for all $T_{k,\mathbf{x}}$ -EDF-combinations using an error metric like the mean squared error (MSE). However, this approach becomes impractical for large RIR sets due to the high number of combinations that need to be analyzed.

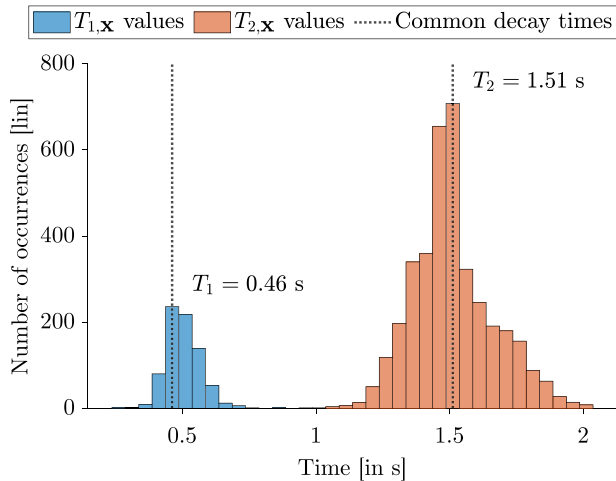


Fig. 3. K-means clustering of decay times (1 kHz octave band) in a coupled room geometry consisting of an acoustically-treated meeting room and a reverberant hallway. The decay times were obtained from simulations contained in the Extended Room Transition dataset (variant 1, c.f. Section IV). The DecayFitNet [45] was used for the decay analysis.

4) *K-Means Clustering of $T_{k,x}$ Values*: To overcome the issues with the previous approaches, decay times can be automatically clustered based on their mode group using the k-means clustering algorithm. This algorithm divides a set of data points into multiple clusters (determined by the number of mode groups, κ , in this case) based on their distance (e.g. squared Euclidean distance) to the cluster mean [49], [50].

To determine the common decay times using the k-means clustering approach, the EDFs of the RIR set are first analyzed using an established decay analysis method [45], [46]. The resulting $T_{k,x}$ values are then pooled and clustered into κ clusters, corresponding to the mode groups expected for the room geometry. The number of mode groups can be visually determined from a histogram of the $T_{k,x}$ values, where distinct clusters appear as separate groups. The largest histogram bin for each cluster is identified, and the centers of these bins correspond to the common decay times T_k . Fig. 3 illustrates this approach.

IV. DATASETS UNDER INVESTIGATION

We use two datasets throughout this article to derive the common-slope decay analysis, illustrate some of its concepts, and evaluate the proposed approach. Both datasets are based on the room geometry illustrated in Fig. 4.

The first dataset is part of the Room Transition dataset (RTD) by McKenzie et al. [47], [48], which contains higher-order Ambisonic room impulse responses of transitions between coupled rooms. The responses were measured with a coaxial loudspeaker (Genelec 8331 A) and a higher-order spherical microphone array (mh acoustics em32 Eigenmike). We limit our analyses to measurements from the “meeting room to hallway” transition, where the first room is an acoustically-treated meeting room ($4.6 \text{ m} \times 6.6 \text{ m} \times 2.8 \text{ m}$; volume: approx. 85 m^3) and the second room is a more reverberant hallway ($4.5 \text{ m} \times 18 \text{ m} \times 2.8 \text{ m}$; volume: approx. 227 m^3). The measured transition is 5 m long, centered

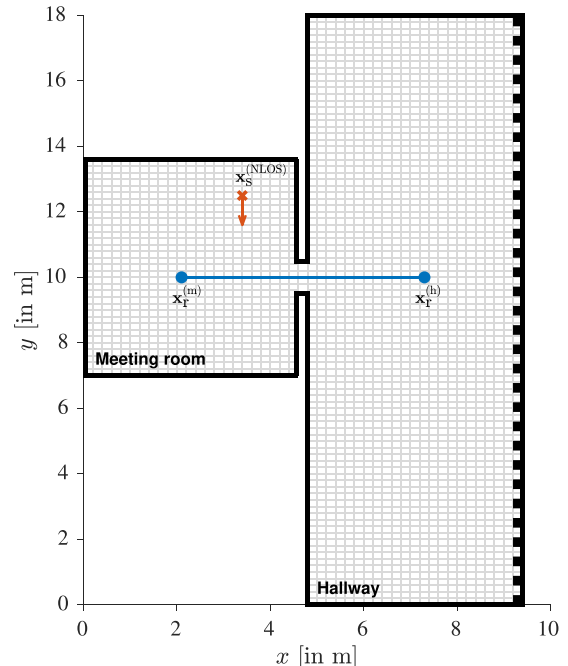


Fig. 4. Room geometry of the investigated Room Transition dataset (RTD) [47], [48] and Extended RTD (ERTD). The orange \times indicates the source position (no line-of-sight), the orange arrow indicates the main axis of the RTD’s loudspeaker, the blue dots indicate the start- and endpoints of the transition measured in the RTD (0.05 m resolution), the blue line indicates the transition, and the gray grid indicates the receiver grid in the ERTD (0.2 m resolution on the horizontal plane at 1.55 m height). The irregularities on the right hallway wall indicate the more absorbent wall material in the ERTD2 variant of the dataset.

around the door connecting the rooms, and features receiver positions every 0.05 m. Consequently, the transition consists of $\mathcal{N}^{(\text{RTD})} = 101$ RIRs in total. We only consider the “source in meeting room, no line-of-sight” (NLOS) configuration,² i.e., the sound source remains at the position $\mathbf{x}_s^{(\text{NLOS})}$ for all measurements. The energy decay functions of this transition feature a distinct double-slope characteristic, which is typical for coupled room geometries [27], [39].

The second dataset is the Extended Room Transition dataset (ERTD). It was compiled specifically for this article and contains finite-difference time-domain (FDTD) simulations. It extends the “meeting room to hallway” transition of the RTD by additional receiver positions in both rooms. The ERTD contains $\mathcal{N}^{(\text{ERTD})} = 2833$ RIRs in total. For all RIRs, the sound source position was chosen according to the “source in meeting room, no line-of-sight” (NLOS) configuration of the RTD. While the RTD only features receiver positions on a straight line between the rooms, the Extended Room Transition dataset samples the entire room geometry using a uniform grid with 0.2 m resolution. All simulations are performed on the horizontal plane at 1.55 m

²The terminology “no line-of-sight” was chosen to be consistent with [47]. Please note that the line-of-sight is only obstructed when the receiver is in the hallway, whereas other receiver positions feature a clear line-of-sight.

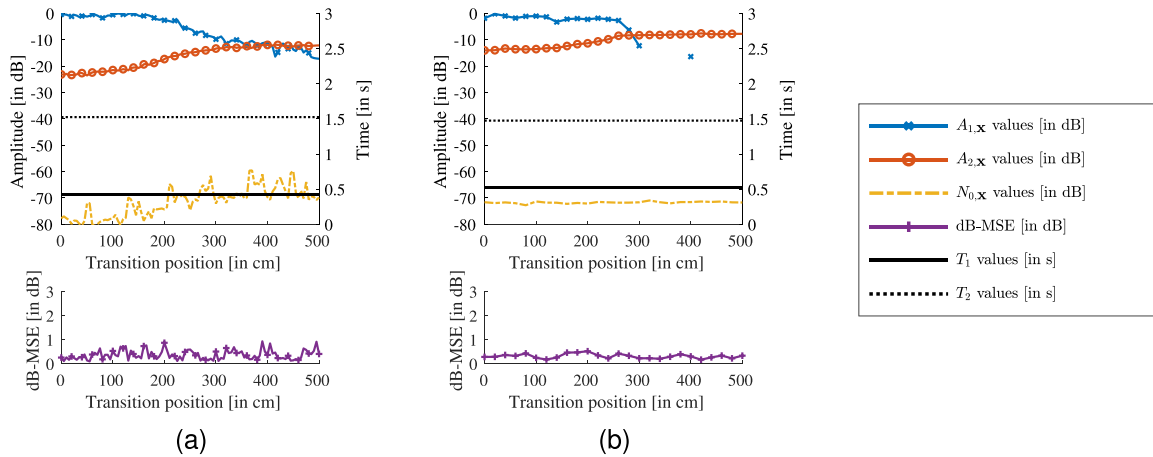


Fig. 5. Common-slope analysis results (1 kHz octave band) of a transition between coupled rooms. The investigated geometry consists of a meeting room and a more reverberant hallway (c.f. Section IV; the transition is indicated by the blue line in Fig. 4). The analyses are based on (a) measurements from the Room Transition dataset [47], [48] and (b) FDTD simulations from the Extended Room Transition dataset. The common decay times were obtained with the k-means clustering approach (c.f. Section III-B4). In both plots, the $A_{2,x}$ values indicate that the slope with the longer decay time T_2 is getting stronger while approaching and entering the more reverberant hallway. The common-slope model introduces only a small decibel-based mean squared error (dB-MSE) between modeled and true energy decay functions [c.f. (16)]. A perfect fit would yield a dB-MSE of 0 dB.

height. An open-source FDTD solver³ [51] (3D standard rectangular scheme, sampling frequency 80 kHz, omnidirectional soft source) was used for the simulations. The room geometry was modeled according to the geometry specifications described by McKenzie et al. [47]. We simulated two variants of the dataset. In variant ERTD1, the wall absorption properties were assigned to approximately match the corresponding measurements. The absorption coefficient was assigned uniformly to all walls in the respective rooms, where the meeting room walls exhibited an absorption coefficient $\alpha_{R1} = 0.12$, while the hallway was modeled less absorbent with $\alpha_{R2} = 0.042$. In variant ERTD2, a significantly higher absorption coefficient $\alpha_{R3} = 0.48$ is assigned to the right wall of the hallway, while all other surfaces are modeled analogous to ERTD1. We use ERTD2 in our evaluation to demonstrate that the common-slope model can also be used in environments with very non-uniform absorption distributions.

V. EVALUATION: COUPLED ROOMS AND NON-UNIFORM ABSORPTION DISTRIBUTION

A. Room Transition Along a Straight Line: Spatial Analysis

We first apply the common-slope analysis to RIRs from the room transition along a straight line, as indicated by the blue line in Fig. 4. Our analysis is based on the 1 kHz octave band of the RTD (omnidirectional channel) and ERTD1.

The following steps were carried out analogously for both RIR sets. The DecayFitNet [45] was used to analyze the EDFs $d(\mathbf{x}, t)$. The resulting $T_{k,x}$ were clustered using the k-means approach with $\kappa = 2$, as coupled room geometries usually have two mode groups and the histogram of the $T_{k,x}$ values showed two clear clusters. The common decay times for the RTD and the ERTD1 amounted to $\{T_1^{(\text{RTD})} = 0.43 \text{ s}, T_2^{(\text{RTD})} = 1.52 \text{ s}\}$ and $\{T_1^{(\text{ERTD1})} = 0.53 \text{ s}, T_2^{(\text{ERTD1})} = 1.48 \text{ s}\}$, respectively, indicating good agreement between measured and simulated sound

energy decays. T_1 and T_2 correspond to the decay times of the acoustically treated meeting room and the reverberant hallway, respectively. Finally, the decay and noise amplitudes $A_{k,x}$ and $N_{0,x}$ were determined via a linear least-squares fit [c.f. (13)].

Fig. 5(a) and (b) summarize all results of the common-slope analysis for the RTD and ERTD1, respectively. The common decay times are indicated by straight horizontal lines because they are by definition modeled constant along the entire transition.

The decay amplitudes $A_{1,x}$ and $A_{2,x}$ vary considerably with the transition position. Generally, $A_{2,x}$, which corresponds to the more reverberant hallway, is smaller for receiver positions in the meeting room and gradually increases during the transition through the door. The decay amplitudes $A_{1,x}$ exhibit an inverse effect, i.e., they decrease while transitioning from the meeting room to the hallway. However, the faster-decaying slope is quickly masked by the slower-decaying slope for receiver positions further in the hallway, thus making $A_{1,x}$ difficult to detect. Therefore, we deleted $A_{1,x}$ values below -40 dB from Fig. 5.

Interestingly, the $N_{0,x}$ values in Fig. 5(a) also vary considerably over the entire transition, likely due to a localized noise source during the measurements. After consulting with the authors of the RTD, the noise source was identified as a printer located in the hallway, which explains why the noise values there are consistently higher than in the meeting room. Although no explanation was found for the intermediate peaks, manual analysis of EDFs confirmed that the increased noise floor is part of the measurements and not due to fitting errors. To demonstrate the robustness of the common-slope analysis, a constant noise floor was added to the ERTD1 simulations. Fig. 5(b) shows that the $N_{0,x}$ values are

³Available at: <https://github.com/AaltoRSE/ParallelFDTD>, Accessed: 25 April 2022

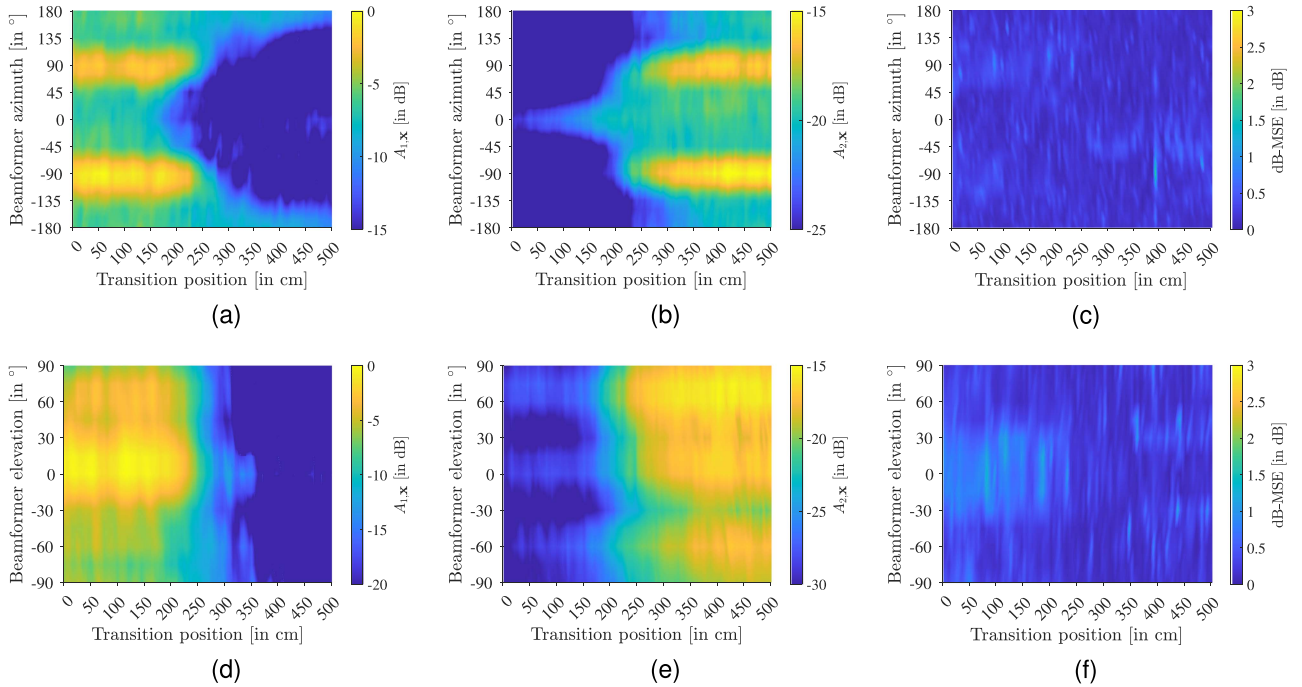


Fig. 6. Common-slope analysis results (1 kHz octave band) of a transition between coupled rooms. The analysis is based on measurements from the Room Transition dataset [47], [48] (c.f. Section IV; the transition is indicated by the blue line in Fig. 4). The investigated geometry consists of a meeting room ($T_1 = 0.43$ s) and a more reverberant hallway ($T_2 = 1.53$ s). Directional analysis results are obtained from the higher-order Ambisonic RIRs via beamforming. Plots (a) and (b) show the $A_{1,x}$ and $A_{2,x}$ values, respectively, for various azimuth angles and fixed elevation angle $\theta = 0^\circ$. Plots (d) and (e) show the $A_{1,x}$ and $A_{2,x}$ values, respectively, for various elevation angles and fixed azimuth angle $\phi = 90^\circ$. The directional analysis highlights the anisotropy of the late reverberant sound field and demonstrates how the two decay processes cross-fade during the transition between the rooms. The common-slope model introduces only a small decibel-based mean squared error (dB-MSE) between modeled and true energy decay functions [c.f. (16)], as depicted in Plots (c) and (f). A perfect fit would yield a dB-MSE of 0 dB.

estimated correctly as approximately constant over the entire transition.

The common-slope analysis does not introduce significant errors, as evidenced by the decibel-based mean squared error (dB-MSE) curves in Fig. 5. We define the dB-MSE as

$$\text{dB-MSE} = \frac{1}{L} \sum_{t=1}^L \left[d_{\kappa}^{(\text{dB})}(\mathbf{x}, t) - \hat{d}^{(\text{dB})}(\mathbf{x}, t) \right]^2, \quad (16)$$

where the true EDF $d^{(\text{dB})}(\mathbf{x}, t)$ and its common-slope model $\hat{d}_{\kappa}^{(\text{dB})}(\mathbf{x}, t)$ are represented on a logarithmic scale in dB.

The plots show that the dB-MSE is close to 0 dB with no considerable peaks along the transition. The average dB-MSE for the RTD and the ERTD1 amount to 0.37 dB and 0.31 dB, respectively, indicating that the common-slope decay model is a suitable description of the inhomogeneous energy decays featured in both datasets. The common-slope model reduces the number of source-receiver-configuration-dependent parameters from 5 (2 decay times, 2 decay amplitudes, 1 noise amplitude) to 3 (2 decay amplitudes, 1 noise amplitude), and increases the interpretability of results compared to a representation where decay times and amplitudes vary.

B. Room Transition Along a Straight Line: Spatial and Directional Analysis

In this section we extend the preceding analysis with directional information, i.e., the analyzed RIR set now exhibits both

inhomogeneity and anisotropy. Directional sound energy decay variations can be described with directional EDFs (DEDFs), as proposed by Berzborn and Vorländer [28]. We obtain the J directionally constrained beamformer output signals $\mathbf{S} \in \mathbb{R}^{J \times L}$ by steering axisymmetric beamformers

$$\mathbf{S} = \mathbf{A}\mathbf{h}, \quad (17a)$$

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_J]^T, \quad (17b)$$

where \mathbf{s}_{ξ} is the ξ th beamformer output, and $\mathbf{A} \in \mathbb{R}^{J \times (N+1)^2}$ and $\mathbf{h} \in \mathbb{R}^{(N+1)^2 \times L}$ denote the analysis matrix and the N th-order Ambisonic RIR,⁴ respectively. The analysis matrix \mathbf{A} can be calculated from the axisymmetric beamformer weights $\mathbf{c}_N \in \mathbb{R}^{N \times 1}$ and the spherical harmonics $\mathbf{y}(\boldsymbol{\Omega}_{\xi}) \in \mathbb{R}^{(N+1)^2 \times 1}$ evaluated at the beamformer steering directions $\boldsymbol{\Omega}_{\xi}$ as

$$\mathbf{A} = \mathbf{Y} \text{diag}_N(\mathbf{c}_N), \quad (18a)$$

$$\mathbf{Y} = [\mathbf{y}(\boldsymbol{\Omega}_1), \mathbf{y}(\boldsymbol{\Omega}_2), \dots, \mathbf{y}(\boldsymbol{\Omega}_J)]^T, \quad (18b)$$

$$\mathbf{c}_N = [c_0, c_1, \dots, c_N]^T. \quad (18c)$$

The term $\text{diag}_N(\cdot)$ formalizes μ times repeating the ν th axisymmetric beamformer weight, with μ and ν being the spherical harmonic (SH) degree and order, respectively. We employ a

⁴We do not explicitly define a channel ordering here because it is not relevant for this work as long as it is used consistently along the entire analysis pipeline. The maximum spherical harmonic order N shall not be confused with the noise term N_0 of the decay model.

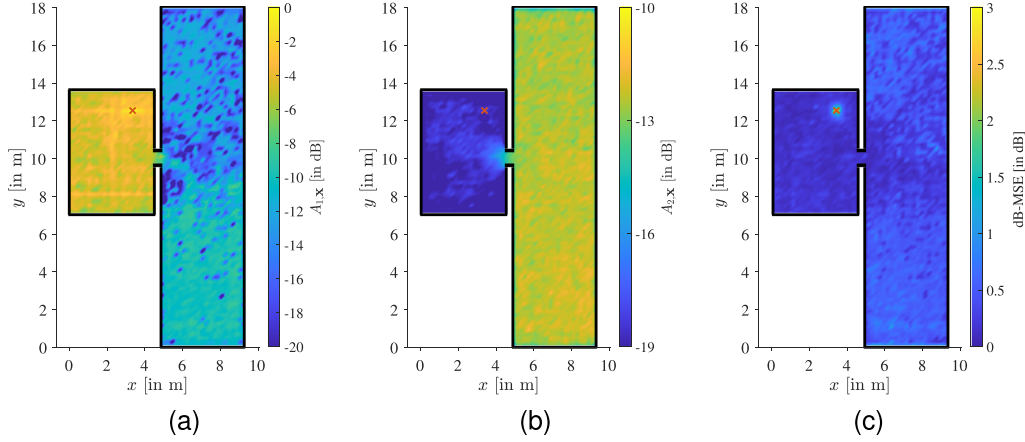


Fig. 7. Common-slope analysis results (1 kHz octave band) of a scene consisting of two coupled rooms. The analysis is based on the Extended Room Transition dataset variant 1 (c.f. Section IV) and the common decay times were obtained with the k-means clustering approach (c.f. Section III-B4). The orange \times indicates the sound source position. (a) and (b) show how the decay amplitudes of the common-slope model [c.f. (9)] vary for different receiver positions, with (a) depicting the $A_{1,x}$ and (b) the $A_{2,x}$ variations, respectively. The common decay times amount to $T_1 = 0.46$ s and $T_2 = 1.51$ s. The variations of the $A_{2,x}$ values highlight how the reverberation of the more reverberant hallway spreads into the less reverberant meeting room. (c) indicates that the common-slope model introduces only a small decibel-based mean squared error (dB-MSE) between the modeled and true energy decay [c.f. (16)]. A perfect fit would yield a dB-MSE of 0 dB.

spherical Butterworth beamformer [52, Table 3.1], which exhibits a great front-to-back-separation and has been successfully used in room acoustic analysis and Ambisonic RIR processing applications [53], [54]. The axisymmetric weights of the spherical Butterworth beamformer are given by

$$c_\nu^{\text{Butterworth}} = \frac{1}{\sqrt{1 + (\nu/\nu_c)^{2\gamma}}}, \quad (19)$$

where we set the Butterworth beamformer order $\gamma = 5$, and the cuton SH order $\nu_c = 3$.

Finally, the DEDFs $d(\mathbf{x}, t)$ are obtained by applying the Schroeder backwards integration procedure [c.f. (3)] on the beamformer outputs. The source-receiver configuration \mathbf{x} now represents source and receiver position as well as DOD and DOA.

The common-slope analysis follows the same procedure as in the previous section. Fig. 6 summarizes all analysis results, revealing how the two decay processes of the sub-rooms cross-fade during the transition. The reverberation of the more reverberant hallway ($T_2 = 1.53$ s) gradually fades in while approaching the door and eventually passing through it. In the hallway, the longer decay predominates and masks the reverberation of the meeting room ($T_1 = 0.43$ s).

Fig. 6(a), (b), (d), and (e) show anisotropic late reverberation in the investigated environment for various azimuth and elevation angles. In the meeting room, the reverberation of the hallway is most prominent when steering the beamformer toward the door. Higher decay amplitudes can be observed when the beamformer is steered toward $\phi = \pm 90^\circ$. This effect might be caused by the non-uniform absorption distribution and uneven excitation of the room due to the directional loudspeaker. Fig. 6(d) and (e) highlight the anisotropy of the late reverberation with respect to different elevation angles, which is caused by the non-uniform absorption distribution in the rooms (e.g., carpet in the meeting room).

Fig. 6(c) and (f) show that the common-slope model introduces only little dB-MSE between the modeled and true EDFs, with average and 95% quantile values of 0.25 dB and 0.63 dB, respectively.

C. Two-Dimensional Receiver Grid

In this section we extend the analysis of Section V-A to the entire room geometry depicted in Fig. 4. We carry out the common-slope analysis analogously to the previous section.

For the ERTD1 variant, we obtain the common decay times $T_1 = 0.46$ s and $T_2 = 1.51$ s. Fig. 7(a) shows the spatial variations of the $A_{1,x}$ values. They are generally higher in the meeting room, where they slightly decrease with increasing distance to the sound source. The $A_{1,x}$ values are somewhat larger in the lower half of the hallway, potentially due to shadowing caused by the wall. While the wall shadows the sound for the upper half of the hallway, some sound can directly pass through the door to reach the lower half of the meeting room.

Fig. 7(b) shows the spatial variations of the $A_{2,x}$ values. In the meeting room, they are on average 6 dB to 8 dB lower than in the hallway. The plot also shows the previously described fade-in behavior when transitioning from the meeting room to the hallway, with higher amplitudes in front of the door that gradually decrease towards the sides.

Lastly, Fig. 7(c) illustrates the dB-MSE [c.f. (16)] introduced by the common-slope analysis. The dB-MSE is well below 3 dB for almost all receiver positions, with an average dB-MSE of 0.36 dB. The highest errors occur near the sound source, caused by steep sound energy drops after the direct sound that cannot be described by the multi-exponential model. Slightly increased errors can be observed at receiver positions far from the door due to longer horizontal sections in the early part of the EDFs due to the obstructed direct sound.

Fig. 8 summarizes the common-slope analysis results for the ERTD2 variant. A new mode group emerges due to the

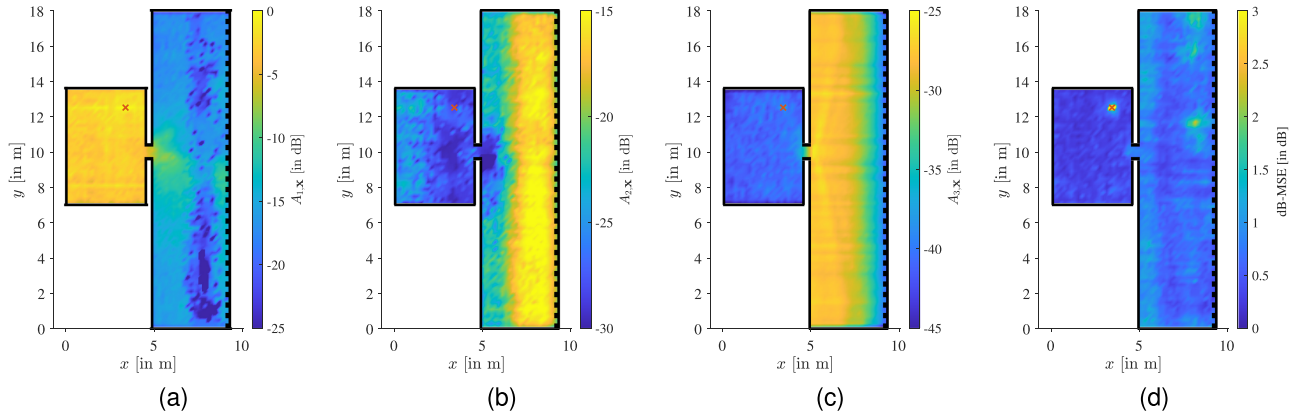


Fig. 8. Common-slope analysis results (1 kHz octave band) of a scene consisting of two coupled rooms, with the rightmost wall being highly absorptive. The analysis is based on the Extended Room Transition dataset variant 2 (c.f. Section IV) and the common decay times were obtained with the k-means clustering approach (c.f. Section III-B4). The orange \times indicates the sound source position. (a), (b), and (c) show how the decay amplitudes of the common-slope model [c.f. (9)] vary for different receiver positions, with (a) depicting the $A_{1,x}$, (b) the $A_{2,x}$, and (c) the $A_{3,x}$ variations, respectively. The common decay times amount to $T_1 = 0.44$ s, $T_2 = 0.84$ s, and $T_3 = 1.89$ s. In addition to the reverberation cross-fade that was already observed in Fig. 7, the amplitude variations in (b) and (c) highlight the effect of the highly absorptive right hallway wall. (d) indicates that the common-slope model introduces only a small decibel-based mean squared error (dB-MSE) between the modeled and true energy decay [c.f. (16)]. A perfect fit would yield a dB-MSE of 0 dB.

TABLE I
DECIBEL-BASED MEAN SQUARED ERROR [dB-MSE, C.F. (16)] BETWEEN ENERGY DECAY FUNCTIONS OF THE EXTENDED ROOM TRANSITION DATASET (C.F. SECTION IV) AND VARIOUS MODELS. A PERFECT FIT WOULD YIELD A dB-MSE OF 0 dB

	ERTD1			ERTD2				
	κ	mean	median	95% q.	κ	mean	median	95% q.
Common-slope model								
250 Hz octave band	3	0.52 dB	0.48 dB	0.95 dB	3	0.68 dB	0.63 dB	1.18 dB
500 Hz octave band	3	0.40 dB	0.36 dB	0.76 dB	3	0.50 dB	0.47 dB	0.88 dB
1 kHz octave band	2	0.36 dB	0.31 dB	0.72 dB	3	0.64 dB	0.61 dB	1.14 dB
2 kHz octave band	2	0.28 dB	0.21 dB	0.66 dB	3	0.66 dB	0.49 dB	1.75 dB
Traditional multi-exponential model								
250 Hz octave band	3	0.36 dB	0.32 dB	0.68 dB	3	0.59 dB	0.53 dB	1.06 dB
500 Hz octave band	3	0.26 dB	0.23 dB	0.48 dB	3	0.50 dB	0.43 dB	1.00 dB
1 kHz octave band	2	0.22 dB	0.19 dB	0.41 dB	3	0.63 dB	0.51 dB	1.38 dB
2 kHz octave band	2	0.19 dB	0.18 dB	0.35 dB	3	0.31 dB	0.27 dB	0.59 dB
Traditional single-exponential model								
250 Hz octave band	1	1.21 dB	0.57 dB	4.26 dB	1	3.22 dB	2.40 dB	8.22 dB
500 Hz octave band	1	1.03 dB	0.32 dB	4.00 dB	1	7.12 dB	5.88 dB	17.65 dB
1 kHz octave band	1	1.03 dB	0.23 dB	3.84 dB	1	8.89 dB	5.68 dB	21.77 dB
2 kHz octave band	1	1.02 dB	0.21 dB	3.84 dB	1	2.88 dB	2.04 dB	6.61 dB

highly non-uniform absorption distribution in the hallway, thus requiring three common decay times ($T_1 = 0.44$ s, $T_2 = 0.84$ s, and $T_3 = 1.89$ s) for describing all EDFs with low dB-MSE. The decay amplitude plots in Fig. 8(a) and (c) show the reverberation cross-fade that was already observed previously. Additionally, Fig. 8(b) and (c) illustrate the effect of the absorptive right hallway wall, causing the decay slope with the shorter decay time to be more prominent at positions close to the wall, whereas the longer decay slope dominates positions in the left half of the hallway. Fig. 8(d) shows that the dB-MSE values introduced by the common-slope analysis are well below 3 dB for most of the environment, amounting to an average dB-MSE of 0.64 dB.

Table I shows the dB-MSE between EDFs of the ERTD and various models. It compares the traditional multi-exponential [15] and single-exponential model [55] with the

proposed common-slope model [c.f. (9) and (10)]. While the traditional multi-exponential model yields the lowest mean and median dB-MSEs for all octave bands, the fitting performance of the proposed common-slope model is only slightly inferior, despite requiring only approximately half of the source-receiver-configuration-dependent parameters. In contrast, the traditional single-exponential model is clearly not general enough for some source-receiver configurations in the coupled room geometry, as indicated by the considerably increased mean and 95% quantile values. However, the median error of the traditional single-exponential model is comparable to the other two models due to the large number of receiver positions in the more reverberant hallway ($\mathcal{N}_{\text{hallway}}^{(\text{ERTD})} = 2070$, $\mathcal{N}_{\text{meeting}}^{(\text{ERTD})} = 759$). Despite the room coupling, most EDFs with receiver positions in the hallway can be described with a single-exponential model because the faster

energy decay of the meeting room is masked by the slower energy decay of the hallway (c.f. Fig. 2). This effect is less pronounced for the ERTD2.

The common-slope model yields good fitting results for all octave bands with mean dB-MSEs well below 1 dB. The increased model order in the 250 Hz and 500 Hz bands can be explained by additional mode groups. At lower frequencies, axial and tangential modes are predominant, whereas the higher frequencies are dominated by oblique modes [56]. The different mode types are characterized by the number of walls that are involved during their formation, and consequently, their decay times may vary considerably [56].

VI. DISCUSSION

Previous studies have often reported inhomogeneous or anisotropic sound energy decay, decay times or reverberation times [18], [19], [20], [21], [22], [23], [24], [25], [26], [28]. While this sounds at first like a contradiction compared to our proposed common-slope analysis approach, two explanations for the compatibility of previous study results and our approach are given in the following.

Firstly, the multi-exponential decay model is well-established for describing sound energy decays in rooms [15], [44]. Therefore, EDFs of different source-receiver configurations can be easily expressed with distinct model parameter combinations, i.e., by varying the decay times, decay amplitudes, and noise amplitudes. In fact, modifying all parameters at the same time may be an overparameterization. According to Lanczos [57], the separation of exponentials, i.e., the decomposition of a decay function into a linear combination of exponentials with initially unknown decay times [c.f. (9) and (10)], is a highly ill-conditioned problem.⁵ In other words, due to an inherent lack of measurement accuracy, two decay models with considerably different decay times, decay amplitudes, and model orders can result in numerically equivalent fits to a given measurement. It is therefore not surprising that previous studies found a combination of decay time and amplitude variations to yield suitable EDF fits. However, the common-slope analysis provides a more compact representation of the variations in terms of only decay amplitudes, thus making the results easier to interpret. For example, the evaluation in Section V demonstrated that phenomena like the fade-in of slopes in coupled room geometries become evident by using the common-slope analysis.

Secondly, for certain EDF variations, variable decay times would actually yield a numerically better fit than the common-slope model. For example, let us assume a room that features modes with decay times uniformly distributed between 1.2 s and 1.8 s. Furthermore, we consider two different source-receiver configurations in this room. For the first configuration, $\mathbf{x}^{(1)}$, the mode amplitudes are zero for all modes with decay times above 1.5 s, whereas, for the other configuration, $\mathbf{x}^{(2)}$, they are zero for all modes below 1.5 s. In such an admittedly quite

extreme scenario, we would expect the two configurations to exhibit numerically best fits with different decay times, such as $T_{1,\mathbf{x}^{(1)}} = 1.35$ s and $T_{1,\mathbf{x}^{(2)}} = 1.65$ s, respectively. This observation is also the reason why the unconstrained decay analysis yields decay time distributions as displayed in the histogram of Fig. 3. Moreover, it explains why the common-slope model introduces a small error at certain positions.

In summary, the goal of the common-slope analysis is not to find a numerically perfect fit to the measurement, because such a fit may not even exist due to the ill-conditioned nature of the fitting problem [57]. In contrast, the goal is to find a set of common decay times, which are representative of the underlying room geometry and introduce a low error for all considered source-receiver configurations. This objective can be seen analogous to the reverberation time measurement according to ISO [16], [17], which aims to find a reverberation time value that is representative of the measured room. The common-slope analysis extends this concept by introducing multiple slopes and directional analysis.

The common-slope analysis is heavily inspired by the common-acoustical-pole and residue (CAPR) model by Haneda et al. [10]. However, the common-slope analysis abstracts away the underlying modal nature of the sound field, and describes only the energy decay variations in terms of exponential slopes. In the common-slope model, the degrees of freedom and consequently the number of model parameters is fairly small. More precisely, the common-slope model has $(2\kappa + 1)$ parameters for each frequency band, thus enabling a compact representation of inhomogeneous and anisotropic energy decay. In contrast, the CAPR model requires a large number of parameters, because all room modes must be modeled individually. Due to the compactness of the common-slope model, it can be used to drive computationally-efficient room acoustic simulators, such as [31].

For the special case of model order $\kappa = 2$, the common-slope model can be transformed into the parametric diffusion equation solution proposed by Luizard et al. [32]. Their parametric model was intended to describe the inhomogeneous energy decay in coupled rooms by adjusting the decay amplitudes of two exponentials with fixed decay times according to the source-receiver distance and two heuristically determined parameters. Our model extends the model of Luizard et al. to general geometries and reverberation that is both inhomogeneous and anisotropic.

Lastly, the applicability of the common-slope model in various frequency bands should be discussed. The analyses presented in the preceding section showed that the common-slope model works well in the octave bands between 250 Hz to 2000 Hz. In these frequency bands, it yields fitting results that are comparable to the traditional multi-exponential model regarding the dB-MSE between the modelled and true EDF [c.f. (16)]. However, in very low frequencies, room modes are usually more sparse. In such cases, it may be necessary to model the individual modes with distinct decay times, analogously to the CAPR model by Haneda et al. [10]. Hence, for very low frequencies, it may not be possible to leverage the full potential of the common-slope model regarding its compactness.

⁵A similar case of ill-conditioning can also be observed in Prony's method, whose goal is to determine decay rates and amplitudes of damped sinusoids [58]. For sinusoids with zero oscillation, the problems are identical.

VII. CONCLUSION

This article introduced the common-slope model for late reverberation. Its main idea is to use one common set of decay times to model large EDF sets, whose EDFs describe different source-receiver configurations within the same environment. Consequently, all directional and spatial energy decay variations are described as a weighted sum of exponential functions with fixed decay times and a noise term. Different approaches for determining the common decay times were explored, finding that the k-means clustering of decay times is the most general of them. It was shown that the common-slope model reduces the degrees of freedom in energy decay analysis considerably, while introducing only a small error between the modeled and the true EDF. Furthermore, the common-slope model enables a compact representation of inhomogeneous and anisotropic energy decay, thus making its analysis results easy to interpret. For example, in our evaluation, the common-slope analysis revealed acoustic phenomena like the fade-in of reverberation during the transition between coupled rooms, or the inhomogeneous energy decay caused by a highly non-uniform absorption distribution. The proposed model leverages its full potential in frequencies above the Schroeder frequency with a dense modal overlap, whereas traditional mode-wise processing might be advantageous in lower frequencies.

The common-slope model may benefit future research on room acoustic analysis and modeling. Furthermore, it will be valuable for all research fields relying on late reverberation models, such as dereverberation, echo cancellation, source separation, sound field equalization, and parametric spatial audio rendering.

Companion Page: A companion page with additional information, animations, and source code related to this article can be found at: <http://research.spa.aalto.fi/publications/papers/ieeetaslp-common-slope/>

ACKNOWLEDGMENT

We acknowledge the computational resources provided by the Aalto Science-IT project.

REFERENCES

- [1] H. Kuttruff, *Room Acoustics*, 4th ed. London, U.K.: Spon Press, 2000.
- [2] R. Badeau, "Common mathematical framework for stochastic reverberation models," *J. Acoust. Soc. Amer.*, vol. 145, no. 4, pp. 2733–2745, 2019.
- [3] J. A. Moorer, "About this reverberation business," *Comput. Music J.*, vol. 3, no. 2, pp. 13–28, 1979.
- [4] N. Raghuvanshi and J. Snyder, "Parametric directional coding for pre-computed sound propagation," *ACM Trans. Graph.*, vol. 37, no. 4, 2018, Art. no. 108.
- [5] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 205–216, Jan. 2014.
- [6] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Amer.*, vol. 143, no. 6, pp. 3770–3779, 2018.
- [7] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. Van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 10, pp. 1929–1941, Oct. 2017.
- [8] E. Fernandez-Grande, "Sound field reconstruction using a spherical microphone array," *J. Acoust. Soc. Amer.*, vol. 139, no. 3, pp. 1168–1178, 2016.
- [9] P. Samarasinghe, T. Abhayapala, M. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2217–2227, Dec. 2015.
- [10] 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 Process.*, vol. 7, no. 6, pp. 709–717, Nov. 1999.
- [11] B. Xu and S. D. Sommerfeldt, "A hybrid modal analysis for enclosed sound fields," *J. Acoust. Soc. Amer.*, vol. 128, no. 5, pp. 2857–2867, 2010.
- [12] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. Van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, Jul. 2017.
- [13] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, 1st ed. Berlin/Heidelberg, Germany: Springer, 2008.
- [14] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environment," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.
- [15] N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," *J. Acoust. Soc. Amer.*, vol. 110, no. 3, pp. 1415–1424, 2001.
- [16] *Acoustics - Measurement of Room Acoustic Parameters — Part 1: Performance Spaces*, ISO Standard 3382-1, International Organization for Standardization, Geneva, Switzerland, 2009.
- [17] *Acoustics - Measurement of Room Acoustic Parameters — Part 2: Reverberation Time in Ordinary Rooms*, ISO Standard 3382-2, International Organization for Standardization, Geneva, Switzerland, 2008.
- [18] T. Akama, H. Suzuki, and A. Omoto, "Distribution of selected monaural acoustical parameters in concert halls," *Appl. Acoust.*, vol. 71, no. 6, pp. 564–577, 2010.
- [19] X. Pelorson, J.-P. Vian, and J.-D. Polack, "On the variability of room acoustical parameters: Reproducibility and statistical validity," *Appl. Acoust.*, vol. 37, no. 3, pp. 175–198, 1992.
- [20] K. Fujii et al., "Spatial distribution of acoustical parameters in concert halls: Comparison of different scattered reflections," *J. Temporal Des. Archit. Environ.*, vol. 4, no. 1, pp. 59–68, 2004.
- [21] M. Berzborn, J. Balint, and M. Vorländer, "On the estimation of directional decay times in reverberation rooms," in *Proc. Euronoise Online Conf.*, 2021, pp. 1–9.
- [22] B. Alary and V. Välimäki, "A method for capturing and reproducing directional reverberation in six degrees of freedom," in *Proc. IEEE Int. Conf. Immersive 3D Audio*, 2021, pp. 1–8.
- [23] M. Garai, S. De Cesaris, and D. D’Orazio, "Spatial distribution of monaural acoustical descriptors in historical Italian theatres," in *Proc. Int. Symp. Room Acoust.*, 2013, pp. 1–13.
- [24] M. Horvat, K. Jambrosic, and H. Domitrovic, "The examination of the influence of standing waves on reverberation time measurements in small reverberant rooms," *J. Acoust. Soc. Amer.*, vol. 123, no. 5, 2008, Art. no. 3615.
- [25] A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the use of a diffusion model for acoustically coupled rooms," *J. Acoust. Soc. Amer.*, vol. 120, no. 4, pp. 2043–2054, 2006.
- [26] A. Billon, "On the spatial variations of double-sloped decay in coupled-volume concert halls," in *Proc. Int. Symp. Room Acoust.*, 2007, pp. 1–5.
- [27] 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. Amer.*, vol. 133, no. 6, pp. 3975–3985, 2013.
- [28] M. Berzborn and M. Vorländer, "Directional sound field decay analysis in performance spaces," *Building Acoust.*, vol. 28, no. 3, pp. 249–263, 2021.
- [29] R. San Martín, I. B. Witew, M. Arana, and M. Vorländer, "Influence of the source orientation on the measurement of acoustic parameters," *Acta Acustica United Acustica*, vol. 93, pp. 387–397, 2007.
- [30] L. Cremer and H. A. Müller, *Principles and Applications of Room Acoustics*, vol. 1. Westport, CT, USA: Peninsula, 2016.
- [31] 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.
- [32] P. Luizard, J.-D. Polack, and B. F. G. Katz, "Sound energy decay in coupled spaces using a parametric analytical solution of a diffusion equation," *J. Acoust. Soc. Amer.*, vol. 135, no. 5, pp. 2765–2776, 2014.

- [33] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, no. 3, pp. 409–412, 1965.
- [34] O. Das, P. Calamia, and S. V. Amengual Garí, "Room impulse response interpolation from a sparse set of measurements using a modal architecture," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 960–964.
- [35] M. Zaunschirm, F. Zagala, and F. Zotter, "Auralization of high-order directional sources from first-order RIR measurements," *Appl. Sci.*, vol. 10, no. 11, 2020, Art. no. 3747.
- [36] G. H. Koopmann, L. Song, and J. B. Fahline, "A method for computing acoustic fields based on the principle of wave superposition," *J. Acoust. Soc. Amer.*, vol. 86, no. 6, pp. 2433–2438, 1989.
- [37] J. Escolano, J. J. López, and B. Pueo, "Directive sources in acoustic discrete-time domain simulations based on directivity diagrams," *J. Acoust. Soc. Amer.*, vol. 121, no. 6, pp. EL256–EL262, 2007.
- [38] H. Kuttruff, "Eigenschaften und Auswertung von Nachhallkurven [characteristics and analysis of sound energy decay curves]," *Acta Acustica United Acustica*, vol. 8, no. 1, pp. 273–280, 1958.
- [39] C. F. Eyring, "Reverberation time measurements in coupled rooms," *J. Acoust. Soc. Amer.*, vol. 3, no. 2, pp. 181–206, 1931.
- [40] F. V. Hunt, L. L. Beranek, and D. Y. Maa, "Analysis of sound decay in rectangular rooms," *J. Acoust. Soc. Amer.*, vol. 11, no. 1, pp. 80–94, 1939.
- [41] E. Nilsson, "Decay processes in rooms with non-diffuse sound fields Part I: Ceiling treatment with absorbing material," *Building Acoust.*, vol. 11, no. 1, pp. 39–60, 2004.
- [42] M. Meissner, "Acoustic energy density distribution and sound intensity vector field inside coupled spaces," *J. Acoust. Soc. Amer.*, vol. 132, no. 1, pp. 228–238, 2012.
- [43] J. Balint, F. Muralter, M. Nolan, and C.-H. Jeong, "Bayesian decay time estimation in a reverberation chamber for absorption measurements," *J. Acoust. Soc. Amer.*, vol. 146, no. 3, pp. 1641–1649, 2019.
- [44] N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayesian decay time estimation," *J. Acoust. Soc. Amer.*, vol. 117, no. 6, pp. 3707–3715, 2005.
- [45] G. Götz, R. F. Pérez, S. J. Schlecht, and V. Pulkki, "Neural network for multi-exponential sound energy decay analysis," *J. Acoust. Soc. Amer.*, vol. 152, no. 2, pp. 942–953, 2022.
- [46] T. Jasa and N. Xiang, "Efficient estimation of decay parameters in acoustically coupled-spaces using slice sampling," *J. Acoust. Soc. Amer.*, vol. 126, no. 3, pp. 1269–1279, 2009.
- [47] T. McKenzie, S. J. Schlecht, and V. Pulkki, "Acoustic analysis and dataset of transitions between coupled rooms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 481–485.
- [48] T. McKenzie, S. J. Schlecht, and V. Pulkki, "A dataset of measured spatial room impulse responses for the transition between coupled rooms (Version 1.2)," Feb. 2021, doi: [10.5281/zenodo.4636068](https://doi.org/10.5281/zenodo.4636068).
- [49] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probability*, 1967, vol. 1, pp. 281–297.
- [50] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, 4th ed. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [51] J. Saarelma, "Finite-difference time-domain solver for room acoustics using graphics processing units," M.S. thesis, Sch. Elect. Eng., Aalto Univ., Espoo, Finland, 2013.
- [52] B. Devaraju, "Understanding filtering on the sphere," Ph.D. dissertation, Univ. of Stuttgart, Stuttgart, Germany, 2015, doi: [10.18419/opus-3985](https://doi.org/10.18419/opus-3985).
- [53] G. Götz, C. Hold, T. McKenzie, S. J. Schlecht, and V. Pulkki, "Analysis of multi-exponential and anisotropic sound energy decay," in *Proc. 48th Jahrestagung für Akustik*, 2022, pp. 1055–1058.
- [54] C. Hold, T. McKenzie, G. Götz, S. J. Schlecht, and V. Pulkki, "Resynthesis of spatial room impulse response tails with anisotropic multi-slope decays," *J. Audio Eng. Soc.*, vol. 70, no. 6, pp. 526–538, 2022.
- [55] N. Xiang, "Evaluation of reverberation times using a nonlinear regression approach," *J. Acoust. Soc. Amer.*, vol. 98, no. 4, pp. 2112–2121, 1995.
- [56] M. Meissner, "Acoustics of small rectangular rooms: Analytical and numerical determination of reverberation parameters," *Appl. Acoust.*, vol. 120, pp. 111–119, 2017.
- [57] C. Lanczos, *Applied Analysis*. New York, NY, USA: Dover, 1988.
- [58] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 1991.



Georg Götz received the B.Sc. and M.Sc. degrees in media technology from the Ilmenau University of Technology, Ilmenau, Germany, specialising in spatial audio and psychoacoustics in 2016 and 2019, respectively. He is currently a Doctoral Researcher with the Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Espoo, Finland. He was the deputy Chair of the AES Student Section Ilmenau from 2013 to 2016, and in 2023, he was part of the convention committee of the 154th AES Convention in Espoo. In 2022,

he completed a research internship with Reality Labs Research, Meta. The research of his doctoral thesis is about machine-learning-based virtual acoustics rendering. His other research interests include room acoustics, psychoacoustics, and VR/AR technology.



Sebastian J. Schlecht (Senior Member, IEEE) received the Diploma in applied mathematics from the University of Trier, Trier, Germany, in 2010, and the M.Sc. degree in digital music processing from the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K., in 2011, and the Doctoral degree with the International Audio Laboratories Erlangen, Erlangen, Germany, on artificial spatial reverberation and reverberation enhancement systems in 2017. He is currently a Professor of practice for sound in virtual

reality with the Acoustics Lab, Department of Information and Communications Engineering, and Media Lab, Department of Art and Media, Aalto University, Espoo, Finland. From 2012 to 2019, he was also an External Research and Development Consultant and Lead Developer of the 3D Reverb algorithm with Fraunhofer IIS, Erlangen.



Ville Pulkki is currently a Professor with the Department of Information and Communications Engineering, Aalto University, Helsinki, Finland. He has been working in the field of spatial audio for more than 20 years. He developed the vector-base amplitude panning (VBAP) method in his Ph.D. in 2001 and directional audio coding after the Ph.D. with his research group. He also has contributions in perception of spatial sound, laser-based measurement of room responses, and binaural auditory models. He was the recipient of the Samuel L. Warner Memorial

Medal Award from the Society of Motion Picture and Television Engineers and AES Silver Medal Award. He enjoys being with his family, building his summer house, and performing in musical ensembles.