ESTIMATION OF MULTI-SLOPE AMPLITUDES IN LATE REVERBERATION

Jeremy B. Bai and Sebastian J. Schlecht

Multimedia Communications and Signal Processing
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
jeremy.bai@fau.de

ABSTRACT

The common-slope model is used to model late reverberation of complex room geometries such as multiple coupled rooms. The model fits band-limited room impulse responses using a set of common decay rates, with amplitudes varying based on listener positions. This paper investigates amplitude estimation methods within the common-slope model framework. We compare several traditional least squares estimation methods and propose using LINEX regression, a Maximum Likelihood approach using log-squared RIR statistics. Through statistical analysis and simulation tests, we demonstrate that LINEX regression improves accuracy and reduces bias when compared to traditional methods.

1. INTRODUCTION

The advancement of Virtual Reality (VR) and Augmented Reality (AR) technology saw great demand in room acoustics modeling. VR/AR applications require perceptually accurate and computationally efficient sound field rendering to support highly dynamic real-time environmental changes and user interactions.

The recently proposed common-slope model by [1] shows potential for real-time late reverberation rendering, particularly in complex acoustic environments. It is well-established that room impulse responses (RIRs) exhibit exponentially decaying tails [2, 3]. Traditional statistical models represent late reverberation as exponentially decaying Gaussian noise, assuming a single dominant decay rate within a frequency band [3, 4]. Empirical studies suggest that multiple decay rates often coexist, especially in coupled spaces or non-diffuse environments [5, 6], and the common-slope model extends on this finding.

The common-slope model builds on the modal decomposition of RIRs, which reveals that amplitudes and decay rates are separable components [7, 8, 9]. The key assumption of this model states that decay rates across different RIRs depend only on the property of the room, leading to the phenomenon that prominent modes within the same frequency band tend to cluster around a few significant decay rates, hence the term *common slopes* [1, 2]. In contrast, the amplitudes of these decay components vary with the source and receiver locations. This parameterization enables an efficient representation of room acoustics: all RIRs in a given space can be described using a small set of common decay rates and location-dependent amplitudes, significantly reducing the storage requirements for acoustic scene rendering while maintaining perceptual accuracy.

The model offers flexibility in representing room acoustics, particularly in accounting for inhomogeneity and anisotropy. By allowing for multiple decay slopes, it can reflect variations in reverberation characteristics, which preserves perceptual cues related to spaciousness and surface reflectivity. Prior studies have demonstrated perceptual relevance in complex acoustic environments such as coupled rooms [10] and in scenarios involving smooth transitions between spaces, supporting the rendering of changes in listener position [11].

It is straightforward that the model is parametric and requires the identification of both the decay rates and amplitudes of the slopes. Numerous studies have explored decay rate estimation under different acoustic modeling frameworks. Classic methods for estimating reverberation time (RT) rely on Energy Decay Curves (EDCs) derived via Schroeder backward integration. Schroeder applies linear regression to estimate a single decay rate [4], while Xiang and Goggans analytically derive the EDC and employ Bayesian nonlinear least squares regression to handle multiple decay components [5]. However, both approaches neglect amplitude parameter estimation. In blind RT estimation, least squares regression on RIR at different scales remains a standard approach [12, 13]. This paper provides a comprehensive analysis of these statistical models, emphasizing amplitude estimation through both analytical inference and simulations.

On top of the existing estimation methods mentioned above, we introduce a novel regression framework for the common-slope model. Using Maximum Likelihood Estimation (MLE), we derive the probability distribution of an additive model error in logarithmic squared RIRs. Notably, the resulting negative log-likelihood function aligns with the linear exponential (LINEX) loss function. Under the assumption of correct model specification, our proposed regression gives more efficient amplitude estimators compared to conventional least-squares estimation.

The remainder of this paper is structured as follows. Section 2 revisits the common-slope model, introducing the statistical framework that supports the different estimation methods. Section 3 begins with an overview of least squares regression on different scales of RIRs and some refinements according to the model specification, then motivates and introduces the proposed MLE-based LINEX regression approach, and finishes with a discussion of model fitting and estimation over EDCs. In Section 4, we apply the estimation method to synthetic RIR datasets and evaluate performance. Finally, Section 5 provides concluding remarks and outlines potential future research.

2. REVISITING THE COMMON-SLOPE MODEL

Following the notation from [1], an RIR $h(\mathbf{x}, t)$ models the combined effects of sound waves given a known *source-receiver con-*

figuration $\mathbf{x} = (\mathbf{x}_s, \mathbf{x}_r, \mathbf{\Omega}_s, \mathbf{\Omega}_r)$, where $(\mathbf{x}_s, \mathbf{x}_r)$ are the sound source and receiver position in the room, and $(\mathbf{\Omega}_s, \mathbf{\Omega}_r)$ represents the direction of sound departure and arrival. For discrete time sample index $t \in \mathbb{Z}$ at sample rate f_s, the modal decomposition of an RIR gives:

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

$$= \sum_{k=1}^{\kappa} \sum_{m \in \mathcal{M}_k} \chi_m(\mathbf{x}) \tau_m(t), \tag{2}$$

where $h_b(\mathbf{x}, t)$ is the band-limited RIR filtered at frequency band f_b, and each mode indexed m is an exponentially decaying sinusoid. $\chi_m(\mathbf{x})$ describes the amplitudes and phases of the mode, varying only for different source-receiver configurations $\mathbf{x}$'s, briefly referred to as positions later. The temporal mode

$$\tau_m(t) = \exp\left(-\frac{\delta_m t}{f_s}\right) \cos\left(\frac{\omega_m t}{f_s}\right) \tag{3}$$

includes modal frequency $\omega_m \in \mathbb{R}_+$ and decay rate $\delta_m \in \mathbb{R}_+$ and does not depend on position $\mathbf{x}$.

The modes are grouped into κ clusters of decay rates such that $\{\mathcal{M}_k\}_{k=1}^{\kappa}$ is a partition of all modes $\{1, \ldots, M\}$. For each modal group $\mathcal{M}_k$, we expect the decay rates can be modeled by an average of the cluster $\delta_k \approx \frac{1}{|\mathcal{M}_k|} \sum_{m \in \mathcal{M}_k} \delta_m$, then the sum of modes in the same group with random phases and amplitudes can be modeled by a band-limited white noise with an exponentially decaying envelope, and the RIR can rewrite to:

$$h_b(\mathbf{x}, t) = \sum_{k=1}^{\kappa} A_{b,k}(\mathbf{x}) \exp\left(-\frac{\delta_{b,k} t}{f_s}\right) z_{b,k}(\mathbf{x}, t). \tag{4}$$

The aggregate amplitude $A_{b,k}(\mathbf{x})$ for the mode group k also depends only on position, accounting for all the information of $\chi_m(\mathbf{x})$ for $m \in \mathcal{M}_k$. And the exponential term is the envelope for the decay. Notice that the number of decay rates δ_k's is reduced from the number of modes to modal groups. More often in RT estimation literature, the decay rates take the form of RT60s denoted by T_k, which gives the time needed for the sound energy to decay by 60 dB.[1] The relationship between RT60 and decay rate writes $\delta_k = \frac{3\log(10)}{T_k} \approx \frac{6.9}{T_k}$.

We propose to model the noise component by standard Gaussian processes $z_{b,k}(\mathbf{x}, t) \sim N(0, 1)$ that are i.i.d. across positions $\mathbf{x}$, time t and decay group k for the following argument. Consider a band-pass filtered white Gaussian noise. Since filtering is a linear operation, the resulting random process retains a Gaussian distribution. Band-limited white noise is generally time-correlated because of the correlation introduced by filtering. This can be further addressed by demodulating the signal to a baseband equivalent and downsampling. Specifically, consider a band-pass filtered white Gaussian noise process centered at frequency f_c with bandwidth B. By demodulating this process to baseband and sampling at twice its bandwidth $2B$, the resulting process can be treated as approximately white [14, 15].

The model also inherently accommodates non-decaying noise components such as environmental noise or pickup noise, etc. For a non-decay slope, one can simply denote a slope with $k = 0$, and let the decay rate $\delta_0 = 0$, the exponential envelope then becomes constant 1.

The assumptions above give us the full distribution of a band-limited RIR $h_b(\mathbf{x}, t)$, and further allow us to estimate the parameters using a Maximum Likelihood approach. However, because we assume independence for analytical convenience rather than realism, the model is intended as a descriptive framework.

3. ESTIMATION MODELS

Following the last section, with the Gaussian assumption for the noise component, (4) gives a direct result that the band-limited RIR follows a zero-mean Gaussian distribution with time-varying variance. To simplify notation, we omit the frequency band b and the RIR position $\mathbf{x}$ after having established that the model applies for each frequency band and each instance of RIR. The distribution of $h(t)$ then writes

$$h(t) \sim \left(0, \sum_{k=1}^{\kappa} A_k^2 \exp\left(-\frac{2\delta_k t}{f_s}\right)\right) \tag{5}$$

This implies that all information about the amplitudes A_k is encoded in the second-order moment of $h(t)$, which represents the expected energy decay profile in squared amplitude. This characterization, as shown in Fig. 1(a) in energy, namely squared RIR $h^2(t)$, forms the basis of the modeling approaches introduced in the following subsections.

3.1. Nonlinear Least Squares Regression

Directly from (5), we have that the variance of $h(t)$, also the expectation of $h^2(t)$, is a function of time:

$$\mathrm{Var}[h(t)] = \mathbb{E}\big[h^2(t)\big] = f(t; \mathbf{A}^2) = \sum_{k=1}^{\kappa} A_k^2 \exp\left(-\frac{2\delta_k t}{f_s}\right), \tag{6}$$

where $\mathbf{A}^2 = [A_1^2, \ldots, A_\kappa^2]^\top$ is the vector notation of squared amplitude parameters. All estimation later is performed on the squared amplitudes because the model function $f(t; \mathbf{A}^2)$ is linear in A_k^2.

From here, it is intuitive to rewrite the squared RIR as a functional model with a zero-mean error term:

$$h^2(t) = f(t; \mathbf{A}^2) + \xi_{\text{sq}}(t; \mathbf{A}^2). \tag{7}$$

Since the error term $\mathbb{E}[\xi_{\text{sq}}(t; \mathbf{A}^2)] = 0$ has zero mean by design, we could run a naive nonlinear least squares (NLS) to estimate the amplitudes. The least squares estimators write:

$$\hat{\mathbf{A}}^2_{\text{LS}} = \arg\min_{\mathbf{A}^2} \sum_{t=1}^{T} \xi^2_{\text{sq}}(t; \mathbf{A}^2). \tag{8}$$

However, this method yields poor results, since the error is highly skewed due to $h^2(t)$ being nonnegative, and it exhibits an exponential decaying envelope, making contributions from later time samples heavily underweighted, as shown in Fig. 1 (a). This leads to a regression that is heavily biased toward the early part

[1] Note that the RT60 only provides a standardized and interpretable representation of the decay rate for each modal group, the 60 dB energy decay is not really observed in the RIR in general.

of the signal. Such behavior is expected, as NLS yields asymptotically efficient estimators only under the assumption of homoskedastic Gaussian errors—conditions under which NLS coincides with MLE [16]. From the Gaussian assumption on $h(t)$, it follows that $h^2(t)$ is Gamma-distributed with variance

$$\mathrm{Var}[h^2(t)] = 2\left(\mathrm{Var}[h(t)]\right)^2 = 2\left(f(t;\mathbf{A}^2)\right)^2, \tag{9}$$

which shows significant heteroskedasticity.

One way to tackle this is to adopt the weighted least squares (WLS), where a time-dependent weight function $w(t)$ compensates for the heteroskedasticity introduced by the exponentially decaying envelope:

$$\hat{\mathbf{A}}^2_{\text{WLS}} = \arg\min_{\mathbf{A}^2} \sum_{t=1}^{T} \left[w(t)\xi_{\text{sq}}(t)\right]^2. \tag{10}$$

However, selecting an appropriate weight function is nontrivial: the variance of the noise is dependent on the decay amplitudes A_k^2, making it difficult to prescribe weights *a priori*. As a result, WLS typically requires iterative schemes, which increase both computational complexity and analytical intractability [16].

As an alternative strategy, transforming the RIR into a scale with weaker time-dependent variance can improve LS robustness. Error structure becomes more homoskedastic, but a corresponding bias correction must be introduced into the functional model to ensure consistency of the estimator [17]. In the following subsections, we examine two such transformations: the power-law scale and the logarithmic scale.

3.1.1. Nonlinear Least Squares on Power-Law Scale

A power-law scale model was proposed in [12], which models the absolute power of RIR $|h(t)|^\alpha$ as a generalization of modeling the second-order RIR $h^2(t)$. This method attenuates the drastic variance change in the RIR, meanwhile avoiding the numerical problem of taking the logarithmic values near zero. Their empirical results show that choosing power value around $\alpha = 0.5$ is a useful compromise between linear and logarithmic amplitude scaling.

With the power-law scaling, we need to revise the functional model analytically before performing least squares. Consider first a simpler case of a zero-mean Gaussian variable $Z \sim N(0, \sigma^2)$, the absolute moments of Z are given by

$$\mathbb{E}[|Z|^\nu] = \left(2\sigma^2\right)^{\frac{\nu}{2}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}}, \tag{11}$$

where $\Gamma(\cdot)$ is the Gamma function [18]. Applying this for $\nu = 0.5$ to the RIR model, we arrive at the following functional form

$$\mathbb{E}\left[|h(t)|^{0.5}\right] = \left(2f(t;\mathbf{A}^2)\right)^{0.25} \frac{\Gamma(0.75)}{\sqrt{\pi}}, \tag{12}$$

with a zero-mean i.i.d. amplitude dependent error term ξ_{pow}:

$$|h(t)|^{0.5} = \left(2f(t;\mathbf{A}^2)\right)^{0.25} \frac{\Gamma(0.75)}{\sqrt{\pi}} + \xi_{\text{pow}}(t;\mathbf{A}^2). \tag{13}$$

Note that the functional model still takes the form of variance at the corresponding power level with a constant scaling factor. The least squares estimator for amplitudes in this model is analogously given by

$$\hat{\mathbf{A}}^2_{\text{LS-PL}} = \arg\min_{\mathbf{A}^2} \sum_{t=1}^{T} \xi^2_{\text{pow}}(t;\mathbf{A}^2). \tag{14}$$

An example of the regression model is shown in Fig. 1(b).

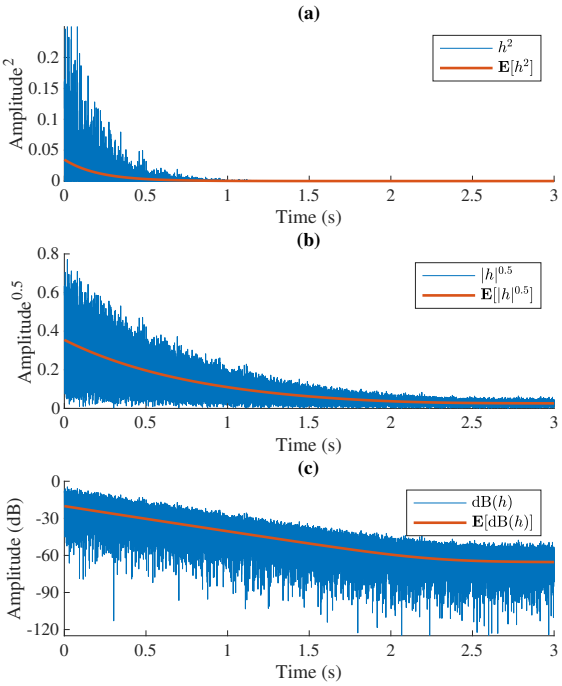


Figure 1: *Regression models on different scales of an example RIR, where the statistical expectation is the model function.* ***(a)*** *shows the model for squared RIR gives poor results because of disproportionate errors over time,* ***(b)*** *shows the attenuated change in variance in power-law scaled RIR, and* ***(c)*** *shows the equivariance of log-squared RIR, which makes regression more eligible.*

3.1.2. Nonlinear Least Squares on Log-Energy Scale

Even better than the power-law transformation, the squared RIR display full equivariance property on the logarithmic scale (or in dB unit), hence running a regression on log-squared RIR, i.e., log-energy scale is also intuitive.[2] The equivariance of log-squared RIR makes sure that we can have a homoskedastic model error when correctly specified.

Similarly to the power-law transformation, again, we need to derive the functional model analytically to perform a least squares regression. Consider again a zero-mean Gaussian variable $Z \sim N(0, \sigma^2)$, the transformed variable on log-energy scale then writes $Y = \log Z^2$. For a chi-square distribution variable $X_k \sim \chi^2(k)$ with k degree of freedom (DOF), the first moment of log chi-square gives

$$\mathbb{E}[\log X_k] = \log 2 + \psi\left(\tfrac{k}{2}\right),$$

where $\psi(\cdot)$ is the digamma function [19]. From this, we have the log-energy

$$Y = \log Z^2 = \log(\sigma^2 X_1) = \log \sigma^2 + \log X_1 \tag{15}$$

[2]The relationship between RIR in dB and natural logarithmic squared RIR is by a scaling factor of $\frac{10}{\log 10} \approx 4.343$. All analytical discussions in the paper are conducted using logarithms; however, the results are presented in a dB scale. The figures on a logarithmic scale are also shown in dB units.

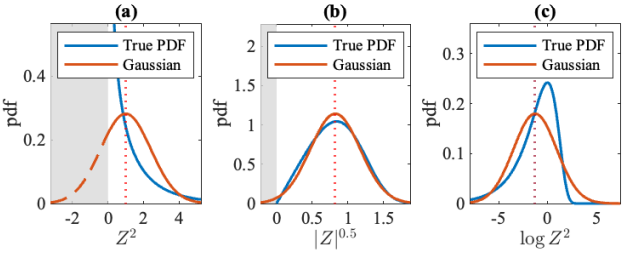


Figure 2: Distributions of model error on the three scales corresponding to Fig. 1—chi-square, square-root absolute Gaussian, and log chi-square—each exemplified by a standard Gaussian variable $Z \sim \mathcal{N}(0,1)$. The grey area falls outside the domain of the variable. For comparison, the red curves overlaying show Gaussian approximations matched by the true mean and variance, as would result from a least squares fit under model misspecification.

follows a shifted log chi-square distribution of DOF 1, the expectation writes

$$\mathbb{E}[Y] = \log \sigma^2 + \log 2 + \psi(0.5)\,, \tag{16}$$

which takes the form of the log-variance, and adds a constant.

Analogously, we can write the log-energy RIR in a similar functional form into a deterministic model

$$\mathbb{E}\big[\log h^2(t)\big] = \log f(t; \mathbf{A}^2) + \log 2 + \psi(0.5) \tag{17}$$

with a zero-mean i.i.d. error term $\xi_{\log}$:

$$\log h^2(t) = \log f(t; \mathbf{A}^2) + \log 2 + \psi(0.5) + \xi_{\log}(t; \mathbf{A}^2), \tag{18}$$

and the least squares estimator for the amplitudes is given by

$$\hat{\mathbf{A}}^2_{\text{LS-LE}} = \arg\min_{\mathbf{A}^2} \sum_{t=1}^{T} \xi^2_{\log}(t; \mathbf{A}^2). \tag{19}$$

An example of the regression model is shown in Fig. 1(c).

3.2. Maximum Likelihood and LINEX Regression

The LINEX regression model is also motivated by the equivariance property of squared RIR in logarithmic scale, but with a different loss function on the model error. From the Gaussian assumption of RIR $h(t)$, we can write the full distribution of $\log h^2(t)$ and run a Maximum Likelihood (ML) estimation, which results in our regression model.

To gain intuition, consider again the case of a zero-mean Gaussian variable $Z \sim \mathcal{N}(0, \sigma^2)$. From Sec. 3.1.2, the log-energy variable $Y = \log Z^2$ can then be modeled with $\log \sigma^2$ with a log chi-square distributed error which does not depend on σ^2, as shown in 15. Define the error on the log-energy model $U = Y - \log \sigma^2$, then we can confirm this by writing the probability density function using a mapping of variables from the Gaussian density,

$$f_U(u) = f_Z(z) \cdot \left| \frac{dz}{du} \right| \tag{20}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e^{u+\log\sigma^2}}{2\sigma^2} \right) \cdot \exp\left(\frac{u + \log\sigma^2}{2} \right) \tag{21}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{e^u - u}{2} \right). \tag{22}$$

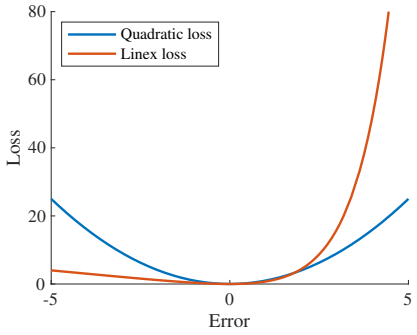


Figure 3: A comparison between quadratic and LINEX loss function. Error on the positive side is heavily penalized by the LINEX loss function.

As expected, all σ^2 cancel out in the pdf of the error term U, which means that we can separate the model parameter σ^2 from the model error distribution. The corresponding negative log-likelihood for U is also not dependent on σ^2:

$$-\ell(u) = -\log f_U(u) \sim e^u - u, \tag{23}$$

which highlights a convex loss function with global minimum at $u = 0$. This loss function is called the Linear-Exponential (LINEX) function, and is mostly used for optimization problems where error is distributed and penalized asymmetrically.

Returning to the log-energy variable Y, we can then rewrite it into a deterministic model with an error term described above:

$$Y = \log \sigma^2 + U. \tag{24}$$

The same method applies to the log-squared RIR. Analog to finding the variance in a scalar Gaussian case, we aim to estimate the amplitude parameters A_k that determine the time-varying variance structure of $h(t)$. Since $h(t)$ is modeled as a zero-mean Gaussian process, the log-energy $\log h^2(t)$ follows a scaled log chi-squared distribution and admits the following generative form, analogously to Y in (24):

$$\log h^2(t) = \log f(t; \mathbf{A}^2) + \xi_{\text{linex}}(t; \mathbf{A}^2) \tag{25}$$

where $\xi_{\text{linex}}(t)$ denotes an i.i.d. error additive to the log-variance model, corresponding to U in (24). The resulting ML estimator for the amplitude parameters is then obtained by minimizing the aggregate LINEX loss

$$\hat{\mathbf{A}}^2_{\text{LINEX}} = \arg\min_{\mathbf{A}^2} \sum_{t=1}^{T} e^{\xi_{\text{linex}}(t;\mathbf{A}^2)} - \xi_{\text{linex}}(t; \mathbf{A}^2), \tag{26}$$

This formulation can be interpreted as a functional regression model, hence the name LINEX regression. In contrast to the conventional NLS approach in (19), it is distinguished by the use of a non-quadratic, asymmetric error penalty metric, as shown in Fig. 3. The LINEX regression-based estimator aligns with the true data structure under the Gaussian assumption of RIR $h(t)$, and thus is more efficient than least squares estimators.

3.3. Fitting over Energy Decay Curve

A common practice for identifying decay parameters in an RIR is to apply Schroeder's backward integration and run a regression on the resulting EDC. Despite its widespread usage, existing literature provides limited discussion on the statistical inference of the estimation from the EDCs, particularly concerning decay amplitude parameters. In this section, we address this gap by explicitly analyzing and evaluating the EDC fitting procedure within the common-slope modeling framework.

The Schroeder's backward integration, defined as shown below, is performed to give the EDC $d(t)$,

$$d(t) = \sum_{\tau=t}^{T} h^2(\tau). \tag{27}$$

By taking expectations and applying the common-slope model assumption, we have

$$\mathbb{E}[d(t)] = \sum_{\tau=t}^{T} \mathbb{E}\big[h^2(\tau)\big] = \sum_{\tau=t}^{T} f(t; \mathbf{A}^2) \tag{28}$$

$$= \sum_{\tau=t}^{T} \sum_{k=1}^{\kappa} A_k^2 \exp\left(-\frac{2\delta_k \tau}{f_s}\right) \tag{29}$$

$$= \sum_{k=1}^{\kappa} A_k^2 \cdot \frac{\exp\left(-\frac{2\delta_k t}{f_s}\right) - \exp\left(-\frac{2\delta_k T}{f_s}\right)}{1 - \exp\left(-\frac{2\delta_k}{f_s}\right)}, \tag{30}$$

where the summation in the last line follows from geometric series properties. Defining the basis functions

$$\Psi_k(t) = \frac{\exp\left(-\frac{2\delta_k t}{f_s}\right) - \exp\left(-\frac{2\delta_k T}{f_s}\right)}{1 - \exp\left(-\frac{2\delta_k}{f_s}\right)}, \tag{31}$$

we have the expectation of the EDC as a linear model in the basis functions,

$$\mathbb{E}[d(t)] = g(t; \mathbf{A}^2) = \sum_{k=1}^{\kappa} A_k^2 \Psi_k(t). \tag{32}$$

In practice, the basis functions $\Psi_k(t)$ effectively act as exponential decay envelopes, due to the negligible magnitude of $\exp(-\frac{2\delta_k T}{f_s})$ for large T [5].

In the case of a non-decay slope $k = 0$, setting the decay rate $\delta_0 = 0$ would result in the basis function (31) being ill-defined, but taking the limit for $\delta_0 \to 0_+$ gives

$$\Psi_0(t) = T - t, \tag{33}$$

and the comprehensive functional model for EDC would include $k = 0$ in the summation in (32).

The main advantage of Schroeder's method is that it transforms model errors asymptotically into Gaussian according to the Central Limit Theorem, while preserving the exponential decay characteristics captured by the basis functions $\Psi_k(t)$.

The variance of the EDC writes

$$\mathrm{Var}[d(t)] = \sum_{\tau=t}^{T} \mathrm{Var}\big[h^2(\tau)\big] = 2\sum_{\tau=t}^{T} \big(f(\tau; \mathbf{A}^2)\big)^2, \tag{34}$$

exhibiting exponential decay characteristics that are analogous to the derivation of its expectation in (30). To address this exponentially decaying variance, we again take the log of EDC, yielding an approximated expression of the variance by Taylor expansion:

$$\mathrm{Var}[\log d(t)] \approx \mathrm{Var}\left[\frac{d(t) - \mathbb{E}[d(t)]}{\mathbb{E}[d(t)]}\right] = \frac{\mathrm{Var}[d(t)]}{\mathbb{E}^2[d(t)]} \tag{35}$$

$$= \frac{2\sum_{\tau=t}^{T} \big(f(\tau; \mathbf{A}^2)\big)^2}{\Big(\sum_{\tau=t}^{T} f(\tau; \mathbf{A}^2)\Big)^2}, \tag{36}$$

where the last line gives the concentration measure of f on the domain of (t, T). From Cauchy-Schwarz, we know that the term reaches lower bound $\frac{2}{T-t}$ when f is completely flat; by expanding the squares, it reaches upper bound 2 when all weight is on one entry of f. Knowing that f in general has the shape of exponential decay, we know that the term is stable in a small range within the bounds before t gets close to T.

With the log-EDC having stable variance across time, running a least squares is feasible. In this case, the expectation of log-EDC is hard to derive analytically, but a second-order Taylor expansion approximation gives

$$\mathbb{E}[\log d(t)] \approx \mathbb{E}\left[\log \mathbb{E}[d(t)] - \frac{(d(t) - \mathbb{E}[d(t)])^2}{2\mathbb{E}^2[d(t)]}\right] \tag{37}$$

$$= \log \mathbb{E}[d(t)] - \frac{\mathrm{Var}[d(t)]}{2\mathbb{E}^2[d(t)]} \tag{38}$$

$$= \log g(t; \mathbf{A}^2) - \frac{\sum_{\tau=t}^{T} \big(f(\tau; \mathbf{A}^2)\big)^2}{\Big(\sum_{\tau=t}^{T} f(\tau; \mathbf{A}^2)\Big)^2}, \tag{39}$$

where the last fraction term is a smooth function bounded in $(0, 1)$, hence does not contribute much to the estimation of amplitudes. Ignoring this term, the resulting regression model is given by

$$\log d(t) = \log g(t; \mathbf{A}^2) + \xi_{\text{edc}}(t; \mathbf{A}^2), \tag{40}$$

where ξ_{edc} denotes the model error that is approximately zero-mean and Gaussian. Note that the errors are highly correlated because of cumulative summation; hence, downsampling is common in practice. The amplitude parameters can still be estimated by a naive nonlinear least squares:

$$\hat{\mathbf{A}}^2_{\text{EDC}} = \arg\min_{\mathbf{A}^2} \sum_{t=1}^{T} \xi^2_{\text{edc}}(t; \mathbf{A}^2). \tag{41}$$

4. EVALUATION

4.1. Simulation Based on Decaying Gaussian Noise

We compare the estimators from Sec. 3 using synthetic data to evaluate their accuracy and robustness under controlled decay conditions.

We generate multiple datasets with $N = 1000$ synthetic RIRs of chosen length and sample rate. Each RIR $h(t)$ is modeled as a sum of exponentially decaying Gaussian noise:

$$h(t) = \sum_{k=1}^{\kappa} A_k e^{\frac{-\delta_k t}{f_s}} z_k(t), \quad z_k(t) \sim \mathcal{N}(0, 1), \tag{42}$$

where the decay rate corresponds to chosen RT60 values.

Table 1: *Estimation bias and standard deviation [dB] in single-slope cases. LINEX achieves lowest variance with low bias.*

Estimator	Bias (std. dev.) [in dB]	
	Concert Hall	Classroom
LINEX	0.005 (0.094)	0.015 (0.357)
LS-PL	−0.009 (0.301)	−0.029 (0.689)
LS-LE	0.002 (0.216)	−0.002 (1.122)
EDC	−0.017 (0.097)	−0.193 (0.369)

We evaluate the four methods introduced in Section 3 to estimate the amplitude of the decays $\mathbf{A}^2$: LINEX regression, power-law nonlinear least squares (LS-PL), log-energy nonlinear least squares (LS-LE), and EDC-based nonlinear least squares. Optimization for LINEX regression is performed with MATLAB R2024b `fmincon` function (interior-point algorithm), while other least squares are solved with `lsqnonlin` (trust-region reflective algorithm). Convergence is declared when both step norm and change in objective function are below a set threshold, and both optimization has per step complexity $\mathcal{O}(T\kappa)$. Each method produces an estimate per RIR, and with the ground truth value, we can evaluate the distribution of estimation errors across the datasets.

First, we investigate single-slope cases, where we can drop k in the data-generating process (42). To align the synthetic data with empirical conditions, we consider two sets of parameters with contrasting levels of data richness:

- **Concert Hall:** We consider a typical mid-frequency RT60 of approximately 2 s, with a high-quality recording environment with SNR of 60dB [20]. The RIR length is set to 2 s to fully capture the decay before reaching the noise floor. A sampling rate of $f_s = 2$ kHz is used, which is approximately double the effective bandwidth to match the baseband equivalent. The decay rate is set to $\delta = 3.45$, corresponding to an RT60 of 2 s. The amplitude is fixed at -10 dB.[3]
- **Classroom:** We assume a low-frequency RT60 of approximately 0.8 s and a lower SNR. The RIR length is set to 0.6 s to capture the decay before it is overwhelmed by noise. A sample rate of $f_s = 250$ Hz is chosen to reflect the narrower low-frequency bandwidth. The decay rate is set to $\delta = 8.63$, and the amplitude gain is fixed at -10 dB.

Fig. 4 shows the empirical pdf of the estimated amplitudes. The LINEX and EDC methods yield narrow distributions, though EDC shows a small bias likely due to approximation in (39). NLS-based methods exhibit heavier tails. Table 1 summarizes performance.

The results indicate that, all estimators are approximately unbiased when sufficient data are available. However, LINEX and EDC-based methods are more efficient with a lower estimation variance. Note that in single-slope setting, both the log-squared RIR and the log-EDC regression models reduce to linear least squares formulations, making them computationally trivial.

Next, we consider a more complex scenario with three components in (42): two decaying slopes (δ_1, δ_2) and one noise term ($\delta_0 = 0$). Similarly to a single slope setting, we choose a sample rate f_s of 2 kHz to limit data richness, but a longer RIR length of 3 s to include the tail of noise floor.

[3]All amplitudes are transformed via logarithm and reported in dB for visual and interpretive clarity, specifically, $10\log_{10}(A^2)$.

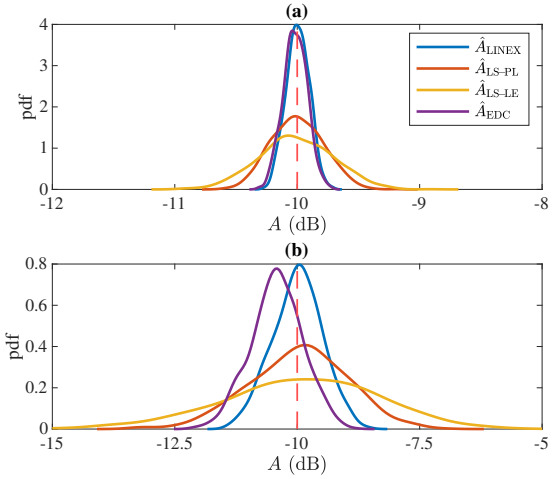


Figure 4: *Estimated amplitude distributions in single-slope cases, where **(a)** is data-richer than **(b)**.*

Let the two slopes have RT60s of 1 and 2 s, and the noise amplitude -40 dB. We consider two amplitude schemes: $A_1 = -15$ dB and $A_2 = -25$ dB where the lower-energy slope get masked between higher-energy slope and the noise; and $A_1 = A_2 = -20$ dB where the faster decaying slope gets masked since they start with same energy level.

Fig. 5 shows the empirical pdf of the four estimators of the three amplitude parameters A_1, A_2, and A_0 in dB in the two parameter settings described above. The results are consistent with our findings for the single-slope scenario, with LINEX and EDC methods being more efficient. Importantly, for less prominent slopes like A_2 in **(a)** and A_1 in **(b)**, all estimators are skewed to the right with a larger variance compared to the others. This is due to the dominance of different decays that take up parts of the model function; thus, it is intrinsically harder to estimate the amplitude of the decays that are masked. An example of a fitted model is shown in Fig. 6.

4.2. Simulation Based on a Synthetic Three-Room RIRs

A high-quality RIR dataset is synthesized in a coupled three-room scene using the Treble[4] suite by [10]. We preprocess the RIRs with DecayFitNet and perform a K-means clustering to identify the band-specific RT60s [21], and then perform a LINEX regression for each RIR to estimate the decay amplitudes of the slopes.

Fig. 7 shows the map of estimated amplitudes for each slope in the frequency band 500 Hz and 1000 Hz. The results show a clear separation of slope characteristics in each room, with the middle room being the most reverberant and the left room the most absorbent. Energy leakage across rooms is also visible, showing a smooth interpolation between the estimates. The results are also comparable to those of [10] numerically. This shows the application potential of the LINEX regression model.

[4]`https://www.treble.tech/`

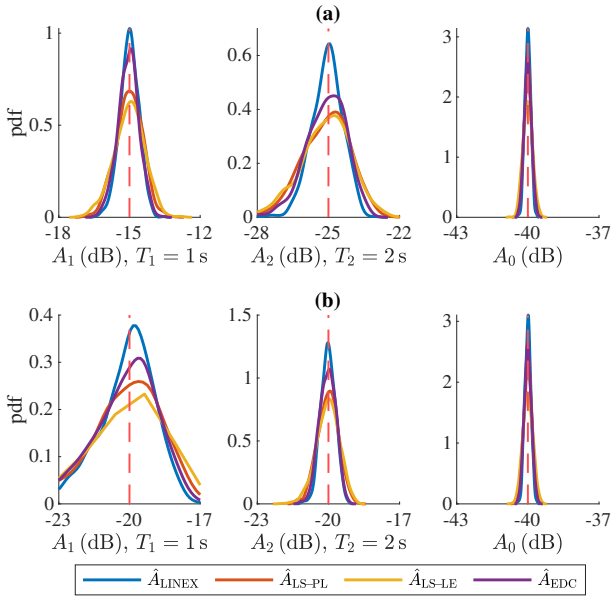


Figure 5: *Distribution of estimates of decay amplitudes in multi-slope cases.* ***(a)*** *highlights a slope weak in amplitude, and* ***(b)*** *highlights a slope masked because of fast decay rate.*

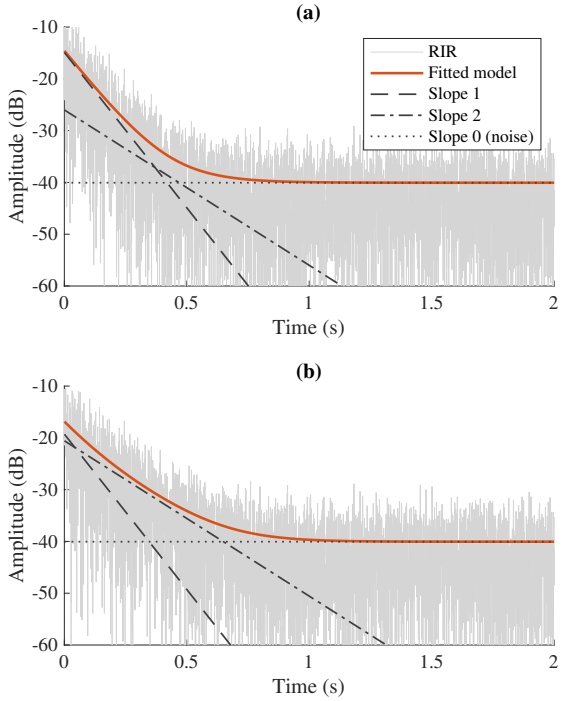


Figure 6: *An example of fitted LINEX regression model for both cases corresponding to Fig. 5. Slight masking can be seen in both cases.*

5. CONCLUSION

This paper presents a comprehensive study on amplitude estimation methods within the common-slope model of late reverberation. We refined and analyzed several classic estimation methods, including nonlinear least squares on linear, logarithmic, and power-law scales, regression based on energy decay curves (EDC), and proposed a novel LINEX regression derived from a Maximum Likelihood framework. Our theoretical formulation makes use of the statistical structure of RIRs, enabling more accurate modeling of decay amplitudes and better inferences.

Simulation results confirm that the LINEX and EDC-based approaches consistently outperform others in terms of efficiency, particularly in low-SNR or complex multi-slope settings. We also demonstrated the practical applicability of our approach to capture spatially varying decay structures. However, we acknowledge that our work is done on synthetic datasets only, and validation work using real-world recordings of complex acoustic scenes can be beneficial.

Future work may extend the use of the LINEX loss function in joint estimation of decay rates and amplitudes, which would further complete the analysis-to-synthesis pipeline on late reverb generation.

6. REFERENCES

[1] Georg Götz, Sebastian J. Schlecht, and Ville Pulkki, "Common-Slope Modeling of Late Reverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 3945–3957, 2023.

[2] Heinrich Kuttruff, *Room acoustics*, Taylor & Francis, London New York, 4. ed., transferred to digital printing edition, 2006.

[3] Jean-Dominique Polack, "Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics," *Applied Acoustics*, vol. 38, no. 2-4, pp. 235–244, 1993.

[4] M. R. Schroeder, "New Method of Measuring Reverberation Time," *The Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, Mar. 1965.

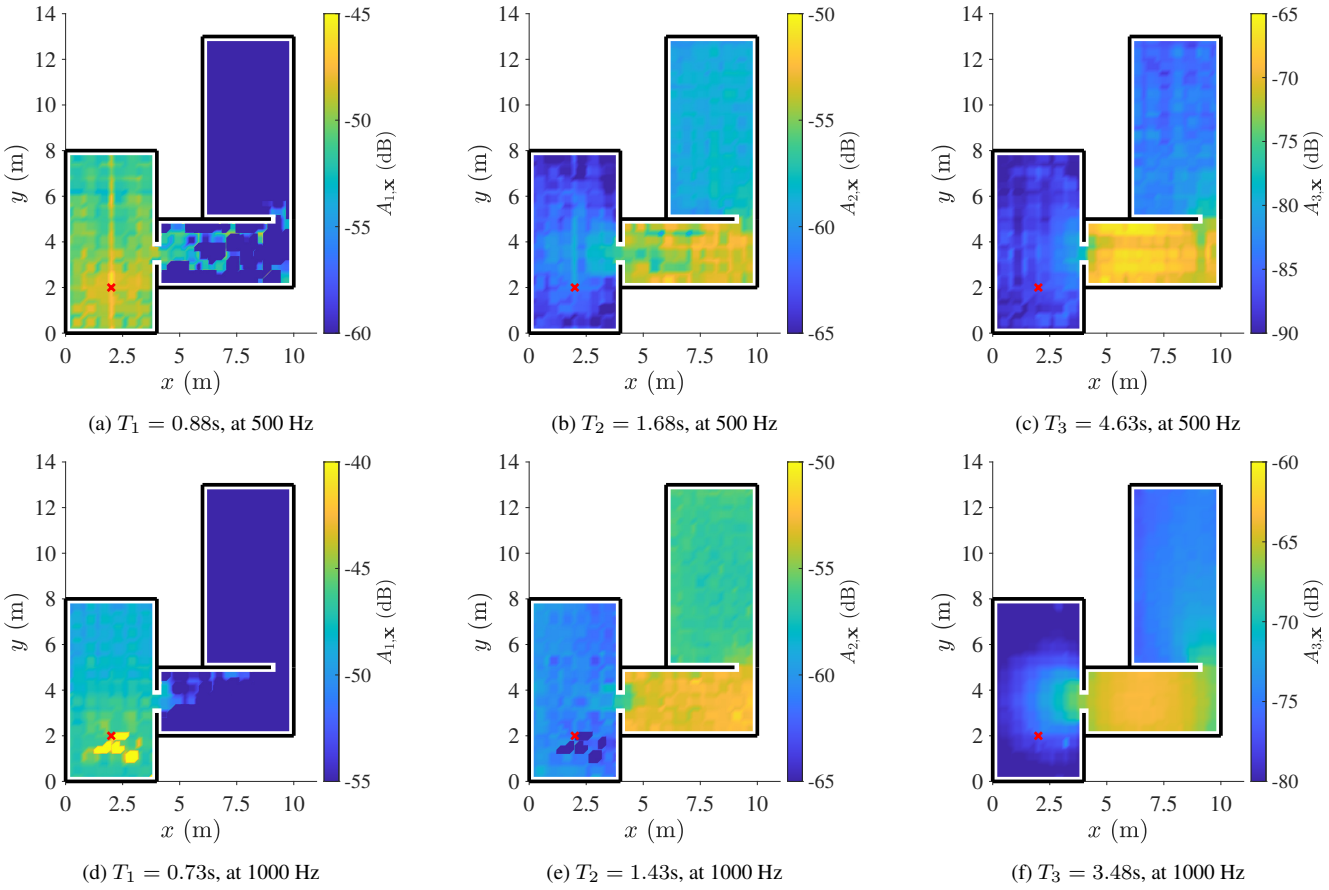
[5] Ning Xiang and Paul M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1415–1424, Sept. 2001.

[6] Jason E. Summers, Rendell R. Torres, and Yasushi Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 958–969, Aug. 2004.

[7] 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 Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 709–717, Nov. 1999.

[8] Mirosław Meissner, "Acoustic energy density distribution and sound intensity vector field inside coupled spaces," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 228–238, July 2012.

[9] Erling Nilsson, "Decay Processes in Rooms with Non-Diffuse Sound Fields Part I: Ceiling Treatment with Absorb-



(a) $T_1 = 0.88$s, at 500 Hz (b) $T_2 = 1.68$s, at 500 Hz (c) $T_3 = 4.63$s, at 500 Hz

(d) $T_1 = 0.73$s, at 1000 Hz (e) $T_2 = 1.43$s, at 1000 Hz (f) $T_3 = 3.48$s, at 1000 Hz

Figure 7: The amplitude map for each slope in the 500 Hz and 1000 Hz frequency band, estimated by LINEX regression. The red cross marks the sound source.

ing Material," *Building Acoustics*, vol. 11, no. 1, pp. 39–60, Mar. 2004.

[10] Georg Götz, Teodors Kerimovs, Sebastian J Schlecht, and Ville Pulkki, "Dynamic late reverberation rendering using the common-slope model," 2024.

[11] Kyung Yun Lee, Nils Meyer-Kahlen, Georg Götz, U. Peter Svensson, Sebastian J. Schlecht, and Vesa Välimäki, "Fade-in Reverberation in Multi-room Environments Using the Common-Slope Model," July 2024.

[12] Matti Karjalainen and Timo Peltonen, "Estimation of Modal Decay Parameters from Noisy Response Measurements," .

[13] Christian Schüldt and Peter Händel, "Decay Rate Estimators and Their Performance for Blind Reverberation Time Estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 8, pp. 1274–1284, Aug. 2014.

[14] A.V. Oppenheim and R.W. Schafer, *Discrete-time Signal Processing*, Prentice-Hall. Pearson, 2010.

[15] W.A. Gardner, *Introduction to Random Processes: With Applications to Signals and Systems*, Macmillan Publishing Company, 1986.

[16] Steven M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, Pearson Education.

[17] Wayne A. Fuller, *Measurement error models*, Wiley series in probability and mathematical statistics. Wiley, New York, 1987.

[18] Andreas Winkelbauer, "Moments and Absolute Moments of the Normal Distribution," July 2014.

[19] Steven E. Pav, "Moments of the log non-central chi-square distribution," Mar. 2015.

[20] William Cavanaugh, Gregory Tocci, Joseph Wilkes, and Philip Robinson, "Architectural Acoustics: Principles and Practice, 2nd Edition," *Noise Control Engineering Journal*, vol. 58, pp. 557, Jan. 2010.

[21] Georg Götz, Ricardo Falcón Pérez, Sebastian J. Schlecht, and Ville Pulkki, "Neural network for multi-exponential sound energy decay analysis," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 942–953, Aug. 2022.