PROBABILISTIC REVERBERATION MODEL BASED ON ECHO DENSITY AND KURTOSIS

Champ C. Darabundit *
CIRMMT †
CAML‡ , McGill University
Montréal, Canada
champ.darabundit@mail.mcgill.ca

Jonathan S. Abel
CCRMA §
Stanford University
Stanford, CA
abel@ccrma.stanford.edu

Wieslaw Woszczyk
CIRMMT
Sound Recording, McGill University
Montréal, Canada
wieslaw.woszczyk@mcgill.ca

ABSTRACT

This article proposes a probabilistic model for synthesizing room impulse responses (RIRs) for use in convolution artificial reverberators. The proposed method is based on the concept of echo density. Echo density is a measure of the number of echoes per second in an impulse response and is a demonstrated perceptual metric of artificial reverberation quality. As echo density is related to the statistical measure of kurtosis, this article demonstrates that the statistics of an RIR can be modeled using a probabilistic mixture model. A mixture model designed specifically for modeling RIRs is proposed. The proposed method is useful for statistically replicating RIRs of a measured environment, thereby synthesizing new independent observations of an acoustic space. A perceptual pilot study is carried out to evaluate the fidelity of the replication process in monophonic and stereo artificial reverberators.

1. INTRODUCTION

The reverberation of acoustic space is characterized by how a sound is reflected and absorbed as it travels from a source to a listener. Typically, a listener will observe the direct sound from a source, followed by a series of distinct echoes. These echoes are early reflections signifying the apparent geometries — such as nearby walls — of the acoustic space. The echoes in the space will rapidly build up, layering upon one another until giving way to a dense late reverberation. Over the same period, due to the absorptive properties of air and the surrounding environment, the acoustic energy will decay until the environment is actuated again [1].

The same observations can be made by a time domain analysis of a room impulse response (RIR), which may be procured through techniques such as balloon pop [2], sine sweep [3], or maximum-length sequence [4] measurements. Further analysis of RIRs in the frequency domain can highlight the behavior of modes and their decay. In the time domain, measures such as decay time (T_{60}) and echo density can be obtained, the latter of which is the focus of this article.

Absolute echo density (AED) measures the number of echoes per second, that is to say, the rate of non-zero impulses, in an RIR. In 1961, Schroeder noted that a property of "acoustically good rooms" was a high echo density and focused on designing artificial reverberators with a high echo density [5, 6]. Moorer, in 1979, observed that the distribution of echoes in an RIR becomes Gaussian in nature as the reflections become well-mixed. Echo density has been evaluated as a perceptually relevant reverberation parameter in [7, 8], and has been used as a measure to control the mixing time in feedback delay networks in [9]. Synthesis of RIRs based on a desired echo density for use in convolution reverberators has been pursued using a Poisson process [7, 8, 10], sparse FIR filters [11, 12], and velvet noise [13]. Echo density-focused methods permit RIRs synthesis from early reflections onward. In comparison, Gaussian noise synthesis methods described in [14, 15], only accurately model the late reverberation.

This article builds upon work by previous collaborators on the concept of normalized echo density (NED) [7, 8, 10, 16]. NED is a measure that compares the distribution of an RIR in a sliding time window to the Gaussian distribution and, as a result, estimates the echo density of an RIR. NED is inversely proportional to the statistical measure of kurtosis. In our contribution, we propose modeling the statistics of an RIR with a probabilistic mixture model. A mixture model characterizes a relatively complex probability distribution by modeling the distribution as a weighted sum of more basic distributions. Mixture models, particularly Gaussian mixture models (GMMs), have found wide usage in audio machine learning for tasks such as speaker identification [17–19]. GMMs have also been used to remove reverberation from sonar signals [20]. However, using mixture models to synthesize RIRs is — to the authors' knowledge — a novel approach. The proposed method has the potential to be an efficient and adaptive algorithm for synthesizing RIRs which provides better characterization of the RIR compared to prior methods.

Section 2 will review the proposed measures of echo density and the relationship between echo density and kurtosis. Section 3 will describe the method of moments used to derive the weights of the proposed mixture model and Section 4 will describe the mixture parameters used when modeling RIRs. Section 5 will detail example applications of the proposed method. Section 6 will detail a perceptual study with results presented in Section 7. Section 8 will conclude.

2. ECHO DENSITY MEASURES

Echo density, in units of echoes per second, was first proposed as a measure of reverberation quality in [5] and is now referred to as the absolute echo density (AED). This is to distinguish AED from the measure of normalized echo density (NED) [10].

* Research partially conducted while a Masters student at CCRMA
† Center for Interdisciplinary Research in Music, Media, and Technology
‡ Computational Acoustic Modeling Laboratory
§ Center for Computer Research in Music and Acoustics

2.1. Normalized and absolute echo density

NED, $\eta(t)$, was first proposed by Abel and Huang in [16] and is a statistical measure that estimates, within a time window, the departure of an RIR's distribution from the Gaussian distribution.

$$\eta(t) = \frac{1}{\text{erfc}(1/\sqrt{2})} \sum_{\tau=t-\beta}^{t+\beta-1} w(\tau) \mathbf{1}\left\{|h(\tau)| > \sigma\right\}. \quad (1)$$

Here, $h(t)$ is a time windowed portion of the RIR centered around t and 2β long in samples, $w(t)$ is a window function, and σ is the standard deviation of the $h(t)$. $\mathbf{1}\{\cdot\}$ is the indicator function, producing one when the argument is true and zero when false. We assume $h(t)$ to have zero-mean such that the standard deviation is:

$$\sigma = \left[\sum_{\tau=t-\beta-1}^{t+\beta} w(\tau) h^2(\tau)\right]^{\frac{1}{2}}. \quad (2)$$

The NED, η, can be related to AED, ρ, in echoes per second, through the following expression,

$$\rho = \frac{1}{\tau_d} \frac{\eta}{1-\eta}, \quad (3)$$

where τ_d is the echo duration in seconds. As η approaches 1, the AED in the window will increase describing a well-mixed reverberation. Conversely, as η approaches 0 the AED will decrease describing the sparse reflections found during early reflections.

2.2. Kurtotic measure of echo density

Abel and Huang, also proposed an alternative measure of echo density with a close similarity to the NED [16]. Their metric is related to the statistical metric of kurtosis α_4,

$$\eta_k(t) = \frac{\sigma}{\left[\frac{1}{3}\sum_{\tau=t-\beta}^{t+\beta-1} w(\tau) h^4(\tau)\right]^{\frac{1}{4}}} \propto (\alpha_4)^{-\frac{1}{4}}. \quad (4)$$

A measure of echo density based only on kurtosis was proposed independently by Usher in [21].

Kurtosis is defined as the fourth standard moment and has been incorrectly ascribed to the peakedness of a distribution or, alternatively, the fatness of a distribution's tails. The measure, however, has no direct bearing on these components. Kurtosis is better thought of as a movement of a distribution's mass from its shoulders to its center and tails [22, 23].

In this regard, kurtosis is an ideal measure of echo density. A window with a sparse number of echoes will have a distribution concentrated about its center and tails as the signal is dominated by silence and the occasional reflection. The distribution of said window will be highly kurtotic. Correspondingly, a well-mixed window will not be kurtotic, as its distribution is Gaussian.

3. A MIXTURE MODELING APPROACH

A probabilistic reverberation model aims to generate a synthetic RIR based on desired statistical properties. In [10], this was done using a Poisson process. This process, however, becomes computationally expensive for dense reverberations as echoes must be generated on an echo-by-echo basis. We propose using a mixture modeling approach to capture the statistics of an RIR.

The objective of a mixture model is — most often — to approximate the probability density of an empirical observation using a linear superposition of components with simpler density functions. As such, the primary task of mixture modeling is to derive the mixture parameters which best approximate a given observation. These parameters are the number of components, the parameters of individual component density functions, and the weights of each component within the mixture model.

The distribution of a RIR within a time window is relatively simple and assumptions can be made regarding the necessary number of components and their individual parameters. These parameters are discussed in Section 4.1. We are then chiefly concerned with deriving the weights of a finite set of components with known probability densities. To accomplish this we will use the method of moments, similarly described in [24, 25] instead of the more common maximum-likelihood-based expectation maximization (EM) method [17]. The method of moments is based on matching the raw moments of a mixture distribution to the raw moments of a desired distribution. Raw, central, and standard moments are common statistical measures used to evaluate statistical properties such the mean, variance, skewness, and kurtosis [26].

3.1. Raw, central, and standard moments

The n$^{\text{th}}$ raw moment of a random variable X with a continuous probability density function $f(x)$ can be computed using the expectation operator $E[\,\cdot\,]$:

$$\nu_n = E\left[X^n\right] = \int_{x\in\mathbb{R}} x^n f(x) dx. \quad (5)$$

For a discrete random variable, $E[\,\cdot\,]$ is given by:

$$E\left[X^n\right] = \sum_{l=0}^{L-1} x_l^n f(x_l). \quad (6)$$

For an empirical observation, $f(x_l) = \frac{1}{L}\ \forall x_i$. The first raw moment is the mean, represented by $\bar{X} = \nu_1$, and the central moment is a raw moment evaluated about the mean:

$$\mu_n = E\left[(X - \bar{X})^n\right]. \quad (7)$$

If the distribution has zero-mean, the central moment is equivalent to the raw moment. The second central moment is the variance $\mu_2 = \sigma^2$, the square of the standard deviation. The standard moment is the central moment scaled by the standard deviation,

$$\alpha_n = E\left[\left(\frac{x-\bar{x}}{\sigma}\right)^n\right]. \quad (8)$$

Kurtosis is, as previously mentioned, the fourth standard moment and can be expressed in terms of the second and fourth central moments

$$\alpha_4 = \frac{\mu_4}{\mu_2^2}, \quad (9)$$

In the context of modeling an RIR, parameterizing our model based on desired moments is intuitive if we aim to synthesize a response with a desired echo density and therefore kurtosis.

3.2. Method of moments

The method of moments aims to match the moments of a desired distribution, $\hat{Z}$, to the moments of a mixture distribution. The probability density function, $f_Z(x)$, of a mixture distribution Z is generated by the linear superposition of a finite set of M components with probability density functions $f_m(x)$ weighted by π_m,

$$f_Z(x) = \sum_{m=1}^{M} \pi_m f_m(x). \tag{10}$$

The weights, π_m, represent the amount by which each component is sampled. As such, the total sampling must sum to unity and each weight bounded:

$$\sum_{m=1}^{M} \pi_m = 1, \quad 0 \leq \pi_m \leq 1. \tag{11}$$

Applying (5) to (10), consider that the n$^{\text{th}}$ raw moment of the mixture distribution is the dot product of the mixture weights and the n$^{\text{th}}$ raw moment of the components:

$$E\left[Z^n\right] = \sum_{m=1}^{M} \pi_m \int_{x\in\mathbb{R}} x^n f_m(x) dx = \langle \pi \mid \nu_n \rangle \tag{12}$$

The method of moments is then formulated as a linear system relating the raw moments of a mixture's M components to the desired raw moments and accounting for the constraint in (11):

$$\boldsymbol{M}\boldsymbol{\pi} = \boldsymbol{\nu}, \tag{13}$$

where

$$\boldsymbol{M} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \nu_{1,1} & \nu_{1,2} & \cdots & \nu_{1,M} \\ \nu_{2,1} & \nu_{2,2} & \cdots & \nu_{2,M} \\ & & & \vdots \\ \nu_{n,1} & \nu_{n,2} & \cdots & \nu_{n,M} \end{bmatrix} \begin{matrix} \boldsymbol{\pi} = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_M \end{bmatrix}^T \\ \boldsymbol{\nu} = \begin{bmatrix} 1 & \hat{\nu}_1 & \cdots & \hat{\nu}_n \end{bmatrix}^T \end{matrix}$$

The moment matrix, $\boldsymbol{M}$, is a $n+1$ x M matrix formed by n$^{\text{th}}$ raw moments $\nu_{n,m}$ of the m$^{\text{th}}$ component. The weight vector, $\boldsymbol{\pi}$, is found by applying the inverse moment matrix to the desired moment vector, $\boldsymbol{\nu}$, of $\hat{Z}$ with moments $\hat{\nu}_n$. For M components, a unique system necessarily requires the computation of $n = M-1$ moments. After solving (13) for $\boldsymbol{\pi}$, the desired distribution is synthesized by random sampling of pseudo-random sequences with the same distribution as each component.

In this article, we utilize a mixture model based on the Gaussian distribution $X_i \sim \mathcal{N}(\nu_i, \sigma_i)$. Each component is parameterized by its mean μ_i and standard deviation σ_i. We chose Gaussian components because the desired response is generated by random sampling of scaled and shifted white Gaussian noise sequences. The Gaussian distribution has the following probability density function,

$$f\left(x \mid \nu, \sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\nu)^2}{\sigma^2}}. \tag{14}$$

The normal distribution, along with other distributions such as the binomial, gamma, and the Poisson families of distributions, has the property that its variance is — at most — a quadratic function of its mean value $\bar{x}$ [24, 27]. For these distributions, the moment matrix is readily generated based on only the parameters of the component distributions. Table 1 provides the first four raw moments of the Gaussian distribution.

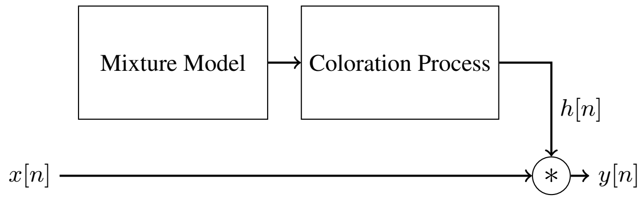


Figure 1: *Block diagram of mixture model-based RIR synthesis. A mixture model generates the desired "colorless" distribution which is filtered to generate the RIR $h[n]$. This impulse response is convolved with the input signal $x[n]$ to generate the reverberated signal $y[n]$.*

Take note that the method of moments in (13) does not guarantee $0 \leq \pi_m \leq 1$ and assumes the moments, and therefore the parameters, of the components are known. Tuning the component parameters to generate non-negative weights is an iterative optimization process [24]. However, based on prior knowledge of RIR signals, assumptions can be made regarding the necessary number of components and their individual parameters. These assumptions are described in the next section.

4. PROBABILISTIC MODELING BASED ON ECHO DENSITY

Our proposed probabilistic RIR model is compromised of two separate modeling processes, described in a block diagram in Figure 1. The first model is the mixture model which will generate what we denote the "echo response," a noise-like "colorless" impulse response based on the desired moment vector $\boldsymbol{\nu}$. Once the echo response has been generated, the color of the reverberation is modeled by a coloration process. Because the distribution of silence in each echo response is stochastic, RIRs synthesized with the same color will not be statistically correlated. The parameters of both models evolve with time.

This article is primarily concerned with the synthesis of the echo response, and we utilize an STFT-based frequency domain convolution to impose the spectra of a measured RIR onto the echo response. Similar RIR synthesis procedures have used equal rectangular bandwidth (ERB) band-wise exponential decay to color Poisson process generated noise [8]. Time-varying lowpass coloration filters have been used to color velvet noise [28]. In this section, we determine the parameters of the proposed RIR mixture model based on assumptions regarding the distribution of RIRs.

Table 1: *First four raw moments of $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$*

ν_1	μ_i
ν_2	$\mu_i^2 + \sigma_i^2$
ν_3	$\mu_i^3 + 3\mu_i\sigma_i^2$
ν_4	$\mu_i^4 + 3\sigma_i^4 + 6\mu_i^2\sigma_i^2$

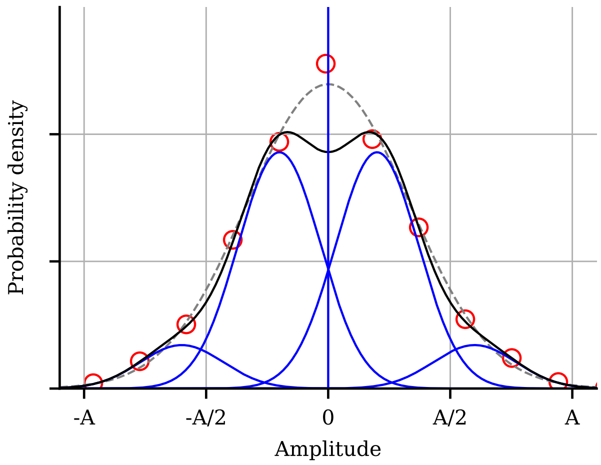


Figure 2: *Approximation of a Gaussian distribution* $f\left(x \mid 0, A/3\right)$ *(dashed grey) with the proposed mixture components (blue). The summation of the mixture components (black) matches the desired distribution except near zero where the model fails to capture the contribution from the degenerate distribution. A histogram of noise generated by the mixture — with bin centers and values indicated by the red scatter plot — demonstrates close adherence to the desired normal distribution.*

4.1. Mixture model parameters for RIR synthesis

In regards to the number of components, we assume that echo density — and consequently kurtosis — suitably characterizes the behavior of an RIR and it is unnecessary to consider moments higher than the fourth moment. By the nature of (13), a unique system necessitates a mixture with only five components.

In regards to the component parameters, consider that a large portion of an RIR signal during early reflections and into the late reverberation is the silence between echoes. To characterize this, we suggest that one component of the mixture should be a Gaussian with zero-mean and a variance that approaches an infinitely small value,

$$f_0(x) = \lim_{\sigma \to 0} f\left(x \mid \nu = 0, \sigma\right) = \delta(x). \tag{15}$$

The density function of this component can be abstractly represented by a delta function, and all raw moments of this component are equal to zero: $\nu_n = 0 \ \forall n$. This distribution is also referred to as a degenerate distribution [29].

We then assume — without loss of generalization — that the window has zero-mean and is normalized such that the samples lie within the amplitude range $\pm A$. The remaining components $(M = 4)$ are then evenly spaced in the range:

$$\nu_m = -A + \frac{2A}{M+1} m, \quad m = 1, 2, \ldots, M. \tag{16}$$

We propose that these components exhibit the same standard deviation. Overlapping the distribution is necessary to ensure a continuous distribution, but too much overlapping promotes the generation of negative weights as we are — in a sense — oversampling our distribution. Too little overlapping creates gaps in the amplitude distribution and the distribution is undersampled.

To ensure an ideal overlap of our component density functions, we proposed that the intersection between two neighboring Gaussians should sum to the maximum value of each component's density function. Determining the intersection of Gaussian probability densities $f\left(x \mid \nu_m, \sigma\right)$ and $f\left(x \mid \nu_{m+1}, \sigma\right)$ by substituting the expressions for ν_m and ν_{m+1} from (16) into (14), the point of intersection is,

$$x = \frac{A}{M+1}\left(2m - M\right). \tag{17}$$

The result is substituted back into (14) and evaluated against half the maximum value, $\frac{1}{2}\frac{1}{\sigma\sqrt{2\pi}}$. The proposed standard deviation for the components is:

$$\sigma = \frac{A}{M+1}\sqrt{\frac{1}{2\log(2)}} \tag{18}$$

Even with these parameters, it is possible to generate negative sampling weights. To counteract this, we propose that any negative weights are zero-ed and the resulting $\boldsymbol{\pi}$ vector is rescaled to observe the constraint in (11)

$$\pi_i = \begin{cases} \pi_i, & 0 \leq \pi_i \leq 1 \\ 0, & \pi_i < 0 \end{cases} \tag{19}$$

If the desired amplitude A is normalized, then the parameters of our components are predetermined and the inverse moment matrix $\boldsymbol{M}^{-1}$, can be stored ahead of computation. Since the desired distribution is assumed to have zero-mean, the desired moment vector $\boldsymbol{\nu}$ simplifies to,

$$\boldsymbol{\nu} = \begin{bmatrix} 1 & 0 & \mu_2 & 0 & \mu_4 \end{bmatrix}^T \tag{20}$$

In Figure 2, the proposed mixture is used to approximate a Gaussian distribution with $f\left(x \mid \mu = 0, \sigma = A/3\right)$. The sum of the component density functions, in black, approximates the Gaussian distribution except near zero where the contribution of the degenerate distribution is not well characterized. The histogram of a simulation, indicated by a red scatter plot, instead verifies the mixture behavior.

5. EXAMPLES

Given a desired NED in (4) and the definition in (9), it is difficult to make assumptions about the necessary values of μ_2 or μ_4. In the following examples, we demonstrate how to determine the desired moment vector for a desired AED value based on the properties of a generalized Poisson process. Alternatively, the desired moment vector can be found empirically when statistically replicating an RIR.

5.1. Static echo density based on the Poisson distribution

The generalized Poisson process proposed in [10] forms the process:

$$h(t) = \sum \alpha(t) \cdot \delta(t - \tau(t)). \tag{21}$$

The arrival times are represented as a set of Poisson impulses where $\tau(t)$ is drawn from an exponential distribution $\tau(t) \sim \exp\left\{-t/\rho(t)\right\}$ and ρ is the AED at time t. The amplitude, conversely, is drawn from a Gaussian distribution with a variance also dependent on the AED, $\alpha(t) \sim \mathcal{N}\left(0, \sqrt{1/\rho(t)}\right)$.

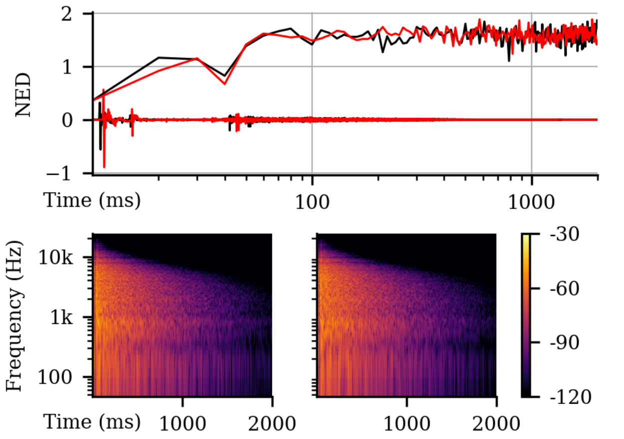


Figure 3: *Comparison of the left (black) and right (red) RIR signals and NED measured in Pollack Hall at McGill University. Spectrograms of each are displayed on the bottom left and right, respectively. The NED profiles are similar while the spectrograms are nearly identical.*

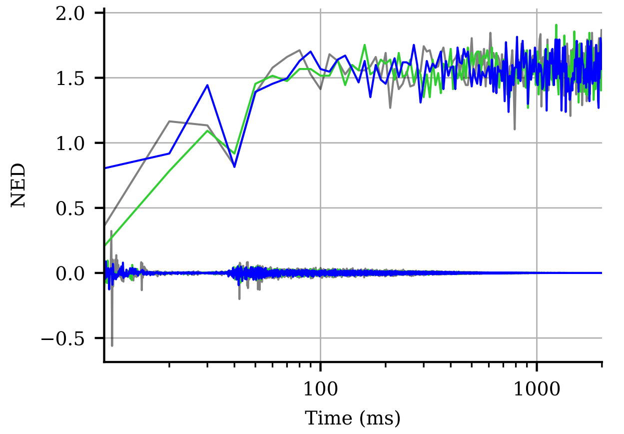


Figure 4: *RIR and NED of a measure (grey) of the Pollack Hall at McGill University replicated using the proposed method without block switching (blue) and with block switching (green). The synthesis window is 5 ms long while the analysis window was 10 ms long.*

The second and fourth moments of the Poisson process described by [10] can be derived based on the properties of a Poisson impulse [26]. A proof deriving the fourth moment of the Poisson process is provided in Section 11. The resulting second and fourth central moments are:

$$\mu_2^{PP} = \frac{1}{\rho} \cdot \left(\left(\frac{\rho}{f_s} \right)^2 + \frac{\rho}{f_s} \right) \tag{22a}$$

$$\mu_4^{PP} = \left(\frac{1}{\rho} \right)^2 \cdot \left(\left(\frac{\rho}{f_s} \right)^4 + 3\frac{\rho}{f_s} \right). \tag{22b}$$

These results can be substituted into the desired moment vector in (20). With these parameters, the proposed mixture described in Section 4.1 can generate a window of impulses with a fixed echo density. This method can be generalized to generate a synthetic RIR with a time-varying NED or AED profile.

5.2. Statistical replication of RIRs based on their moments

Another application of the method is for statistically replicating RIRs of measured rooms. The RIR replicas are akin to an independent observation of the acoustic space. Consider Figure 3, which compares the RIR measurement of Pollack Hall at McGill University with two microphones spaced roughly a human head width apart. The NED profile and spectrogram are nearly identical, and the two responses mainly differ in the arrival time of the reflections. Replicated RIRs can be used artistically in virtual acoustics to simulate different observations of a measured space.

To replicate a measured RIR, we propose using the overlap-add (OLA) algorithm [30] to analyze the raw moments of the measured RIR. Within each window, the echo response is generated and then colored in the spectral domain. In comparison to the Poisson process, the mixture model approach makes no assumptions about the distribution of the measured responses and instead attempts to holistically replicate the measured distribution.

Window size is an important parameter in the proposed method, as the generated echoes are randomly distributed within the window. This can result in temporal smearing when replicating early reflections as energy is no longer concentrated about distinct echoes. Smearing can be decreased by using a smaller window size at the cost of frequency resolution in the coloration process. A possible solution is to use a window size switching scheme akin to how transients are analyzed in audio coding [31]. In this scheme, a small window replicates early reflections and a longer window replicates the late reverberation. The result of the replication process with a single-window size and switched window size is shown in Figure 4. The main window size for both methods was 5 ms, and the small window size was 1.25 ms. The window size was switched after 30ms. These values were heuristically chosen and a formal metric based on NED merits further investigation.

The first reflections are largely deterministic based on the geometry of the acoustic space. When replicating RIRs it is better — in practice — to mix the first reflections from the measured RIR with early reflections onward from the replica. This is achieved by crossfading the measured RIR with the replicated RIR following a few early reflections. For the experiment described in Section 6, the signals were mixed 30ms after the initial direct path impulse with a mixing time of 5ms.

6. PERCEPTUAL EXPERIMENTS

Two informal perceptual pilot studies were conducted on students from CIRMMT and CCRMA. The first study compared the smoothness of noise generated with the Poisson process and the proposed mixture model. The mixture was generated for varying NED values using the parameters discussed in Section 5.1. The second study evaluated the efficacy of RIR replication using the Poisson process and the method discussed in Section 5.2. Both studies were administered online using the BeaqleJS framework [32]. All samples used in the study had their loudness normalized based on the EBU R 128 recommendation [33]. Samples

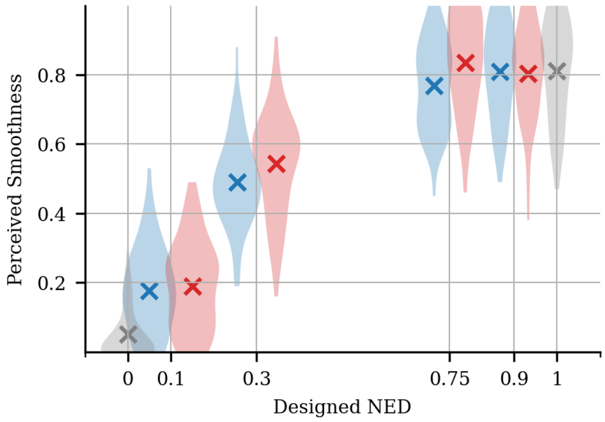


Figure 5: *Violin plot of perceived smoothness versus designed NED value. The Poisson process (blue) and mixture-generated samples (red) evaluations have been, respectively, shifted to the left and right from the designed values. The reference samples are in grey, and the "sputtery" reference with $\eta = 0.05$ is offset to the left.*

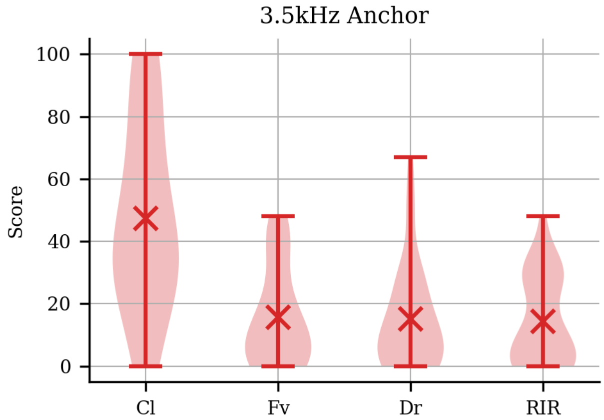


Figure 6: *Comparison of 3.5kHz lowpass anchor responses for clarinet (Cl), female voice (Fv), drums (Dr), and RIRs (RIR).*

used in the experiment can be accessed online [1].

6.1. Noise smoothness evaluation

Participants were asked to rate the smoothness of noise generated using the Poisson process and the proposed mixture for NED values $\eta = 0.1, 0.3, 0.75$, and 0.9 with a bandwidth of 10 kHz. Participants were asked to rate each noise sample on a scale from "sputtery" $[0 - \frac{1}{3})$ to "rough" $[\frac{1}{3} - \frac{2}{3})$ to "smooth" $[\frac{2}{3} - 1]$, terminology borrowed from an earlier study [10]. The participants were given two reference signals to ground their evaluations: a smooth reference — Gaussian noise — and a sputtery reference — Poisson process noise with $\eta = 0.05$.

6.2. RIR replication evaluation

Participants were asked to subjectively rate the quality of monophonic and stereo reverberated signals in a multi-stimulus test with hidden reference and anchor (MUSHRA) style test [34]. The test included a training phase where participants were familiarized with the process using a sample not included in the main evaluation and an artificial impulse response.

For the main evaluation, two environments were measured as references: Pollack Hall at McGill University and Memorial Church at Stanford University. RIR measurements of both environments were captured using two microphones representing the left and right channels. In the monophonic evaluations, only the left channel was utilized. Both channels were used in the stereo evaluations.

Test signals were generated by convolving generated and measured RIRs with anechoic audio samples. The anechoic audio samples consisted of female voice speech, drum, and clarinet signals. The voice and drum signal were chosen as they are largely impulsive in nature and the clarinet was chosen as it is pitched and less impulsive comparatively. The test RIR replicas were created using the proposed method and the Poisson process. Anchor RIRs were generated by filtering the measured RIRs at 3.5 and 7 kHz based on recommendations in [34]. Participants were additionally asked to evaluate the RIR signals by themselves.

[1] https://ccrma.stanford.edu/∼champ/dafx23

7. RESULTS AND ANALYSIS

Both experiments had a total of 10 participants. However, one participant was removed from the RIR replication study as they consistently rated the anchor and reference signals as having an equal perceptual quality.

7.1. Noise smoothness results

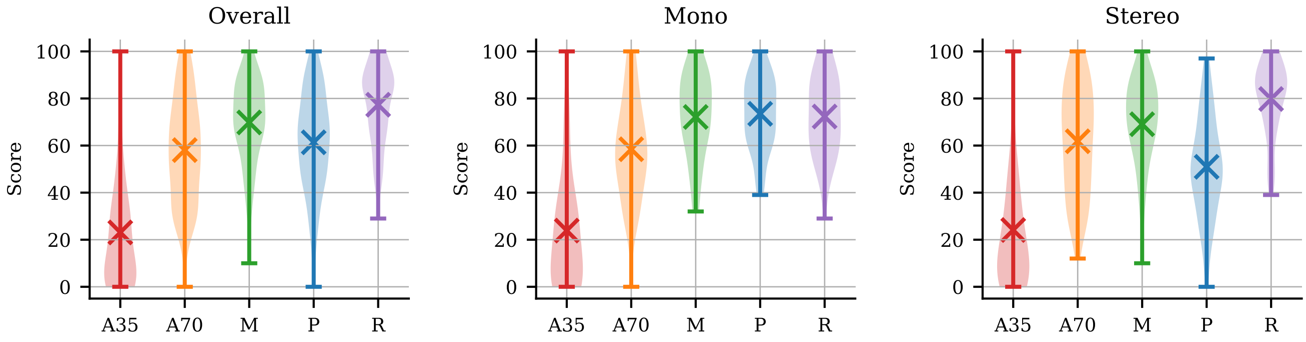
The results of the noise smoothness study are shown in Figure 5. Our experiment demonstrates that samples generated with the proposed mixture model have a similar perceived smoothness as samples generated through the Poisson process. The proposed mixture has a higher mean perceived smoothness compared to the Poisson process for three NED levels. These results suggest that the proposed model performs similarly to prior methods for generating samples with fixed NEDs.

7.2. RIR replication study results

The overall results for all samples are shown in a violin plot in Figure 7a. The overall mean rating for the proposed method is evaluated as being second in quality to the reference. However, it is worth recognizing that participants had difficulty discriminating the hidden anchors and there were a small number of participants in the pilot study.

Evaluation was particularly difficult in the case of the clarinet sample which was the only audio sample with sustained sounds. A comparison of the 3.5kHz lowpass anchor for different samples is shown in Figure 6. This would suggest that the evaluation of RIR quality is better performed with audio signals that are transient or with the RIR signal by itself.

Further analysis is obtained if samples are delineated by monophonic or stereo test signals as in Figure 7b and Figure 7c, respectively. In Figure 7b, the performance of the proposed mixture



(a) (b) (c)

Figure 7: *Violin plot of MUSHRA-style RIR replication study. (a) compares the overall results, (b) the monophonic sample results, and (c) the stereophonic sample results. A35 corresponds to the 3.5kHz lowpass anchor, A70 to the 7.0kHz lowpass anchor, M to the proposed mixture model, P to the Poisson process, and R to the reference.*

model is similar to the Poisson process method for monophonic samples. However, in Figure 7c, the performance of the proposed mixture method is better than that of the Poisson process method stereo samples. This suggests that when rendering virtual acoustic scenes, the proposed method may provide better results.

8. CONCLUSION AND FUTURE WORK

In this article, we proposed a new method for synthesizing artificial RIRs for use in convolution reverberators. The proposed method uses a mixture model with component parameters designed specifically for RIR synthesis. A pilot perceptual study demonstrates the efficacy of the proposed method in comparison to a previously proposed RIR modeling technique. Future work is needed to properly evaluate the quality of the proposed method in a larger study. Frequency domain convolution was used to color the generated echo responses, and future work should evaluate the quality of different coloration techniques.

9. ACKNOWLEDGMENTS

The authors would like to thank all the volunteers at CIRMMT and CCRMA for taking the time to participate in the pilot study. The authors would like to thank Gary Scavone for his discussions on designing the pilot study.

10. REFERENCES

[1] Thomas D. Rossing, Ed., *Springer Handbook of Acoustics*, Springer, New York, NY, 2 edition, 2014.

[2] David Griesinger, "Beyond MLS — occupied hall measurement with FFT techniques," in *101st Convention Audio Eng. Soc.*, November 1996.

[3] Swen Müller and Paulo Massarani, "Transfer-function measurement with sweeps," *J. of the Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, June 2001.

[4] Douglas D. Rife and John Vanderkooy, "Transfer-function measurement with maximum length sequences," *J. of the Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, June 1989.

[5] M. R. Schroeder and B.F. Logan, "'Colorless' artifical reverberation," *J. of the Audio Eng. Soc.*, vol. 9, no. 3, pp. 192–197, July 1961.

[6] M. R. Schroeder, "Natural sounding artificial reverberation," *J. of the Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223, July 1962.

[7] Patty Huang, Jonathan S. Abel, Hiroko Terasawa, and Jonathan Berger, "Reverberation echo density psychoacoustics," in *125th Convention Audio Eng. Soc.*, October 2008.

[8] Kimberly Kawczinski and Jonathan S. Abel, "Perceptually-informed features for room reverberation modeling and prediciton," in *149th Convention Audio Eng. Soc.*, October 2020.

[9] Sebastian J. Schlecht and Emanu el A. P. Habets, "Feedback delay networks: Echo density and mixing time," in *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, 2017.

[10] Patty Huang and Jonathan S. Abel, "Aspects of reverberation echo density," in *123rd Convention Audio Eng. Soc.*, October 2007.

[11] Per Rubak and Lars G. Johansen, "Artificial reverberation based on a pseudo-random impulse response," in *104th Convention Audio Eng. Soc.*, May 1998.

[12] Per Rubak and Lars G. Johansen, "Artificial reverberation based on a pseudo-random impulse response II," in *106th Convention Audio Eng. Soc.*, May 1999.

[13] Matti Karjalainen and Hanna Jäveläinen, "Reverberation modeling using velvet noise," in *30th Int. Conf. Audio Eng. Soci.*, Saariselk a, Finland, March 2007.

[14] Jean-Marc Jot, Laurent Cerveau, and Olivier Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," in *103rd Convention Audio Eng. Soc.*, September 1997.

[15] James A. Moorer, "About this reverberation business," *Comput. Music J.*, vol. 3, no. 2, pp. 13 – 27, 1979.

[16] Jonathan S. Abel and Patty Huang, "A simple, robust measure of reverberation echo density," in *121st Convention Audio Eng. Soc.*, October 2006.

[17] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag, Berlin, 2006.

[18] Geoffrey J. McLachlan and David Peel, *Finite Mixture Models*, John Wiley & Sons, Inc., New York, 2000.

[19] Douglas A. Reynolds, *A Gaussian mixture modeling approach to text-independent speaker identification*, Ph.D. thesis, Georgia Institute of Technology, August 1992.

[20] Guo Haoquan, Chu Fuzhao, and Zhu Daizhu, "Research on Gaussian mixture auto-regressive reverberation modeling and whitening algorithm," in *2021 IEEE Conf. Signal Proces., Commun., and Comput.*, 2021.

[21] John S. Usher, *Subject evaluation and electroacoustic theoretical validation of a new approach to audio upmixing*, Ph.D. thesis, McGill University, Montreal, Canada, September 2006.

[22] J.J.A. Moors, "The meaning of kurotsis: Darlington reexamined," *The American Statistician*, vol. 40, no. 4, pp. 283–284, November 1986.

[23] Kevin P. Balanda and H. L. MacGillivray, "Kurtosis: A critical review," *The American Statistician*, vol. 42, no. 2, pp. 111–119, May 1998.

[24] Bruce G. Lindsay, "Moment matricies: Applications in mixtures," *The Annals of Statistics*, vol. 17, no. 2, pp. 722–740, June 1989.

[25] D. M. Titterington, A.F.M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.

[26] Athanasios Papoulis and S. Unnikrishna Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, 4th edition, 2002.

[27] Carl N. Morris, "Natural exponential families with quadratic variance functions," *The Annals of Statistics*, vol. 10, no. 1, pp. 65–80, 1982.

[28] Vesa Välimäki, Bo Holm-Rasussen, Denoit Alary, and Heidi-Maria Lehtonen, "Late reverberation synthesis using filtered velvet noise," *Appl. Sci.*, vol. 7, 2017.

[29] V. Sundarapandian, *Probability, Statistics, and Queueing Theory*, PHI Learning Private, New Delhi, India, 2009.

[30] Julius O. Smith III, *Spectral Audio Signal Processing*, `ccrma.stanford.edu/~jos/sasp/`, 2011 edition, accessed April 2023.

[31] Marina Bosi, Karlheinz Brandenburg, Schuyler Quackenbush, Louis Fielder, Kenzo Akagiri, Hendrik Fuchs, Martin Dietz, Jügen Herre, Grant Davidson, and Yoshiaki Oikawa, "ISO/IEC MPEG-2 advanced audio coding," *J. of the Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, October 1997.

[32] Sebastian Kraft and Udo Zölzer, "BeaqleJS: HTML5 and JavaScript based framework for the subjective evaluation of audio quality," in *Linux Audio Conf. 2014*, Karlsruhe, Germany, May 2014.

[33] "R128 loudness normalisation and permitted maximum level of audio signals," Recommendation, European Broadcasting Union, Geneva, August 2020.

[34] "BS.1534: Method for the subjective assessment of intermediate quality level of audio systems," Recommendation, International Telecommunication Union, October 2015.

11. APPENDIX

We derive the fourth moment of the Poisson process proposed by [10] and discussed in Section 5.1. The n^{th} moment of the process described in (21) evaluates,

$$E\left[\alpha^n(t)\delta^n(t-\tau(t))\right]. \tag{23}$$

The Gaussian distributed amplitude and Poisson impulse are mutually independent, therefore their expectations are separable

$$E\left[\alpha^n(t)\right]E\left[\delta^n(t-\tau(t))\right]. \tag{24}$$

The n^{th} moment of the $\alpha(t)$ can be readily found based on the Gaussian distribution, and the first four moments are provided in Table 1. The Poisson impulse can be constructed by taking the time derivative of a Poisson process which is modeled as a discrete Poisson distributed variable $X \sim \mathrm{P}\left(k \mid \lambda = \rho/f_s \cdot t\right)$. The Poisson distribution represents the probability there are k events when λ is the mean number of events. From the linear property of the expectation and time derivative operators, one can verify that the mean of the Poisson impulse is ρ/f_s.

Derivation of the second moment of the Poisson impulse is given in [26]. Here, we present the derivation of the fourth moment. The fourth moment of a Poisson process for non-overlapping times $t_1 < t_2 < t_3 < t_4$ is related to the Poisson impulse by the partial derivatives $\frac{\partial}{\partial t_i}$ of the correlation operator R,

$$R_{\delta\delta\delta\delta}(t_1,t_2,t_3,t_4) = \frac{\partial^4 R_{xxxx}(t_1,t_2,t_3,t_4)}{\partial t_1 \partial t_2 \partial t_3 \partial t_4}. \tag{25}$$

The correlation of a randomly distributed variable is equivalent to its expectation,

$$R_{xxxx}(t_1,t_2,t_3,t_4) = E\left[X(t_1)X(t_2)X(t_3)X(t_4)\right], \tag{26}$$

and t_1, t_2, t_3, t_4 are non-overlapping. The expectation above can be re-expressed as,

$$E\left[X(t_1)X(t_2)X(t_3)X(t_4)\right] = \prod_{i=1}^{4} E\left[X(t_i) - X(t_{i-1})\right], \tag{27}$$

with $x(t_0) = 0$. This product can be expanded as the first moment for all $x(t_i)$ is $\rho/f_s t$

$$R_{xxxx}(t_1,t_2,t_3,t_4) = \frac{\rho^4}{f_s^4} \cdot t_1(t_2-t_1)(t_3-t_2)(t_4-t_3) \tag{28}$$

Applying the partial time derivatives in (25), the resulting fourth moment is

$$R_{\delta\delta\delta\delta} = \frac{\rho^4}{f_s^4} + \frac{\rho}{f_s}\left(\delta(t_1-t_2) + \delta(t_2-t_3) + \delta(t_3-t_4)\right) \tag{29}$$

where the $\delta(t)$ terms account for the discontinuities between non-overlapping time segments.