A CORPUS-DRIVEN PARAMETRIC MODAL REVERBERATOR

Michele Ducceschi
Department of Industrial Engineering
University of Bologna
Bologna, Italy
michele.ducceschi@unibo.it

Leonardo Gabrielli
Department of Information Engineering
Università Politecnica delle Marche
Ancona, Italy

Riccardo Simionato
SCRIME
University of Bordeaux
Bordeaux, France

Riccardo Russo
Department of Industrial Engineering
University of Bologna
Bologna, Italy

ABSTRACT

A parametric modal reverberator is presented in which synthesis parameters are derived from a large, curated corpus of room impulse responses (IRs). The collected responses are subjected to modal decomposition, yielding per-mode frequencies, damping coefficients, and residue amplitudes, together with a short early-reflection finite impulse response (FIR) filter. From the decomposed data, a feature table is constructed per IR comprising standard acoustic indices, per-band damping and density statistics, amplitude distributions, and FIR descriptors—50 variables in total. Six acoustically meaningful user controls are selected; since these exhibit substantial pairwise correlations across the corpus, they are orthogonalised via principal component analysis (PCA) prior to regression. Per-band damping and modal density are predicted by robust linear models in log space; residue amplitudes follow the diffuse-field equipartition relation, and early-reflection energy is set directly from the clarity index definition. The resulting method maps a six-number perceptual specification onto thousands of modal parameters driving a bank of second-order resonators, making modal reverberators easier to operate for use in audio engineering. Regression diagnostics and corpus-distribution analysis confirm that the generated impulse responses are acoustically plausible and span the parameter space of the training data, which, in turn, should be sufficiently representative of room IRs used in audio engineering and music production.

1. INTRODUCTION

Artificial reverberation is a basic tool in music production, sound design, and post-production. Convolution reverberators reproduce the character of a measured room with high fidelity, but offer little parametric flexibility. Algorithmic designs based on the Schroeder allpass networks [1] or feedback delay networks [2, 3]—including recent differentiable variants [4, 5, 6]—are efficient and tunable, but the mapping between internal parameters (delay lengths, feedback coefficients, attenuation filters) and standard acoustic descriptors (T_{60} profile, modal density, spectral balance) is indirect, complicating calibration from measured data. Wave-based solvers achieve high physical accuracy [7, 8] but remain too costly for interactive use.

Modal synthesis offers a middle path: the impulse response is represented as a sum of damped sinusoids whose parameters—frequency, damping, amplitude—carry direct acoustic meaning [9]. Earlier modal reverberators have synthesised measured rooms by direct parameter fitting [10, 11] or sub-band ESPRIT analysis with FIR augmentation for early reflections [12]. The difficulty lies in mapping a small set of high-level controls onto potentially thousands of modal parameters in a manner that is both physically grounded and musically useful. Here, this mapping is constructed entirely from data: a large corpus of measured room IRs is decomposed into modal form, aggregated into per-band statistics, and regressed against six user-facing controls; the contribution is the corpus-driven mapping from a small set of perceptual controls to thousands of modal parameters, rather than the re-synthesis of any specific measurement.

The paper is organised as follows: Section 2 develops the modal representation; Section 3 describes data collection, curation, and feature extraction; Section 4 presents parameter selection, PCA, and the parametric model; Section 5 provides results; Section 6 concludes the paper.

2. MODAL REPRESENTATION

The acoustic pressure field in a room with rigid or absorbing boundaries is governed by the three-dimensional wave equation. Under the assumption of linearity and time invariance (LTI), the source-to-receiver transfer function admits a decomposition into a countable set of resonant modes, each corresponding to a complex pole of the system. In such a decomposition, the frequency response of a source-receiver pair is amenable to the following pole-residue form:

$$H(j\omega) = \sum_{k=1}^{M} \left(\frac{r_k}{j\omega - \lambda_k} + \frac{r_k^*}{j\omega - \lambda_k^*} \right), \tag{1}$$

where M, the number of modes, is truncated to a large but finite integer. In the above, $\lambda_k = -\sigma_k + j\omega_{d,k}$ is the complex pole of the k-th mode, $\sigma_k > 0$ is the damping coefficient, $\omega_{d,k} = \sqrt{\Omega_k^2 - \sigma_k^2}$ is the damped angular frequency ($\Omega_k = 2\pi f_k$ the undamped counterpart), and $r_k = a_k + jb_k$ is the complex residue. In practice, the residues and poles may be obtained analytically for a small number of canonical geometries [13]; numerically, such as via finite difference or finite volume discretisation of the governing equations [7, 8]; or experimentally, by fitting a pole–residue model to a measured impulse response [14]. In all cases, the modal parameters encode the physics of the room: frequencies f_k reflect the geometry and boundary conditions; damping coefficients σ_k encode losses due to wall absorption, air viscosity, and coupling to adjacent spaces; and residue amplitudes r_k depend on the mode shapes evaluated at the source and receiver positions. A modal reverberator exploits this structure directly, yielding a formulation of the time-domain impulse response as

$$h(t) = \sum_{k=1}^{M} 2\, e^{-\sigma_k t}\big(a_k \cos \omega_{d,k} t - b_k \sin \omega_{d,k} t\big). \tag{2}$$

In a physical simulation, the quantities $(f_k, \sigma_k, a_k, b_k)$ would be obtained by solving the Helmholtz equation for a given geometry and boundary conditions. In this work, crucially, modal parameters are derived from a handful of perceptually meaningful parameters which, in turn, are mapped *statistically* to distributions learned from a corpus of measured impulse responses.

3. DATA COLLECTION AND CURATION

Approximately 1 500 room impulse responses were assembled from bundled convolution-reverb presets (Apple Logic Pro, Ableton Live), factory content from MeldaProduction plugins, the OpenAIR library [15], and previously measured IRs available to the authors. The combined collection covers a wide range of room types, sizes, and acoustic treatments. Roughly half are measured-room IRs (OpenAIR and authors' measurements); the rest come from commercial libraries, which are themselves mostly measured spaces processed for production use. All IRs were sampled or resampled to 44.1 kHz.

3.1. Modal decomposition, resynthesis and quality screening

Each IR was decomposed into the pole–residue form (1) following a modification of the sub-band procedure described in [14]. No preprocessing—equalisation, normalisation, or time alignment—was applied prior to decomposition. Resonant modes in the 40–12 000 Hz range were identified by peak-picking; their decay rates were estimated using a combination of the -3 dB bandwidth rule (damping $\sigma_k = \pi\Delta f_{-3\mathrm{dB}}$, with $\Delta f_{-3\mathrm{dB}}$ the width of the resonance peak at -3 dB below its maximum) and sub-band Schroeder integration. For each stable pole, the corresponding residue was obtained by linear least-squares fitting against the measured frequency response [14]. The output was a set of M complex pole–residue pairs per IR, from which the four real scalars $(f_k, \sigma_k, a_k, b_k)$ were extracted. A 256-tap FIR correction filter h_{FIR} was then fitted by time-domain least squares to the residual between the original IR and the modal sum, capturing the direct sound and early reflections not resolved by the modes; 256 taps (≈ 5.8 ms at 44.1 kHz) covers the direct sound and first reflections, leaving the bulk of the early-reflection cluster to the dense modal tail.

Every IR was resynthesised from its extracted parameters via a discrete-time implementation of (2), that is, using $t := nT$ for time index $n \in \mathbb{N}$ and sampling period T. The resulting modal sum was then convolved with the corresponding h_{FIR} to restore the early reflections not resolved by the modes. Each synthesised impulse response was compared to the original using the normalised root-mean-square error (nRMSE). Decompositions exceeding a prescribed error threshold were discarded. Most rejections arose from IRs that had undergone heavy nonlinear processing—saturation, fast-attack compression, or spectral shaping—violating the LTI assumption underlying the modal model. Impulse responses of springs and plates were removed manually because they fall outside the target domain of room impulse response and, in many cases, exhibit non-room-like resonance behaviour and nonlinear artefacts incompatible with the assumed LTI room model.. After screening, $N = 1\,151$ IRs were retained.

Table 1: Feature groups extracted per IR. All quantities were computed on the resynthesised signal. Five frequency bands were used throughout: sub-bass (40–125 Hz), bass (125–500 Hz), mid (500–2 kHz), presence (2–8 kHz), brilliance (8–12 kHz). The dataset contains 50 features per IR.

Group	Features
Acoustic indices	$T_{30,\mathrm{mid}}$, T_s, BR, TR, $C_{80,\mathrm{mid}}$; per-octave T_{30} and C_{80} at 125–4 kHz
Damping	Per-band mean σ_b; global mean, median, std
Spacing & density	Mode count M; CV_Δ; per-band density ρ_b
Amplitude	Per-band mean $\lvert r_k \rvert$; global mean, median, std; spectral slope
FIR	Energy, peak value, temporal centroid

3.2. Feature extraction

From each retained IR, a feature vector of 50 scalar variables was computed from the *resynthesised* signal—not the original recording—so that all acoustic descriptors refer to the same modal representation that produced the underlying parameters. This collection of feature vectors, one per IR, constitutes the *feature set*. Table 1 lists the five feature groups, now described.

Acoustic indices. Standard room-acoustic parameters were computed from the resynthesised impulse response following ISO 3382-1 [16]. Reverberation times were estimated from the energy decay curve [13] evaluated by backward integration of $h^2(t)$ (Schroeder integration). The slope of the resulting curve between -5 and -35 dB was extrapolated to -60 dB, yielding the T_{30} estimate of the 60 dB decay time per octave band. The mid-frequency value

$$T_{30,\mathrm{mid}} = \tfrac{1}{2}\big(T_{30,500} + T_{30,1\mathrm{k}}\big) \tag{3}$$

served as the primary decay descriptor. Centre time

$$T_s = T\,\frac{\sum_n n\, h^2(nT)}{\sum_n h^2(nT)} \tag{4}$$

quantified the temporal centre of gravity of the impulse response energy; it tends to increase with room volume and reverberation time [13]. Clarity

$$C_{80} = 10 \log_{10} \frac{\sum_{n=0}^{n_{80}} h^2(nT)}{\sum_{n>n_{80}} h^2(nT)} \tag{5}$$

(with $n_{80} = \lfloor 0.08/T \rceil$) measured the ratio of early to late energy; the mid-frequency average $C_{80,\mathrm{mid}} = \frac{1}{2}(C_{80,500} + C_{80,1\mathrm{k}})$ was retained. Bass and treble ratios

$$\mathrm{BR} = \frac{T_{125} + T_{250}}{T_{500} + T_{1\mathrm{k}}}, \qquad \mathrm{TR} = \frac{T_{2\mathrm{k}} + T_{4\mathrm{k}}}{T_{500} + T_{1\mathrm{k}}} \tag{6}$$

characterised spectral tilt.

Damping statistics. The per-mode damping coefficients σ_k extracted during decomposition were grouped into the five frequency bands and averaged, yielding per-band mean damping σ_b, $b = 1, \ldots, 5$. Global summary statistics (mean, median, and standard deviation across all modes) were also recorded. These per-band means serve as the regression targets for the damping model described in Section 4.1.

Modal spacing and density. The mode frequencies f_k were sorted and differenced to obtain the inter-mode spacings Δf. The coefficient of variation $\mathrm{CV}_\Delta = \mathrm{std}(\Delta f) \,/\, \mathrm{mean}(\Delta f)$ quantifies spacing regularity: values near zero indicate a periodic grid; values near unity indicate the quasi-random distribution characteristic of real rooms. Per-band modal density ρ_b (modes per Hz) was computed as the number of modes falling within each band divided by the bandwidth. The total mode count M was retained as a global descriptor.

Amplitude statistics. The residue magnitude $|r_k|$ was computed for each retained mode and grouped into the five frequency bands, yielding per-band means together with global mean, median, and standard deviation across all modes. A global spectral slope was additionally computed as the OLS slope of $\log |r_k|$ on $\log |f_k|$ across all modes.

FIR statistics. Three scalar descriptors were extracted from the early-reflection correction filter h_{FIR}: total energy $\sum h_{\mathrm{FIR}}^2$, peak absolute value, and temporal centroid (computed as in (4), applied to h_{FIR}).

4. PARAMETRIC MODEL

The dataset of Table 1 contains 50 features per IR. A compact parametric description of reverberation requires selecting a small number of controls from this set. PCA applied to the full feature matrix was investigated as a data-driven reduction strategy; however, this required ≈ 18 components to explain 95% of the variance, with resulting axes—linear combinations of all 50 variables—bearing no direct connection to physically or perceptually meaningful quantities, and was therefore rejected.

A different strategy was adopted. Rather than letting PCA choose the axes, six controls were selected on physical and practical grounds: five of the six correspond to standard room-acoustic quantities defined in ISO 3382-1 [16]; the sixth, diffusion (CV_Δ), is a modal spacing statistic with no ISO counterpart, but a direct perceptual effect on the character of the reverberant tail. The six controls map naturally to the controls found on commercial parametric reverberators (decay time, room size, spectral tilt, diffusion, early/late balance), and connect directly to at least one of the underlying modal synthesis targets—damping, density, amplitude, or FIR energy. The six controls retained are given in Table 2, and are now described:

- *T60* sets the mid-frequency reverberation time via $T_{30,\mathrm{mid}}$ (3).
- *warmth* and *brightness* map to the bass and treble ratios BR and TR (6), giving direct timbral control as a fraction of the decay time in the lower and upper bands compared to the mid band.

Table 2: User controls, corpus range, and dataset mapping. Controls whose physical units are familiar to audio engineers—decay time in seconds, bass and treble ratios—are exposed directly; the remaining three, whose underlying quantities (T_s, CV_Δ, $C_{80,\mathrm{mid}}$) are less intuitive, are mapped linearly to a $[0,\ 1]$ range spanning the corpus $[p_5,\ p_{95}]$ interval.

Control	Range	Dataset param.	Meaning
T60	0.1–6 s	$T_{30,\mathrm{mid}}$	Decay time
warmth	0.5–2.0	BR	Low-freq tilt
brightness	0.3–2.0	TR	High-freq tilt
roomSize	0–1	T_s	Volume proxy
diffusion	0–1	CV_Δ	Spacing regularity
earlyLate	0–1	$C_{80,\mathrm{mid}}$	Early/late ratio

- *roomSize* maps to centre time T_s (4), encoding temporal energy build-up and proxying room volume.
- *diffusion* maps to CV_Δ, the coefficient of variation of the inter-mode frequency spacing: low values produce a regular, metallic-sounding grid; high values reproduce the quasi-random spacing of real rooms.
- *earlyLate* maps to $C_{80,\mathrm{mid}}$ (5), controlling the balance between early reflections and the reverberant tail.

Figure 1 shows the corpus distributions of the six selected parameters. T_{30} is right-skewed, with 50% of the corpus below 1 s and 95% below 4 s. T_s is concentrated below 0.15 s, confirming that most rooms in the collection are relatively compact. BR and TR cluster near unity, indicating spectrally balanced spaces; the long upper tail of BR arises from a small number of heavily bass-loaded environments. CV_Δ peaks sharply near unity, meaning that quasi-random modal spacing—not a regular grid—is the norm. $C_{80,\mathrm{mid}}$ spans roughly -7 to $+17$ dB, with median near 3 dB. Values in the interquartile range of each parameter are well supported by the data; settings outside the p_5–p_{95} envelope extrapolate beyond the training distribution and should be treated as creative extremes.

Before describing how the modal synthesis targets—damping, density, amplitude, and FIR energy—are predicted from these six controls, their statistical independence must be assessed. Although the six parameters are conceptually distinct, they are not statistically independent in the corpus (Fig. 2, left). Hierarchical clustering of the full 50-feature correlation matrix (not shown) grouped the variables into three coarse blocks linked to decay/density/size, amplitude, and damping, and within the six-parameter subset, the strongest pairwise correlations were T_{30}–T_s ($r = 0.83$) and T_s–C_{80} ($r = -0.64$). Correlated predictors inflate the coefficient variance and destabilise least-squares estimates; direct regression on the raw six-parameter vector was found to yield unstable and poorly interpretable coefficients.

The six parameters were therefore robustly standardised as $z_j = (p_j - \tilde{p}_j) \,/\, (\mathrm{IQR}_j/1.349)$, where $\tilde{p}_j$ is the median and $\mathrm{IQR}_j = Q_3 - Q_1$ the interquartile range. The factor 1.349 is the standard outlier-insensitive rescaling that recovers the standard deviation under a Gaussian assumption [17]; for the non-Gaussian parameters in Fig. 1 it is retained purely as a robust scale estimate. The standardised parameters were then rotated into an orthogonal basis by PCA: the eigenvectors of the 6×6 covariance matrix define a rotation $\mathbf{V}$ under which the transformed scores are pairwise uncorrelated by construction. The 6×6 rotation matrix $\mathbf{V}$ was

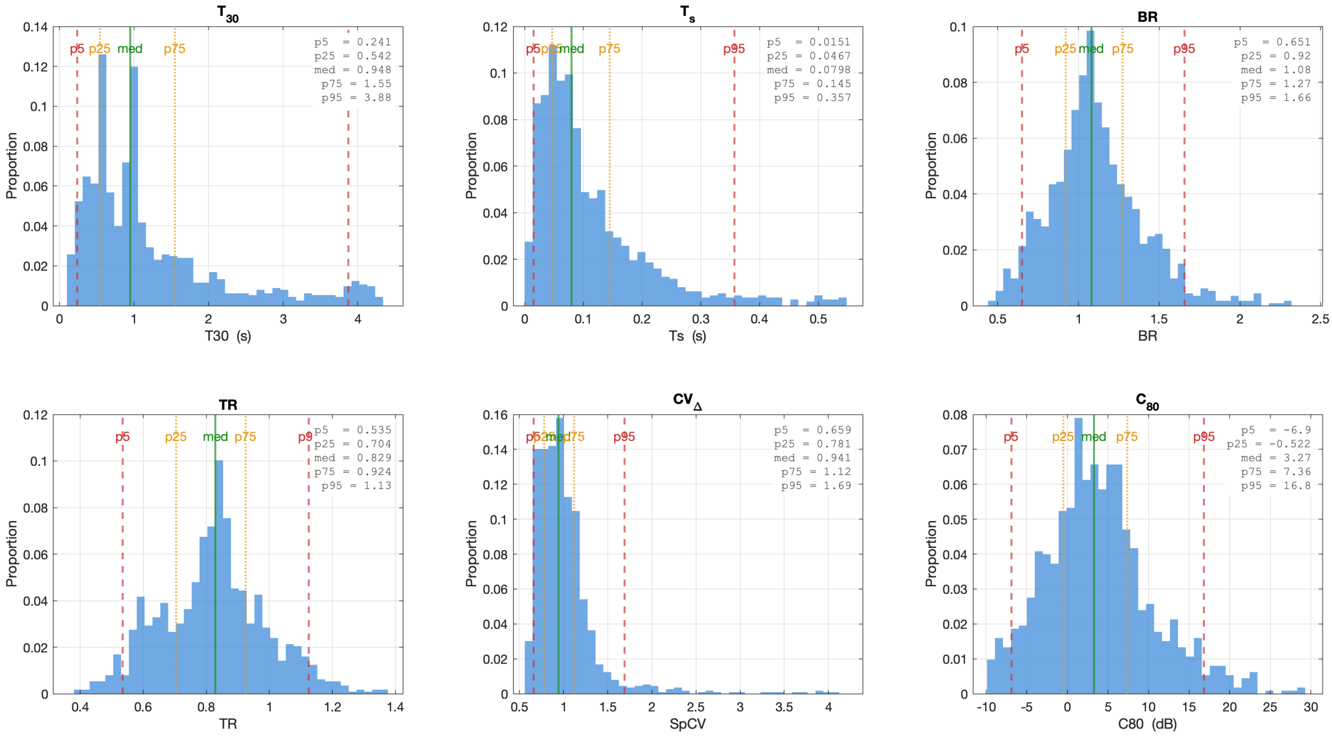

Figure 1: Corpus distributions of the six selected parameters ($N = 1\,151$ IRs, clipped to the 1st–99th percentile range). Vertical lines mark p_5 and p_{95} (red dashed), p_{25} and p_{75} (yellow dotted), and the median (green solid); percentile values are listed per panel.

stored in the model. The first three components accounted for $\approx 95\%$ of variance (Fig. 2, centre); all six were retained so that each control keeps a unique, invertible direction in PC space. Off-diagonal correlations of the PC scores fell below 7×10^{-16} (Fig. 2, right).

For any setting of the six user parameters, collected in the vector $\mathbf{p} = [T_{30},\, T_s,\, \mathrm{BR},\, \mathrm{TR},\, \mathrm{CV}_\Delta,\, C_{80}]^\top$, the corresponding PC scores are obtained by standardising and rotating:

$$[\mathrm{PC}_1, \ldots, \mathrm{PC}_6]^\top = \mathbf{V}^\top \mathbf{D}^{-1}(\mathbf{p} - \tilde{\mathbf{p}})\,, \tag{7}$$

where $\tilde{\mathbf{p}}$ is the vector of corpus medians and $\mathbf{D} = \mathrm{diag}(\mathrm{IQR}_1/1.349, \ldots, \mathrm{IQR}_6/1.349)$. The design vector used in all subsequent regressions is then $\mathbf{x} = [1, \mathrm{PC}_1, \ldots, \mathrm{PC}_6]^\top \in \mathbb{R}^7$, where the leading 1 accommodates the intercept term β_0 of the regressions below.

4.1. Modal parameter generation

Given the six user controls mapped to physical parameters and projected onto the orthogonal basis $\mathbf{x}$ as described above, the modal synthesis targets—per-band damping σ_b and per-band density ρ_b—were predicted by linear models operating in log space.

Damping. In a room, damping arises from surface absorption and air dissipation [13]; both mechanisms are frequency-dependent, so the spectral shape of $\sigma(f)$ varies across rooms. The per-band mean damping σ_b extracted in Section 3.2 captures this shape at five points across the spectrum. To predict σ_b from the six user controls, a multivariate linear model was fitted independently per band. For band b, the model takes the form

$$\log \sigma_b^{(i)} = \beta_0 + \beta_1\, \mathrm{PC}_1^{(i)} + \cdots + \beta_6\, \mathrm{PC}_6^{(i)} + \varepsilon^{(i)}\,, \tag{8}$$

where $i = 1, \ldots, N$ indexes IRs in the corpus and $\varepsilon^{(i)}$ is the residual error. The seven coefficients $\boldsymbol{\beta}_{\sigma,b} = [\beta_0, \ldots, \beta_6]^\top$ were determined by iteratively reweighted least squares (IRLS) with bisquare weights, which down-weight outlying IRs whose absorption profiles deviate strongly from the bulk of the corpus. Given a new parameter setting encoded in the design vector $\mathbf{x}$, the predicted damping—denoted $\hat{\sigma}_b$, where the hat distinguishes the model output from the corpus observations—follows as

$$\log \hat{\sigma}_b = \mathbf{x}^\top \boldsymbol{\beta}_{\sigma,b}\,, \qquad b = 1, \ldots, 5\,. \tag{9}$$

The log transform guarantees $\hat{\sigma}_b > 0$ at all settings.

At synthesis time, the five predicted values are evaluated at the geometric band centres $f_b^c = \sqrt{f_b^{\mathrm{lo}}\, f_b^{\mathrm{hi}}}$ and interpolated to per-mode frequencies by linear interpolation in $(\log_{10} f,\ \log \sigma)$ space, producing a continuous damping profile that is strictly positive everywhere. The profile is then rescaled so that its mean over 500–1000 Hz matches the target decay:

$$\hat{\sigma}(f_k) \leftarrow \hat{\sigma}(f_k) \cdot \frac{3 \ln 10 \,/\, T_{60}}{\bar{\hat{\sigma}}_{500\text{–}1\mathrm{k}}}\,, \tag{10}$$

where $3 \ln 10 \approx 6.908$ is the factor relating T_{60} to the exponential decay constant [13]. A log-linear taper then divides $\hat{\sigma}$ by BR below 500 Hz and by TR above 1 kHz, implementing the timbral controls.

Density. Modal density in a room grows with frequency and with the room volume [13]; it is closely linked to the reverberation time and to the temporal spread of energy, both of which are among the six selected controls. Per-band density ρ_b was modelled by five independent regressions identical in structure to (8)–(9), with $\log \rho_b$ as the response variable:

$$\log \hat{\rho}_b = \mathbf{x}^\top \boldsymbol{\beta}_{\rho,b}\,, \qquad b = 1, \ldots, 5\,, \tag{11}$$

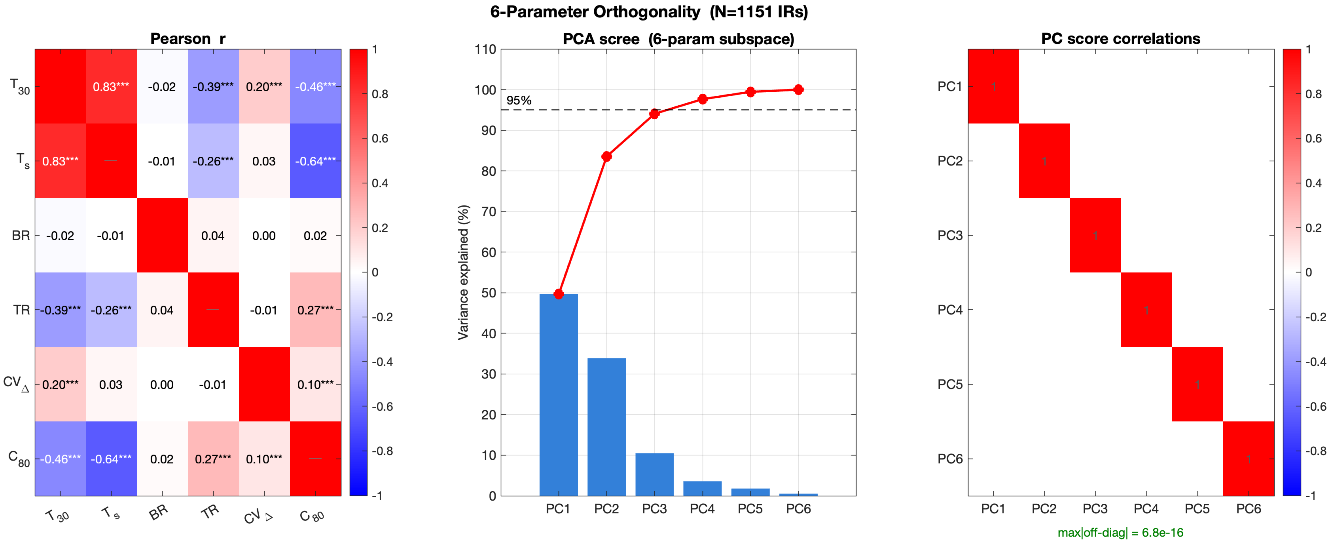

Figure 2: Orthogonality diagnostics ($N = 1\,151$). Left: Pearson r of the six raw parameters. Centre: PCA scree. Right: PC score correlation matrix (all off-diagonal entries at machine-precision zero).

where $\boldsymbol{\beta}_{\rho,b} \in \mathbb{R}^7$ is the coefficient vector fitted by IRLS for band b. As with damping, the prediction is exponentiated before use, guaranteeing $\hat{\rho}_b > 0$ at all settings.

The predicted densities determine the number of modes to be synthesised. The raw total count follows from the bandwidth Δf_b of each band:

$$M_{\text{raw}} = \sum_{b=1}^{5} \hat{\rho}_b \, \Delta f_b \,. \tag{12}$$

To keep the computational cost bounded while preserving perceptual density, M_{raw} is remapped linearly to the range $[M_{\min}, M_{\max}] = [1000, 8000]$, calibrated against the corpus distribution of mode counts. The resulting integer M is allocated across bands in proportion to $\hat{\rho}_b \, \Delta f_b$, with remainders distributed by largest-remainder rounding so that $\sum_b M_b = M$.

Within each band, M_b modes are placed on a uniform frequency grid and perturbed according to the diffusion control d:

$$f_k \leftarrow \text{clip}_{[f_l,\, f_h]}\Big(f_k + \xi_k \cdot d \cdot 0.9\,\overline{\Delta f}_b\Big), \tag{13}$$

where ξ_k is drawn independently from a uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$, $\overline{\Delta f}_b$ is the mean spacing in band b, and the factor 0.9 prevents adjacent modes from coinciding at full jitter. At $d = 0$ the grid is perfectly regular, producing a metallic, comb-like response; at $d = 1$ each mode is displaced by up to $\pm\, 0.45\,\overline{\Delta f}_b$, reproducing the quasi-random spacing observed in real rooms.

Amplitude. The residue amplitudes $|r_k| = \sqrt{a_k^2 + b_k^2}$ depend on the mode shapes evaluated at the source and receiver positions, and vary erratically from mode to mode. No regression target with a useful predictive structure could be identified from the corpus data.

The amplitude model rests instead on an energy argument. The time-integrated energy of the k-th mode is

$$E_k = \frac{|r_k|^2}{2\sigma_k} \,. \tag{14}$$

In a diffuse reverberant field, each mode is expected to carry approximately equal energy [18, 13]—the acoustic counterpart of the equipartition theorem. The condition $E_k = \text{const}$ requires $|r_k| \propto \sqrt{\sigma_k}$. The ratio $|r_k|/\sqrt{\sigma_k}$, computed across all modes in the corpus, concentrates around a median

$$\kappa = \text{median}\left\{ \frac{|r_i|}{\sqrt{\sigma_i}} \right\} \approx 4.86 \,, \tag{15}$$

with standard deviation ≈ 7.07. Amplitudes are accordingly set as $|r_k| = \kappa\sqrt{\hat{\sigma}(f_k)}$, where the corpus constant $\kappa \approx 4.86$ fixes an overall gain that is absorbed by output normalisation. The residue components are obtained from a random phase $\phi_k \sim \mathcal{U}(0, 2\pi)$:

$$a_k = |r_k| \cos\phi_k \,, \qquad b_k = |r_k| \sin\phi_k \,. \tag{16}$$

Early reflections. The direct sound and early reflections are not well represented by the modal sum (2), which describes the reverberant tail. A short FIR correction filter h_{FIR} was introduced in Section 3 to capture this portion of the response. At synthesis time, its *shape* is fixed to the corpus-mean template $\bar{h}_{\text{FIR}}$, averaged across all retained IRs; only its energy is controlled. The energy balance between early and late portions of the IR are governed by the clarity index C_{80} (5). Rearranging the definition yields

$$E_{\text{early}} = E_{\text{late}} \cdot 10^{C_{80}/10} \,, \tag{17}$$

where the late-field energy is computed directly from the modal parameters as

$$E_{\text{late}} = \sum_{k=1}^{M} \frac{|r_k|^2}{2\hat{\sigma}_k} \,. \tag{18}$$

The template is then scaled to match:

$$h_{\text{FIR}} = \bar{h}_{\text{FIR}} \sqrt{\frac{E_{\text{early}}}{\|\bar{h}_{\text{FIR}}\|^2}} \,. \tag{19}$$

The user control earlyLate maps linearly to $C_{80,\text{mid}}$ across the corpus $[p_5,\, p_{95}]$ range: earlyLate $= 0$ selects the most diffuse, tail-dominated responses in the corpus; earlyLate $= 1$ selects the driest, with prominent early energy.

4.2. Synthesis pipeline

The complete signal path from user input to impulse response proceeds as follows. The user specifies six controls: T60, roomSize, warmth, brightness, diffusion, and earlyLate. These are mapped to the physical parameter vector $\mathbf{p}$ according to Table 2: T60 sets T_{30} directly; roomSize interpolates T_s linearly across the corpus $[p_5, p_{95}]$ range; warmth and brightness set BR and TR; diffusion and earlyLate map linearly to CV_Δ and $C_{80,\mathrm{mid}}$ across their respective corpus ranges.

The vector $\mathbf{p}$ is standardised and projected onto the PCA basis via (7), yielding the design vector $\mathbf{x} \in \mathbb{R}^7$. From $\mathbf{x}$, per-band damping $\hat{\sigma}_b$ and density $\hat{\rho}_b$ are obtained from (9) and (11) respectively. The density prediction determines the mode count M via (12) and its allocation across bands; mode frequencies are placed on a jittered grid (13) controlled by the diffusion parameter. The five band-centre damping values are interpolated to per-mode frequencies and anchored to T_{60} (10), with timbral shaping applied via BR and TR. Residue amplitudes follow from equipartition (15)–(16), and FIR energy is set by the C_{80} identity (17)–(19).

The resulting modal parameters $(f_k, \sigma_k, a_k, b_k)_{k=1}^{M}$ and FIR filter h_{FIR} fully specify the impulse response, which can be, thus, synthesised.

5. RESULTS

Table 3 and Fig. 3 summarise the quality of the ten per-band regressions. For each band, ordinary least squares (OLS) R^2, the robust coefficient of determination R^2_{rob} computed on the inliers retained by the IRLS bisquare weights, and the outlier fraction n_{out} are reported. Values of $R^2_{\mathrm{rob}} \gtrsim 0.85$ indicate that the IRLS predictions explain the bulk of the variance in the corpus on the inlier set, with the OLS column quantifying how much robustness contributes over a non-robust baseline. The outlier fraction n_{out} indicates the proportion of IRs assigned negligible bisquare weights by IRLS, i.e., observations whose residuals are large enough to be treated as outliers in the robust fit; values of only 1–4% show that the regressions are driven by a small minority of atypical responses rather than broad model misfit.

The damping fits were tight across all bands, with R^2_{rob} ranging from 0.747 (brilliance) to 0.937 (bass) and inlier fractions exceeding 93%. The density fits were looser at the spectral extremes—sub-bass and brilliance reached $R^2_{\mathrm{rob}} \approx 0.5$—where the corpus contains fewer modes and greater variability. All ten residual distributions exhibited strong leptokurtosis, confirming the appropriateness of the bisquare IRLS estimator.

5.1. Synthesis examples

Figure 4 shows three synthesised IRs spanning a 7.5:1 range of reverberation times: Small Dry ($T_{60} = 0.4$ s, roomSize $= 0.2$ giving $T_s \approx 0.04$ s, earlyLate $= 0.80$), Medium Room (1.2 s, roomSize $= 0.5$, $T_s \approx 0.08$ s, 0.45), and Large Hall (3.0 s, roomSize $= 0.85$, $T_s \approx 0.18$ s, 0.20). Top-row waveforms confirm the progressive lengthening of the decay envelope. Middle-row spectrograms display the frequency-dependent decay controlled by BR and TR: the Medium Room preset (BR $= 1.20$) shows extended low-frequency sustain, while the Large Hall (TR $= 0.80$) exhibits the high-frequency roll-off typical of large absorptive spaces. Bottom-row panels plot per-mode T_{60} (blue, left axis, computed as $6.908/\sigma_k$) and residue amplitude $|r_k|$ (red, right axis) versus log frequency. The equipartition relation $|r_k| \propto \sqrt{\sigma_k}$ is visible as a tracking between the two curves: modes with shorter T_{60} (higher damping) carry proportionally higher amplitudes. The earlyLate control decreases with room size, as expected: in smaller rooms, the direct sound and early reflections dominate, while in large halls, the reverberant tail carries the bulk of the energy.

The generated IRs were compared to the corpus by projecting both into the six-dimensional parameter space, standardised by subtracting the mean and dividing by the standard deviation per axis (Fig. 5). The three synthesis presets (red diamonds) fall within the corpus cloud. Two additional IRs were generated at the extreme corners of the parameter ranges in Table 2 (blue diamonds): these lie well outside the corpus cluster, illustrating the boundary of the trained region. Source code and audio examples are available online.[1]

Table 3: Regression quality for the ten per-band models ($N = 1\,151$ IRs). OLS: ordinary least squares. R^2_{rob}: robust R^2 on IRLS inliers. n_{out}: fraction rejected by bisquare weights. Three additional components require no regression: the T_{60} anchor (10) rescales mid-band damping ($R^2_{\mathrm{rob}} = 0.945$); FIR energy is set from the C_{80} definition (17); and the equipartition constant $\kappa \approx 4.86$ is a corpus median.

Target	OLS	R^2_{rob}	n_{out}
Damping			
$\log\sigma$: sub-bass	0.735	0.881	4%
$\log\sigma$: bass	0.872	0.937	4%
$\log\sigma$: mid	0.861	0.920	3%
$\log\sigma$: presence	0.782	0.878	3%
$\log\sigma$: brilliance	0.579	0.747	5%
Density			
$\log\rho$: sub-bass	0.403	0.621	4%
$\log\rho$: bass	0.583	0.717	4%
$\log\rho$: mid	0.611	0.700	3%
$\log\rho$: presence	0.560	0.645	3%
$\log\rho$: brilliance	0.408	0.497	1%

6. CONCLUDING REMARKS

A data-driven parametric modal reverberator has been presented. Starting from approximately 1 500 measured room IRs, a quality-screened corpus of $N = 1\,151$ IRs was assembled through modal decomposition and nRMSE filtering. A 50-feature table comprising acoustic indices, per-band damping and density statistics, amplitude descriptors, and FIR characteristics was extracted from the resynthesised signals. Six user controls were selected on the basis of their correspondence to ISO 3382-1 quantities, their familiarity in commercial reverberator interfaces, and their direct connection to the modal synthesis targets. Pairwise correlations among the six controls (Pearson $|r|$ up to 0.83) were resolved by PCA before regression.

The model contains ten per-band regressions—five for damping, five for density—fitted by robust IRLS. Per-band damping achieved R^2_{rob} from 0.75 to 0.94; density from 0.50 to 0.72. No parametric assumption on the shape of the damping-versus-frequency curve

[1] https://github.com/Nemus-Project/DAFx26_ModalReverb

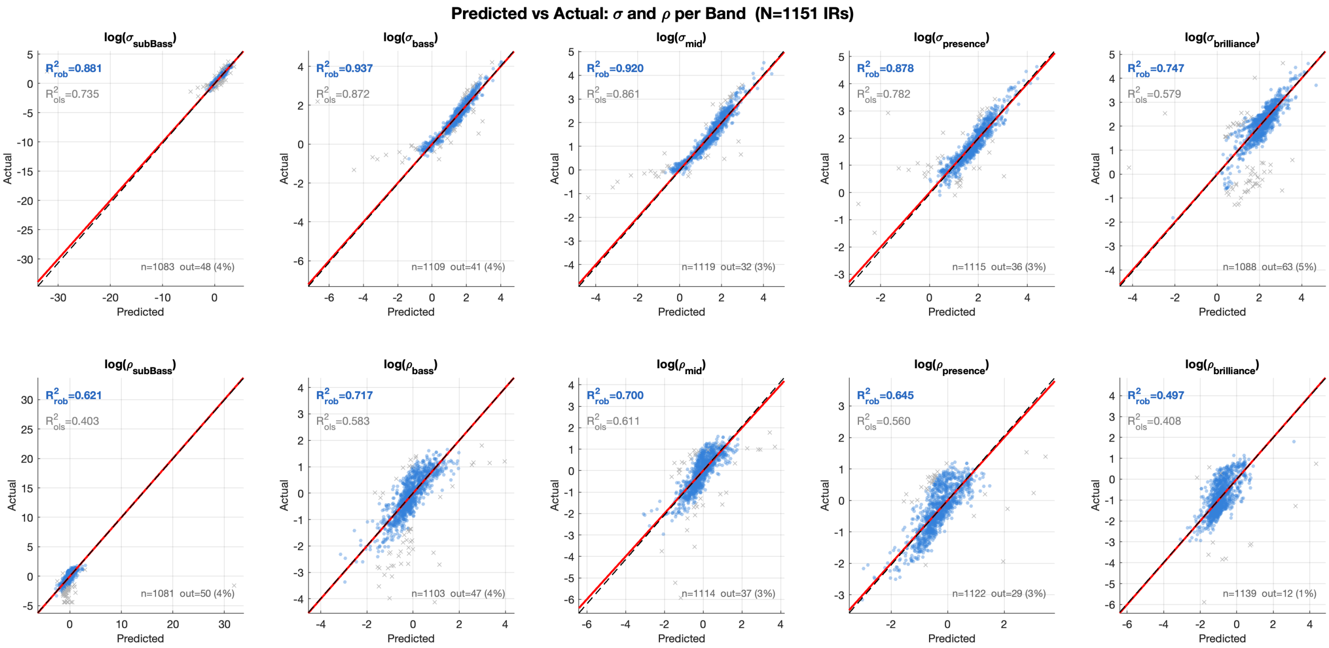

Figure 3: Predicted-vs-actual scatter for the ten per-band regressions ($N = 1\,151$ IRs). Top row: $\log \hat{\sigma}_b$; bottom row: $\log \hat{\rho}_b$. Columns correspond to the five frequency bands (sub-bass to brilliance). Blue: IRLS inliers; grey crosses: rejected outliers; red line: identity; dashed: OLS fit.

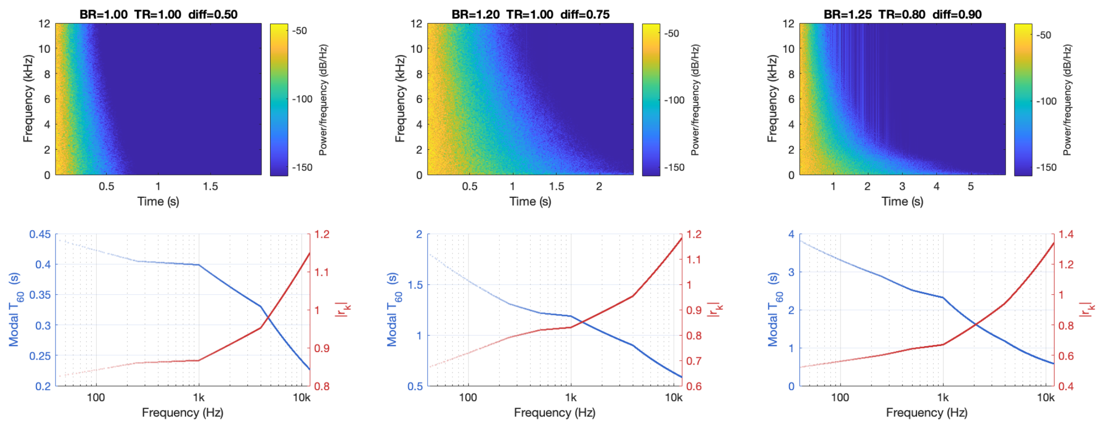

Figure 4: Synthesised IRs for three environments. Top: waveforms. Middle: spectrograms (BR, TR, diffusion annotated). Bottom: per-mode $T_{60} = 6.908/\sigma_k$ (blue, left axis) and $|r_k|$ (red, right axis) vs. log frequency. From left: Small Dry ($T_{60} = 0.4$ s, earlyLate $= 0.80$); Medium Room (1.2 s, 0.45); Large Hall (3.0 s, 0.20).

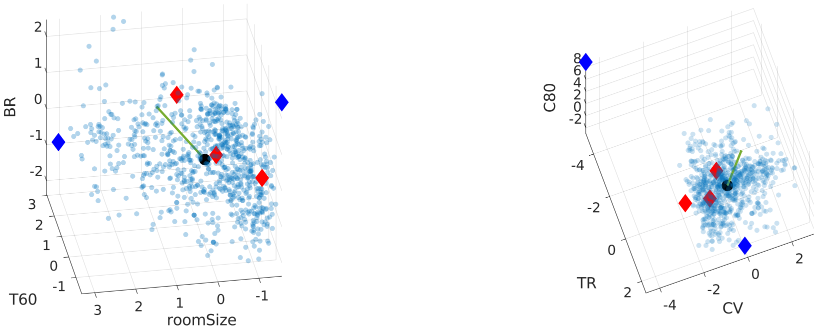

Figure 5: Corpus IRs (blue dots) and synthesised IRs projected into the standardised six-parameter space. Red diamonds: the three presets of Fig. 4; blue diamonds: IRs generated at the extreme ranges of Table 2. Green vector: unit standard deviation.

was imposed: the five band values are interpolated to per-mode frequencies in log–log space, guaranteeing positivity. Residue amplitudes were set via the diffuse-field equipartition relation using a single corpus constant, and early-reflection energy was controlled exactly through the C_{80} definition.

Overall, this work enhances the usability of modal reverberators, allowing the generation of diverse IRs from six parameters familiar to audio engineers, while retaining the qualities of a corpus deemed representative of room IRs used in music production. The current MATLAB prototype generates parameters in well under a second per IR on a standard laptop and runs comfortably in real time at 44.1 kHz inside a host convolver.

Future work includes a formal perceptual evaluation comparing the modal synthesiser against FDN-based methods [5, 19, 20], non-linear regression for bands where R^2 remains moderate, the use of artificial neural networks to learn a direct mapping from physical or perceptual parameters to modal frequencies, dampings, and gains, and extension to multi-channel configurations.

7. ACKNOWLEDGMENTS

This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant agreement No. 950084 NEMUS.

8. REFERENCES

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

[2] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Proc. 90th Conv. Audio Eng. Soc.*, Paris, France, Feb. 1991, paper 3030.

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

[4] G. Dal Santo, K. Prawda, S. J. Schlecht, and V. Välimäki, "Differentiable feedback delay network for colorless reverberation," in *Proc. Int. Conf. Digital Audio Effects (DAFx23)*, Copenhagen, Denmark, Sept. 4–7, 2023, pp. 244–251.

[5] G. Dal Santo, B. Alary, K. Prawda, S. J. Schlecht, and V. Välimäki, "RIR2FDN: An improved room impulse response analysis and synthesis," in *Proc. Int. Conf. Digital Audio Effects (DAFx24)*, Guildford, UK, Sept. 3–7, 2024, pp. 230–237.

[6] A. I. Mezza, R. Giampiccolo, E. De Sena, and A. Bernardini, "Data-driven room acoustic modeling via differentiable feedback delay networks with learnable delay lines," *EURASIP J. Audio, Speech, Music Process.*, vol. 2024, no. 1, p. 51, 2024.

[7] B. Hamilton and S. Bilbao, "FDTD methods for 3-D room acoustics simulation with high-order accuracy in space and time," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 11, pp. 2112–2124, Nov. 2017.

[8] S. Bilbao, "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 7, pp. 1524–1533, Jul. 2013.

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

[10] J. S. Abel, S. Coffin, and K. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Audio Engineering Society Convention 137*, 2014.

[11] E. Maestre, J. S. Abel, J. O. Smith, and G. P. Scavone, "Constrained pole optimization for modal reverberation," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2017.

[12] C. Kereliuk, W. Herman, R. Wedelich, and D. J. Gillespie, "Modal analysis of room impulse responses using subband ESPRIT," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2018.

[13] H. Kuttruff, *Room Acoustics*, 6th ed. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2017.

[14] M. Ducceschi, C. J. Webb, and G. Fratoni, "Efficient synthesis of large room impulse responses in the modal domain," in *2025 Immersive and 3D Audio (I3DA)*. IEEE, 2025.

[15] OpenAIR Library. OpenAIR: The open acoustic impulse response library. Accessed: Mar. 2025. [Online]. Available: https://www.openair.hosted.york.ac.uk.

[16] International Organization for Standardization, *ISO 3382-1:2009 — Acoustics — Measurement of room acoustic parameters — Part 1: Performance spaces*, Std., 2009.

[17] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed. Hoboken, NJ: Wiley, 2009.

[18] R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," *J. Acoust. Soc. Am.*, vol. 45, no. 3, pp. 545–565, 1969.

[19] V. Välimäki, K. Prawda, and S. J. Schlecht, "Two-stage attenuation filter for artificial reverberation," *IEEE Signal Process. Lett.*, vol. 31, pp. 391–395, 2024.

[20] S. J. Schlecht, J. Fagerström, and V. Välimäki, "Decorrelation in feedback delay networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 3478–3487, 2023.