NON-ITERATIVE MODAL PARAMETER ESTIMATION FOR PLATE REVERBS VIA MATRIX-PENCIL-GUIDED STATE SPACE MODEL INITIALIZATION

Matthias Bittner and Axel Jantsch

Christian Doppler Laboratory for Embedded Machine Learning
TU Wien
Vienna, Austria
firstname.lastname@tuwien.ac.at

ABSTRACT

Modal parameter identification for plate reverbs remains a challenging problem in virtual-analog audio effect emulation. Though neural network-based black-box approaches achieve high modeling accuracy, they generally lack interpretability and do not provide access to physically meaningful modal parameters. In this work, we present our solution to Task B of the DAFx Plate Reverb Parameter Estimation Challenge. Our method first estimates the total number of modes and then employs a Matrix Pencil (MP)-guided eigenvalue initialization strategy for a diagonal complex-valued State Space Model (SSM), which can be interpreted as a bank of parallel second-order all-pole filters. Exploiting the linearity of the resulting system, we compute the state impulse responses and replace gradient-based optimization with a closed-form least-squares estimation of the modal gains. The proposed approach enables accurate recovery of the modal parameters while maintaining an interpretable system representation.

1. INTRODUCTION

Classical plate reverbs are commonly modeled through modal resonator formulations [1], where each mode of the plate is represented as an exact discrete-time second-order all-pole biquad filter with the following discrete-time recurrence,

$$q_{m,k+1} = 2r_m \cos(\Omega_m \mathrm{T_s}) q_{m,k} - r_m^2 q_{m,k-1} + b_m f_k. \tag{1}$$

In this equation $q_{m,k}$ denotes the discrete-time amplitude of the m-th resonant mode, Ω_m its angular frequency, $r_m = e^{-\sigma_m \mathrm{T_s}}$ the decay factor derived from the modal damping σ_m, f_k the excitation signal, and b_m the mode-dependent excitation gain and the sampling-time $\mathrm{T_s}$. Applying the z-transform to Equation 1 and considering the full plate reverb as the sum of M modes, the transfer function of the modal reverb architecture results in,

$$G(z) = \sum_{m=1}^{M} \frac{b_m z}{z^2 - 2zr_m \cos(\Omega_m \mathrm{T_s}) + r_m^2}. \tag{2}$$

In our aligned DAFx26 contribution [2], we show that a restricted diagonal complex-valued State Space Model (SSM), such as S-Edge [3], can be an exactly equivalence to a set of parallel second-order all-pole filters. For completeness, a condensed derivation of this equivalence is provided in Section 2.4.

Building on this formulation, our previous attempt to solve the DAFx Plate Reverb Parameter Estimation Challenge (Task B) [4] demonstrated that a Matrix Pencil (MP)-guided eigenvalue initialization enables accurate plate Impulse Response (IR) fitting and meaningful recovery of modal parameters when using gradient-based optimization. However, the method lacked the ability to estimate the actual number of modes and relied on iterative optimization, which becomes computationally demanding for long, high-sampling-rate impulse responses with a high modal count. This challenge solution addresses both limitations by extending the approach proposed in [2] in two ways:

- We introduce a total mode count estimator, to automatically determine the model order used to estimate and extrapolate modes.
- We exploit the linearity of the resulting system and replace iterative gradient-based optimization with least-squares estimation of the modal gains.

2. METHOD

This challenge submission proposes a non-iterative approach for Task B of the DAFx26 parameter estimation challenge. Figure 1 shows an overview of the proposed three-step approach.

In the first step, we fit a model for the task of estimating the total number of modes M_{total}.

In the second step for each individual tested impulse response, we first use the total mode count estimator to get M_{total}. We then estimate frequency and decay rates with the MP method [5], resulting in a set of low-frequency modes $(\boldsymbol{\Omega}_{\mathrm{est}}, \boldsymbol{\sigma}_{\mathrm{est}})$, with frequencies $\boldsymbol{\Omega}_{\mathrm{est}} \in \mathbb{R}^{M_{\mathrm{est}}}$ and decay rates $\boldsymbol{\sigma}_{\mathrm{est}} \in \mathbb{R}^{M_{\mathrm{est}}}$. The M_{est} poles are then used to fit a polynomial function between decay rates and frequencies $\sigma = f(\Omega)$. Beyond the number of initially found poles, we extrapolate by uniformly sampling new frequencies $\boldsymbol{\Omega}_{\mathrm{ex}}$ between $[\max(\boldsymbol{\Omega}_{\mathrm{est}}), \Omega_{\max}]$ until we reach the M_{total} mode count and evaluate the decay rates $\boldsymbol{\sigma}_{\mathrm{ex}}$ based on the previously fitted function.

In a third step, all estimated modes, frequencies $\boldsymbol{\Omega}_{\mathrm{total}} = [\boldsymbol{\Omega}_{\mathrm{est}}, \boldsymbol{\Omega}_{\mathrm{ex}}]$, decay rates $\boldsymbol{\sigma}_{\mathrm{total}} = [\boldsymbol{\sigma}_{\mathrm{est}}, \boldsymbol{\sigma}_{\mathrm{ex}}]$ are used to initialize the eigenvalues of a diagonal complex-valued SSM. We then run the model once by feeding an impulse with height 1 at the initial timestep, monitor the state's response, fit the output weights of the SSM using least-squares, and convert the SSM output weights to gains that match the modal representation of Equation 2.

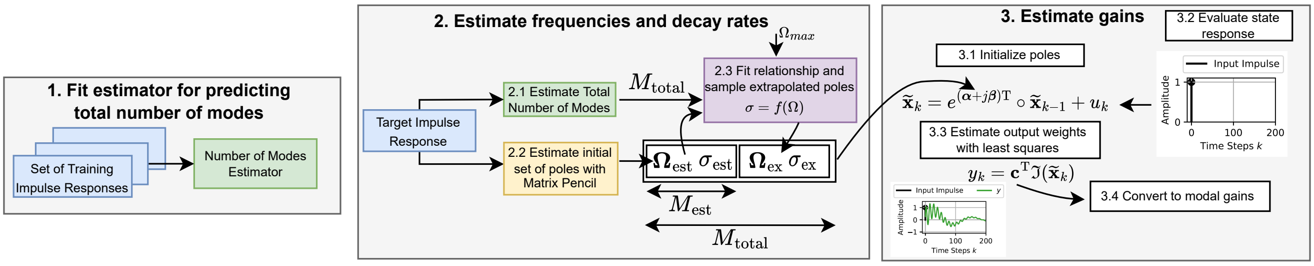

Figure 1: Overview of the proposed three-step approach. First, we fit a regression model that estimates the total number of modes (M_{total}) from the impulse response. Second, we estimate low-frequency poles using the Matrix Pencil (MP) method. We then fit a frequency-dependent decay-rate model, which is subsequently extrapolated to obtain the remaining modes up to (M_{total}). Third, all estimated modes are used to initialize a diagonal complex-valued State Space Model (SSM), whose output weights are fitted via least squares and converted to modal gains.

2.1. Total Mode Count Estimator

By visually inspecting the magnitude transfer function, we observed that impulse responses with a high or a low number of modes are already visually distinguishable, especially by their local structure. This motivated the development of a scalar spectral richness measure that serves as input to a regression task used to estimate the total number of modes M_{total}. For calculating the spectral richness measure, we first logarithmically compress the magnitude response to emphasize weaker spectral components. We then use this compressed spectrum and divide it into overlapping local patches to capture short-range spectral structure. All patches are stacked into a data matrix and independently mean-centered to remove global energy offsets. Singular Value Decomposition (SVD) is subsequently applied to this matrix in order to analyze the diversity of local spectral patterns contained in the impulse response.

If the impulse response contains only a small number of resonant modes, many local spectral patches appear similar, and the data matrix can be represented using only a few dominant singular vectors. Conversely, systems with more modes exhibit more diverse local spectral structures, leading to a broader singular value distribution. To quantify this effect, the entropy of the normalized singular values is computed and exponentiated to yield a scalar spectral richness measure. Given a set of impulse responses (128 in our experiments), we fit a regression model (isotonic regression in our implementation[1]) to estimate the total number of modes with the proposed spectral richness measure serving as a scalar input.

2.2. Frequency and Decay Rate Estimation

Frequencies and decay rates are estimated in a three-step approach:

1. Given a IR, we estimate the total number of modes M_{total}.
2. Based on the IR we then estimate, $(\boldsymbol{\Omega}_{\mathrm{est}}, \boldsymbol{\sigma}_{\mathrm{est}})$, the first subset of modes, with the MP method [5].
3. The M_{est} poles are then used to fit a polynomial function between decay rates and frequencies $\sigma = f(\Omega)$. Afterward, we extrapolate by uniformly sampling new frequencies between $[\max(\boldsymbol{\Omega}_{\mathrm{est}}), \Omega_{\max}]$ until we reach the estimated M_{total} mode count and evaluate the decay rates based on the previously fitted function. The maximum frequency in our experiments is set to $\Omega_{\max} = 10\,\mathrm{kHz}$.

The MP method provides reliable pole estimates for modes that are sufficiently separated from noise. However, it requires the SVD of large Hankel matrices whose dimensions scale with the impulse response length T. Due to the cubic cost of the SVD, this is a bottleneck. To reduce computational cost, we subsample the impulse response by a factor of four and restrict the analysis to the first second of the IR for the initial MP pole estimation stage. The order M_{est} is determined from the Hankel matrix's singular values. Specifically, the cumulative energy contribution of the singular values is computed. The selected model order corresponds to the smallest number of singular values required to explain at least 99.999% of the total singular-value energy.

2.3. Gain Estimation

We use the total number of estimated frequencies $\boldsymbol{\Omega}_{\mathrm{total}}$ and decay rates $\boldsymbol{\sigma}_{\mathrm{total}}$ to initialize the poles of a diagonal complex-valued SSM layer defined by the difference equation,

$$\begin{aligned} \widetilde{\mathbf{x}}_k &= e^{\widetilde{\boldsymbol{\lambda}}\mathrm{T_s}} \circ \widetilde{\mathbf{x}}_{k-1} + u_k, \\ y_k &= \mathbf{c}^{\mathrm{T}} \Im(\widetilde{\mathbf{x}}_k). \end{aligned} \tag{3}$$

This Single Input Single Output (SISO) system maps a real-valued input $u_k \in \mathbb{R}^1$ to a real-valued output $y_k \in \mathbb{R}^1$, via a complex-valued state $\widetilde{\mathbf{x}}_k \in \mathbb{C}^{M_{\mathrm{total}}}$. The dynamic of the states is defined by the continuous-time eigenvalue representation $\widetilde{\boldsymbol{\lambda}} = \boldsymbol{\alpha} + j\boldsymbol{\beta}$, we initialize with the estimates $\boldsymbol{\beta} = \boldsymbol{\Omega}_{\mathrm{total}}$ and $\boldsymbol{\alpha} = -\boldsymbol{\sigma}_{\mathrm{total}}$. The sampling rate $\mathrm{f_s} = 1/\mathrm{T_s} = 44.1\,\mathrm{kHz}$ of the investigated IR scales the eigenvalues back to the frequency normalized domain.

As a next step, we evaluate the state response $\mathbf{X} \in \mathbb{R}^{M_{total} \times T}$ by sending in a single impulse $u_0 = 1$. Using the target impulse response $\mathbf{Y} \in \mathbb{R}^{1 \times T}$ we can derive the output weights $\mathbf{c}$ for the SSM layer by least squares,

$$\mathbf{c} = \mathbf{Y}\mathbf{X}^{\mathsf{T}} \left(\mathbf{X}\mathbf{X}^{\mathsf{T}}\right)^{-1}. \tag{4}$$

This resulting vector corresponds to the individual weighting of the state's contribution but needs to be converted to match the gains b_m in the all-pole filter representation. The two transfer

[1] https://scikit-learn.org/stable/modules/isotonic.html

functions representations follow as,

$$G_{\mathrm{all-pole}}(z) = \sum_{m=1}^{M} \frac{b_m z}{z^2 - 2zr_m\cos(\Omega_m \mathrm{T_s}) + r_m^2},$$
$$G_{\mathrm{SSM}}(z) = \sum_{m=1}^{M} c_m \frac{ze^{\alpha_m \mathrm{T_s}}\sin(\beta_m \mathrm{T_s})}{z^2 - 2ze^{\alpha_m \mathrm{T_s}}\cos(\beta_m \mathrm{T_s}) + e^{2\alpha_m \mathrm{T_s}}}. \tag{5}$$

We see that the eigenvalues directly match the frequency and decay rates, and the gains in the all-pole representation are defined by,

$$b_m = c_m e^{\alpha_m \mathrm{T_s}} \sin(\beta_m \mathrm{T_s}). \tag{6}$$

2.4. Diagonal Complex-Valued State Space Models as Parallel Second Order All-pole Filters

For completeness and as already similarly proposed in our DAFx26 contribution [2], we derive the input-output mapping of our restricted complex-valued diagonal SSM and show its equivalence to a parallel second-order all-pole filter structure. Applying the z-transform to the input-output mapping (Equation 3, for now considering the full complex output) allows us to go from the discrete-time domain k into the z-domain,

$$\begin{aligned} \widetilde{\mathbf{x}}_z(z) &= z^{-1} e^{\widetilde{\boldsymbol{\lambda}}\mathrm{T_s}} \circ \widetilde{\mathbf{x}}_z(z) + u_z(z), \\ \widetilde{y}_z(z) &= \mathbf{c}^{\mathrm{T}} \widetilde{\mathbf{x}}_z(z), \end{aligned} \tag{7}$$

with the known relationship of $z = e^{j\omega \mathrm{T_s}}$ and the angular frequency ω. Based on this representation and the diagonal structure, we are able to formalize the complex-valued discrete-time transfer function as a sum of first-order complex resonators,

$$\widetilde{G}(z) = \frac{\widetilde{y}_z(z)}{u_z(z)} = \sum_{m=1}^{M} c_m \frac{z}{z - e^{\widetilde{\lambda}_m \mathrm{T_s}}}. \tag{8}$$

Applying the inverse z-transform of this representation and taking the imaginary part as proposed in Equation 3 gives the real-valued discrete-time impulse response,

$$\begin{aligned} g_k &= \Im(\widetilde{g}_k) = \Im\left(\sum_{m=1}^{M} c_m e^{\widetilde{\lambda}_m k \mathrm{T_s}}\right) \\ &= \Im\left(\sum_{m=1}^{M} c_m e^{\alpha_m k \mathrm{T_s}} \left(\cos(\beta_m k \mathrm{T_s}) + j\sin(\beta_m k \mathrm{T_s})\right)\right). \\ &= \sum_{m=1}^{M} c_m e^{\alpha_m k \mathrm{T_s}} \sin(\beta_m k \mathrm{T_s}), \end{aligned} \tag{9}$$

which represents a weighted sum of exponentially decaying sinusoids. The real part of the eigenvalue α_m defines the decay rate, and the imaginary part β_m defines the angular frequency for each individual mode. If we now again apply the z-transform on g_k we get the actual real-valued transfer function representation of our proposed input-output mapping,

$$G_{\mathrm{SSM}}(z) = \sum_{m=1}^{M} c_m \frac{ze^{\alpha_m \mathrm{T_s}}\sin(\beta_m \mathrm{T_s})}{z^2 - 2ze^{\alpha_m \mathrm{T_s}}\cos(\beta_m \mathrm{T_s}) + e^{2\alpha_m \mathrm{T_s}}}. \tag{10}$$

3. EXPERIMENTS

While the ground-truth data for the 16 test IRs in the DAFx challenge are hidden and not publicly available, we present results for our approach based on a self-generated set of 17 random IRs.

Ducceschi and Gabrielli propose a baseline for the modal parameter Estimation (Task B) [4]. They use a two-stage procedure. First, they identify modal peaks in the magnitude response of the IR to estimate modal frequencies and bandwidths, from which damping (loss) coefficients are derived using the half-power bandwidth rule. Second, modal gains are estimated by evaluating the imaginary part of the frequency response at the detected peak locations. The baseline is used as the default comparison.

Total Mode Count Estimation: Figure 2 compares the actual number of modes, the number of identified modes using the challenge's baseline method, and the proposed spectral richness regression approach (our-method). We see that our proposed M_{total} mode-count estimator (fitted on an independent set of 128 IRs, sampled with the same hyperparameter settings as the later tested IRs) clearly outperforms the baseline.

Frequency and Decay Rate Estimation: Figure 3 shows a comparison between the identified modes vs actual ground truth modes for the best estimated IR sample (ID=11), showing decay rates vs frequencies of the individual modes. We see that the low-frequency modes are estimated relatively well using the MP method. Given accurate estimates of low-frequency modes, our fitted relationship enables accurate extrapolation.

Challenge Metric Comparison: Figure 4 compares the challenge evaluation metric RE combining the relative per plate error metrics for the identified modal frequencies RE_Ω, decay rates RE_σ, and gains RE_b, where, for $p \in \{\Omega, \sigma, b\}$, the per-plate relative error is $RE_p = \frac{1}{M}\sum_{m=1}^{M} \min\left(1, \frac{|p_m^{\mathrm{est}} - p_m^{\mathrm{ref}}|}{p_m^{\mathrm{ref}}}\right)$, averaged over the M Hungarian-matched modes (unmatched modes contribute an error of 1). The true mode count is unknown a priori, RE_Ω, RE_Σ, and RE_b are therefore computed only over the matched-mode subset, while the mismatch between the true and estimated mode count is captured separately by ΔM (see Table 1). Due to the relatively accurate total mode count estimation paired with the fact that the identified and actual modes are matched one-to-one by optimal (Hungarian) assignment on the log-frequency distance, we observe a good improvement on RE_Ω. For IRs where RE_σ is higher, we have a weaker match between the MP estimated modes and therefore a weaker extrapolation performance.

Compute cost and time: The analyzed impulse responses are 5 s long at 44.1 kHz. We perform MP pole estimation on the first second, subsampled by a factor of four ($\sim$11k samples), taking around 30 s per IR. State-response evaluation and least-squares fitting run on CPU, since state sizes ($\gg$2k) exceeded available GPU memory for the parallel scan. For the challenge submission, the 16-file test set took about 42 minutes on a server with 2$\times$ Intel Xeon Gold 5118 (2.3 GHz), 512 GB RAM.

Table 1: Average metrics (as proposed in [4]) over the 17 analyzed IRs: $RE = RE_0 + \min(1, \Delta M/M)$, with $RE_0 = \frac{1}{3}(RE_\Omega + RE_\sigma + RE_b)$, and the missed mode count ΔM.

	RE	RE_0	RE_Ω	RE_σ	RE_b	ΔM
baseline	1.96	0.98	0.98	0.98	0.98	3239
our-meth.	0.78	0.67	0.27	0.77	0.97	335

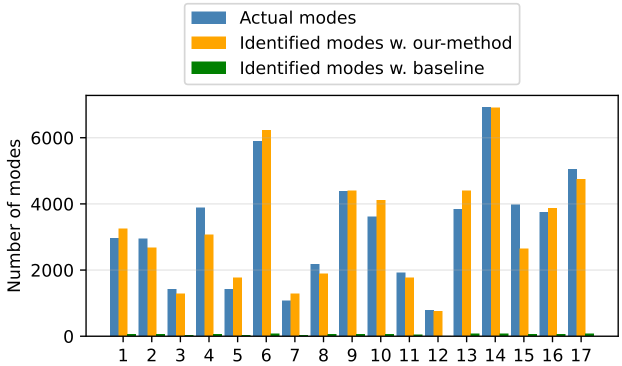

Figure 2: Comparison between the actual number of modes, the number of identified modes with the challenge's baseline method, and the proposed spectral richness regression approach (our-method) for a tested set of 17 IRs.

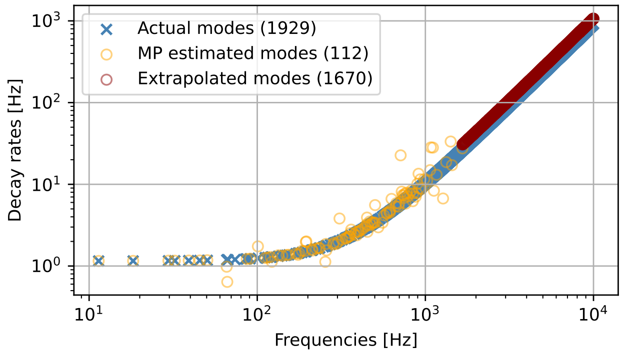

Figure 3: Estimated modes vs actual ground truth modes for the IR sample (ID=11) reaching the best evaluation metric RE. All estimated modes are composed of the MP-estimated modes and the extrapolated modes.

4. DISCUSSION AND CONCLUSION

The proposed pipeline decouples total mode count estimation, pole initialization, and gain fitting into closed-form stages, removing the need for gradient-based optimization. While we adopt the diagonal complex-valued SSM formulation of [2, 3], the closed-form gain fit in Equation 4 does not strictly require it. In practice, any linear basis of mode responses works, including one generated directly via the biquad recurrence of Equation 1. We favor the SSM formulation since its linear diagonal structure enables efficient parallel-scan inference, reducing the compute cost of evaluating mode responses to $\mathcal{O}(\log T)$, and its complex eigenvalues map directly to frequency/decay pairs. This comes at a memory cost, since the parallel scan materializes a state trajectory of size $M_{total} \times T$, which exceeded GPU memory for our high-mode-count IRs and forced CPU evaluation. The naive sequential recurrence has a much smaller memory footprint, at a computational cost of $\mathcal{O}(T)$. In summary, we presented a non-iterative solution for Task B of the modal parameter estimation challenge, combining a fitted mode-count estimator, MP-guided pole initialization with extrapolation, and closed-form gain fitting. This approach avoids iterative training while improving overall performance relative to the baseline, most notably in mode-count and frequency estimation.

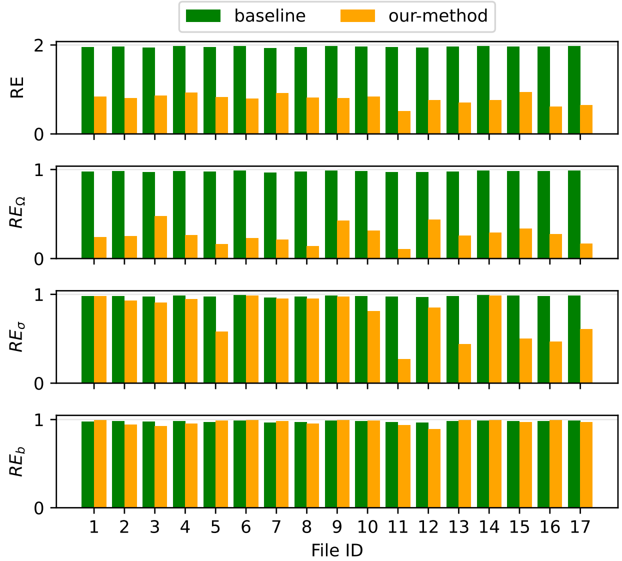

Figure 4: Evaluation metric comparison between the baseline and our challenge proposal (our-method). We show per plate evaluation metrics RE combining the relative per plate error metrics for the identified modal frequencies RE_Ω, decay rates RE_σ, and gains RE_b.

5. ACKNOWLEDGMENTS

This work is supported by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology, and Development, the Christian Doppler Research Association.

6. REFERENCES

[1] 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: from Architecture to Automotive (I3DA)*, 2025, pp. 1–8.

[2] M. Bittner, M. Wess, D. Dallinger, , D. Schnöll, and A. Jantsch, "Diagonal complex-valued state space models for system identification and modeling of metal plate reverbs," in *Proceedings of the 29th International Conference on Digital Audio Effects (DAFx26)*, Cambridge, USA, September 2026.

[3] M. Bittner, D. Schnöll, M. Wess, and A. Jantsch, "Efficient and interpretable raw audio classification with diagonal state space models," *Mach. Learn.*, vol. 114, no. 8, p. 175, 2025.

[4] M. Ducceschi and L. Gabrielli, "The 1st dafx parameter estimation challenge: A benchmark for plate reverb system identification," 2026.

[5] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.