THE 1ST DAFX CHALLENGE: PHYSICAL AND MODAL PARAMETER ESTIMATION FOR PLATE REVERBERATION

Leonardo Gabrielli *

Dept. of Information Engineering
Università Politecnica delle Marche
Ancona, IT
l.gabrielli@staff.univpm.it

Michele Ducceschi *

Dept. of Industrial Engineering
Università di Bologna
Italy
michele.ducceschi@unibo.it

ABSTRACT

The 1st DAFx Parameter Estimation Challenge is an open initiative to advance the state of the art in parameter estimation for acoustic modeling. Stated as a system identification problem, this first edition focuses on plate reverberation—an archetypal dense, modal and weakly damped acoustic system. Participants tackled two tasks: (A) estimating the physical parameters of a vibrating plate from its impulse response, and (B) recovering the modal parameters of the same system. Both rest on a simulation framework based on the damped Kirchhoff–Love plate equation, and both are posed and scored entirely on synthetic data produced by that framework: no measurement of a real plate is involved. Two participants solved Task A down to machine precision by different strategies: one a neural network trained on a very large dataset, and one gradient-free optimization with many inexpensive evaluations. Task B proved considerably harder: the best submission attains a relative error of 0.33 on a $[0, 2]$ scale, and every method recovers modal frequencies and decay rates far more accurately than modal gains. A complementary frequency-domain evaluation reorders the ranking and exposes a systematic gain bias to which the per-mode metric is blind.

1. INTRODUCTION

Digital emulations of plate reverberation units have a long history in audio effects [1–3]. While the physical behaviour of thin plates is well understood, the inverse problem—identifying either material parameters or modal parameters from audio data—remains a challenging task. Dense modal overlap, frequency-dependent damping, and parameter coupling make the system difficult to invert. In preliminary work by the authors, the task proved challenging enough to motivate the creation of a shared benchmark: a well-defined dataset and evaluation framework that allows fair comparison between diverse methods. The dataset and the benchmark have been framed as an academic challenge, the 1st DAFx Parameter Estimation Challenge. It was announced at the 28th International Conference on Digital Audio Effects in September 2025 and was conducted during the first months of 2026. A detailed statement of the rules, the data format and the submission protocol is given in the challenge white paper [4].

* Identical contributions

This paper summarizes the challenge problem and the dataset, and provides a survey of the methods proposed by the participants. Results are reported here for further reference and will be announced at the conference.

The paper is organized as follows. Section 2 describes the modal plate model, Section 3 the two tasks, and Section 4 the datasets, baselines and evaluation metrics. Section 5 summarizes the submissions and Section 6 reports their results. Section 7 draws more general lessons and considers how future editions may add complexity, and Section 8 closes the paper.

2. THE PHYSICAL MODEL

Among the earliest electromechanical reverberation devices is the plate reverberator, in which a thin metal sheet is driven by an exciter and picked up by contact transducers [1]. A plate of practical size supports a very dense set of weakly damped bending modes, so its impulse response (IR) resembles that of a reverberant room, and plates were widely used as a compact substitute for a physical echo chamber. What the organizers distributed is not a plate but a *digital model* of one: the modal simulator of the damped Kirchhoff–Love plate outlined below, derived in full in the challenge white paper [4], which generated every IR in the challenge.

The plate is rectangular with side lengths L_x, L_y, and its transverse displacement $u(x, y, t)$ is defined on $[0, L_x] \times [0, L_y] \times \mathbb{R}_0^+$. The motion follows the damped Kirchhoff–Love equation [5]:

$$\rho h \frac{\partial^2 u}{\partial t^2} = -\rho h \mathcal{K} u - 2\rho h \left(\eta_0 + \eta_1 \mathcal{K}\right) \frac{\partial u}{\partial t} + \delta(\mathbf{x} - \mathbf{x}_i) f(t), \quad (1)$$

where $\mathbf{x} := (x, y)$ and $\mathcal{K}$ is the Kirchhoff–Love plate operator,

$$\mathcal{K} := -\frac{T_0}{\rho h}\Delta + \frac{D}{\rho h}\Delta\Delta, \quad (2)$$

with Δ and $\Delta\Delta$ the Laplace and biharmonic operators. In Eq. (1), $\delta(\mathbf{x} - \mathbf{x}_i) := \delta(x - x_i)\delta(y - y_i)$ is a two-dimensional Dirac delta placing the excitation at $\mathbf{x}_i := (x_i, y_i)$, and $f(t)$ is the input, generally a dry audio signal. Throughout this work the input is an ideal impulse, $f(t) := \delta(t)$, where $\delta(t)$ is the one-dimensional Dirac distribution on the continuous time axis, defined by $\int_{-\infty}^{\infty} \delta(t)\phi(t)\,\mathrm{d}t = \phi(0)$ for any test function ϕ. It should not be confused with the discrete-time unit sample $\delta[k]$, which is what the simulator applies once the model has been discretised (Sec. 2.3). Output is extracted at a single point $\mathbf{x}_o := (x_o, y_o)$.

2.1. Plate Parameters

The constants appearing in Eq. (1) are: ρ, the volumetric density; h, the thickness; T_0, the applied in-plane tension per unit length;

Table 1: Plate parameter ranges. Rows marked *Fixed* take a single value for all simulations; the remaining rows are drawn independently and uniformly at random from the indicated range.

Parameter	Symbol	Min	Max	Units
Plate length (x)	L_x	1.0		m (Fixed)
Plate length (y)	L_y	1.1	4.0	m
Thickness	h	0.001	0.005	m
Tension per length	T_0	0.01	1000.0	N/m
Material density	ρ	2430.0	21230.0	kg/m^3
Young's modulus	E	6.7×10^{10}	2.2×10^{11}	Pa
Poisson's ratio	ν	0.25		Fixed
Decay times	τ_0, τ_1	6.0, 2.0		s (Fixed)
Reference loss freq.	f_1	500.0		Hz (Fixed)
Input position	(x_i, y_i)	$(0.335\,L_x, 0.467\,L_y)$		Fixed
Output position (x)	x_o	$0.51\,L_x$	L_x	
Output position (y)	y_o	$0.51\,L_y$	L_y	

$D := Eh^3/12(1-\nu^2)$, the flexural rigidity, with E Young's modulus and ν Poisson's ratio; and η_0, η_1, two loss parameters measured in s^{-1} and s. The full set of plate parameters is therefore

$$P = \{\rho, E, \nu, L_x, L_y, h, T_0, \eta_0, \eta_1, x_i, y_i, x_o, y_o\}, \quad (3)$$

which also includes the input and output locations. Table 1 summarises all parameters with their upper and lower bounds.

2.2. Modal Representation

Plate reverb units are typically suspended from tensioned springs attached to a heavy frame, and therefore feature free edges. Here, for simplicity, the effect of the spring tension is incorporated into the global parameter T_0 and the boundaries are assumed simply supported, so that $u = \partial^2 u/\partial n^2 = 0$ along the boundary for all t, with n the direction perpendicular to each edge. This admits an analytical solution of the eigenproblem, formulated in the absence of losses and forcing. The mode shapes are

$$\Phi_m := \frac{2}{\sqrt{L_x L_y}} \sin\frac{m_x \pi x}{L_x} \sin\frac{m_y \pi y}{L_y}, \quad (4)$$

where the mode index m is shorthand for a pair of positive integers (m_x, m_y).They satisfy the boundary conditions and are orthonormal over the plate, i.e. $\int_0^{L_x}\int_0^{L_y} \Phi_m \Phi_{m'}\,\mathrm{d}x\,\mathrm{d}y = \kappa_{m,m'}$, with $\kappa_{m,m'} = 1$ if $m = m'$ and zero otherwise. Applying the plate operator gives $\mathcal{K}\Phi_m = (\rho h)^{-1}(T_0\gamma_m^2 + D\gamma_m^4)\Phi_m$ with $\gamma_m = (m_x^2\pi^2/L_x^2 + m_y^2\pi^2/L_y^2)^{1/2}$, so that the undamped modal angular frequencies are

$$\Omega_m = (\rho h)^{-\frac{1}{2}} \gamma_m (T_0 + D\gamma_m^2)^{\frac{1}{2}}. \quad (5)$$

Writing the general solution as a superposition of modes, $u = \sum_m \Phi_m q_m$, substituting into (1), multiplying by $\Phi_{m'}$ and integrating over the plate's domain, orthonormality reduces the PDE to a bank of independent damped oscillators,

$$\ddot{q}_m(t) = -\Omega_m^2 q_m - 2\sigma_m \dot{q}_m + (\rho h)^{-1}\Phi_m(x_i, y_i)\delta(t), \quad (6)$$

for $m = 1, \ldots, M$, with $\sigma_m := \eta_0 + \eta_1\Omega_m^2$ a frequency-dependent loss coefficient. In the frequency domain each mode is a second-order all-pole section with residue $\Phi_m(x_i, y_i)\Phi_m(x_o, y_o)$, and the plate response measured at $\mathbf{x}_o$ is the parallel sum of M such sections.

In the code, the decay constants η_0, η_1 are not selected directly, but rather by choosing the decay times τ_0, τ_1 at two angular frequencies Ω_0, Ω_1. For model (6) the amplitude decays as $e^{-\sigma_m t}$, so the 60 dB decay time is $\tau_m = 3\log(10)/\sigma_m$, giving

$$\eta_0 = \frac{3\log(10)}{\Omega_1^2 - \Omega_0^2}\left(\frac{\Omega_1^2}{\tau_0} - \frac{\Omega_0^2}{\tau_1}\right), \quad (7a)$$

$$\eta_1 = \frac{3\log(10)}{\Omega_1^2 - \Omega_0^2}\frac{\tau_0 - \tau_1}{\tau_0\tau_1}. \quad (7b)$$

In the code $\Omega_0 = 0$ (DC) and $\Omega_1 = 1000\pi$, i.e. a linear frequency of 500 Hz, so only τ_0 and τ_1 are required. For energy dissipation at all frequencies the decay times must satisfy $\tau_0 > \tau_1$, a condition enforced in the code provided on GitHub.

2.3. Discrete-time model

Updating equation (6) in time is accomplished by discretising each $q_m(t)$ with a *time series* $q_m[k]$, where k is the *time index* such that $q_m[k] \approx q_m(t_k)$, with $t_k := Tk$. Here T is the sampling period, i.e. the reciprocal of the sampling rate $f_s = 1/T$, fixed at $f_s = 44\,100$ Hz throughout the challenge; every distributed IR is one second long.

In the Python code provided on GitHub, an update scheme is already implemented using the exact form for an all-pole biquad discretisation [6, 7]. The update equation is

$$q_m[k+1] = 2r_m\cos(\Omega_m T)q_m[k] - r_m^2 q_m[k-1] + b_m f[k], \quad (8)$$

with

$$r_m := e^{-\sigma_m T},$$

$$b_m := \frac{4T^2\Phi_m(x_i, y_i)\Phi_m(x_o, y_o)r_m}{\rho h L_x L_y}.$$

Note that the output projection $\Phi_m(x_o, y_o)$ is absorbed into the gain b_m, so that the discrete-time output is recovered directly as $u[k] = \sum_m q_m[k]$. In the z-domain, with $z = e^{j\omega T}$, the transfer function from the input f to the contribution of mode m is thus

$$H_m(e^{j\omega T}) = \frac{b_m e^{-j\omega T}}{1 - 2r_m\cos(\Omega_m T)e^{-j\omega T} + r_m^2 e^{-2j\omega T}}, \quad (9)$$

and the full plate response is $H = \sum_m H_m$. This is the discrete-time Fourier transform of mode m for scheme (8). The modal workflow is summarised in Fig. 1.

3. CHALLENGE TASKS

The challenge consisted of two branches, both relying on synthetic data generated by the reference model. The two tasks shared the same inverse-estimation setting, but targeted different parameter families. Task A targeted estimation of a subset of the plate parameters P listed in Eq. (3), i.e. a small set of physical, geometrical and material parameters defining the plate. Task B targeted estimation of the modal parameters describing the plate response, a larger-scale problem with thousands of modal parameters to estimate.

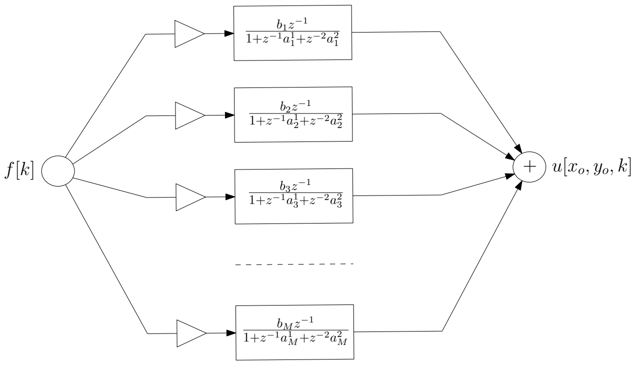

Figure 1: Modal reverb architecture. A sampled input $f[k]$ is projected onto a bank of modes, described by an all-pole biquad filterbank. Each filter is represented by three coefficients, related to the mode's natural frequency, decay time, and gain. The output at location (x_o, y_o) is reconstructed by summing all modal contributions.

3.1. Task A: Physical Parameter Identification

Given an impulse response from the reference simulator, the goal in Task A was to infer the subset $S(P)$ of physical plate parameters in (3):

$$S(P) = \{\mu := \rho h, D/\mu, T_0/\mu, L_y, x_o, y_o\}. \tag{10}$$

The remaining parameters of P—namely L_x, ν, η_0, η_1, x_i and y_i—were held *fixed* across all simulations, at the values given in the rows of Table 1 marked *Fixed*. Each draw of $S(P)$ produces a unique impulse response and modal distribution.

Identifiability and dataset format The raw physical parameters (ρ, h, E, T_0) appear in the equation of motion (1) only through the three invariants $\mu = \rho h$, D/μ and T_0/μ: a one-parameter family of (ρ, h, E, T_0) tuples therefore produces exactly the same impulse response. Working with $S(P)$ rather than the raw parameters eliminates this ambiguity and makes the ground truth unique.

Moreover, all six entries of $S(P)$ are identifiable *only* when the absolute amplitude of the IR is preserved. The numerator of the biquad, b_m in (8), is proportional to $1/(\rho h)$, so the overall scale of the IR carries information about μ, and peak-normalizing the waveform discards it. For this reason, the dataset was generated by synthesizing displacement IRs with absolute amplitude preserved. For each target, two representations were distributed: a NumPy array (.npz) and a normalized audio waveform (.wav).

3.2. Task B: Modal Parameter Estimation

Given an impulse response or frequency response, Task B required retrieval, for each mode, of the modal parameters

$$M_m = \{\Omega_m, \sigma_m, b_m\},\ m = 1, \ldots, M, \tag{11}$$

corresponding to the mode's natural frequency, decay constant (in s^{-1}) and modal gain, as they appear in the update equation (8). The gain b_m is the numerator of the discrete-time biquad in the z-domain, as defined by (8); ground-truth values follow this convention.

4. DATASETS, BASELINES AND EVALUATION METRICS

Since the plate model was supplied to the participants, they were able to produce their own training data for data-driven algorithms and/or a validation set to track their performance.

The test dataset, released later, comprised 16 IRs without the ground-truth parameters, which were kept secret for the purpose of evaluating the submissions. The number of IRs was chosen as a trade-off between the statistical significance of the evaluation metrics and the computational run time of the evaluation. The IRs were generated at random from a secret seed.

4.1. Evaluation for Task A

For Task A, the main metric for plate parameter estimation is the Normalized Mean Squared Error (NMSE):

$$\text{NMSE} = \frac{1}{N_S} \sum_{i=1}^{N_S} \frac{\left(\alpha_i^{(\text{est})} - \alpha_i^{(\text{ref})}\right)^2}{\left(\alpha_i^{(\max)} - \alpha_i^{(\min)}\right)^2}. \tag{12}$$

Here $N_S = 6$ is the number of parameters in $S(P)$, and $\alpha_i^{(\max)} - \alpha_i^{(\min)}$ is the range of the i-th parameter. For L_y, x_o and y_o the ranges are those of Table 1; for the three derived quantities μ, D/μ and T_0/μ they are the bounding boxes induced by the raw-parameter ranges and are computed deterministically by the evaluation script (at the time of writing, $\mu \in [2.43, 106.15]$, $D/\mu \in [0.28, 201.2]$, $T_0/\mu \in [9.4 \times 10^{-5}, 411.5]$).

As an auxiliary signal-domain measure, we also compute the spectral MSE (SMSE) on the log-magnitude FFT of each estimated IR:

$$\text{SMSE} = \frac{1}{K} \sum_{k=1}^{K} \Big(\log\big(|X[k]| + \varepsilon\big) - \log\big(|\hat{X}[k]| + \varepsilon\big)\Big)^2, \tag{13}$$

where $X[k]$ and $\hat{X}[k]$ are the FFT magnitudes of the reference and estimated IRs after length matching, K is the number of retained frequency bins, and $\varepsilon = 10^{-10}$ avoids $\log(0)$.

Table 3 reports the sample mean $E\{\cdot\}$ and median $M\{\cdot\}$ of the per-example scores, for both NMSE and SMSE. The ranking was determined exclusively by the mean. The median is reported alongside it because some methods estimate most targets accurately while failing on a few, in which case the mean is dominated by a handful of outliers; we nonetheless consider $E\{\text{NMSE}\}$ the appropriate criterion, since it privileges consistency across targets.

4.2. Baselines for Task A

The Task A baseline is a Particle Swarm Optimization (PSO) search over the N_S parameters in normalized coordinates. Candidate solutions are scored with a multi-scale spectral loss (MSS) [8, 9] computed on log-magnitude short-time Fourier transforms (STFTs) at three resolutions, with Hann windows of length 512, 2 048 and 8 192 samples (equal to the FFT length) and hop sizes of one quarter of the window, i.e. 128, 512 and 2 048 samples. The three per-resolution losses are mean absolute differences of $\log(|X| + \varepsilon)$, $\varepsilon = 10^{-10}$, and are averaged. Using log magnitude is important because displacement IR amplitudes are very small ($\sim 10^{-8}$): it keeps differences between candidates and ground truth numerically informative. The baseline therefore behaves as a derivative-free global search driven by spectral matching. It was limited to 50 particles and 20 iterations, i.e. 1 000 evaluations in total, to keep the runtime reasonable, and required on average 33 minutes per IR on an i7-6850K CPU, IR synthesis being the most expensive part of the job.

A constant predictor (CP) serves as a lower-complexity reference: the MSE-optimal constant estimate for each parameter is its distribution mean, which for a uniform parameter on $[0, 1]$ yields NMSE $= 1/12$. The three derived parameters μ, D/μ and T_0/μ are products and ratios of uniform raw parameters, hence not uniformly distributed, so their constant-predictor NMSE is typically above $1/12$; the value is a reference order of magnitude rather than a strict threshold.

4.3. Baseline and Evaluation for Task B

For Task B, the baseline follows a two-stage estimator. In the first stage, peaks in the magnitude response are detected and the modal frequency and damping are estimated; frequency is refined by local parabolic interpolation around each peak, while damping is estimated from the half-power bandwidth, giving $\sigma_m \approx \Delta\omega_m/2$. In the second stage, the modal gain b_m is estimated from the imaginary part of the response near each estimated resonance, using the single-mode approximation around $\omega \approx \Omega_m$,

$$H(e^{j\Omega_m T}) \approx -j\frac{b_m}{2\sigma_m T \sin(\Omega_m T)}, \tag{14}$$

which follows from (9) for $\sigma_m T \ll 1$. This baseline works well when modal peaks are sparse and well separated, but it degrades in dense overlapping regimes where peak picking becomes unreliable.

Evaluation is based on the similarity between the discovered and the actual modal parameters. Identified and actual modes are matched one-to-one by optimal (Hungarian) assignment on log-frequency distance, so that confusing a 100 Hz mode with a 110 Hz mode costs the same as confusing 1 kHz with 1.1 kHz. A pair is accepted only if that distance is within a threshold, half an octave by default; pairs beyond it are treated as unmatched, yielding either a missed actual mode or a spurious identified one. Writing $\theta \in \{\Omega, \sigma, b\}$ for the parameter under consideration, the per-mode and per-plate relative errors are

$$\mathrm{re}_\theta^{(m)} = \min\left(1,\ \frac{|\theta_m^{(\mathrm{est})} - \theta_m^{(\mathrm{ref})}|}{\theta_m^{(\mathrm{ref})}}\right), \qquad \mathrm{RE}_\theta = \frac{1}{M}\sum_{m=1}^{M} \mathrm{re}_\theta^{(m)}. \tag{15}$$

For unmatched actual modes, the estimated value is taken to be zero, so each per-component error is exactly 1 by construction. Clipping matched modes at 1 puts all per-component errors on the same $[0, 1]$ scale as unmatched ones and prevents a single outlier from dominating the mean. The combined relative error is then the starting metric:

$$\mathrm{RE}_0 = \tfrac{1}{3}\left(\mathrm{RE}_\Omega + \mathrm{RE}_\sigma + \mathrm{RE}_b\right) \in [0, 1]. \tag{16}$$

Contrary to Task A, the total number of modes M is *unknown*, so if $\tilde{M}$ is the identified count it will generally be true that $\Delta M := |M - \tilde{M}| \geq 0$, and the metric is augmented as

$$\mathrm{RE} := \mathrm{RE}_0 + \lambda \min\left(1,\ \frac{\Delta M}{M}\right), \tag{17}$$

with $\lambda = 1$ by default. Normalizing ΔM by M keeps the mode-count mismatch on the same scale as RE_0; unnormalized, it reaches thousands in dense scenarios and would dominate the ranking. The final metric therefore lies in $[0, 2]$, and the raw ΔM is still reported per file for diagnostics.

5. SUBMISSIONS

This section summarizes all submissions received by the organizers. In total, 13 participants—individuals or teams—proposed 22 methods, 12 for Task A and 10 for Task B. Every entrant was required to supply a short technical report describing the method (available at the repository[1]), alongside code and result files. Participants were also encouraged to submit a short paper that will appear in the conference proceedings to describe the method in more detail (see [10–27]).

Table 2 summarizes the submissions, citing the relevant short paper – when available – or the technical report. Submissions are identified by a label p-Xn, where p is the participant index, $X \in \{\mathrm{A}, \mathrm{B}\}$ the task, and the optional n distinguishes multiple entries by the same participant to the same task.

As can be seen, supervised methods are pervasive, but often coupled with refinement stages such as the meta-heuristics Particle Swarm Optimization (PSO), Latin Hypercube Sampling (LHS) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES), or optimization based on Differentiable Digital Signal Processing (DDSP) [9]: many participants sought to reduce search time by starting from a first guess obtained through training. Others worked only with iterative methods but, where these require IR resynthesis, kept the compute time low, for instance by using a short but informative IR, reducing the model to a parallel filterbank, or holding some parameters fixed.

Inference times consequently vary greatly, from a few milliseconds for a GPU forward pass to thousands of seconds for iterative methods dominated by IR synthesis. They are approximate indicators rather than benchmarks, since each participant used their own hardware and many methods admit further optimization. Although all participants supplied code, we deliberately did not re-run it on common hardware: doing so would have required porting a dozen heterogeneous software stacks—MATLAB and several deep learning frameworks with incompatible GPU requirements—to one machine, and the resulting figures would still have depended strongly on the effort invested in each port. Since the challenge neither ranked on runtime nor incentivized its reduction, we judged that a like-for-like comparison would convey a false sense of precision.

6. RESULTS

Final rankings for Task A and Task B are reported in Tables 3 and 4. This ranking is not intended to establish the overall superiority or the intrinsic value of any particular method, as each approach exhibits distinct advantages and limitations. It reports only the errors observed on the target dataset.

6.1. Task A

The NMSE values in Task A span twelve orders of magnitude. Two methods reach machine precision: 13-A2 and 10-A attain mean NMSE of order 10^{-13} and 10^{-11}, far below any practically relevant tolerance. Both, however, are expensive in an information-theoretic sense: one exploits a very large dataset, the other evaluates the forward model many times.

All QMUL submissions were trained on 327 680 IRs. Since this cost is paid once, at training time, their inference times are

[1] https://github.com/LOGUNIVPM/1st-DAFx-Challenge/tree/main/TechReports

Table 2: Received submissions, sorted by task and then participant index. "FF" is a single feed-forward pass; "Params" counts trainable parameters for learned methods, "–" means none, "n/r" not reported. †Head only: 4-A adds a pretrained Wavegram-CNN14 backbone, 5-A a neural spline flow. ‡Counts for 7-A, 7-B, 12-B1, 12-B2 come from the submitted checkpoints, 1-A and 5-A from the submitted code.

ID	First author (team)	Ref.	Dataset	Params	# iterations	Time [s]	Method comments
Task A — physical parameter identification							
1-A	Yang	[10]	1 000	0.72 M	FF	8.48 (GPU)	Supervised NN (time-domain CNN-GRU)
2-A	Lu	[11]	1 000	–	FF	0.94 (CPU)	Supervised tree ensemble (ExtraTrees, HistGB)
4-A	Guo (WHU-IASP)	[12]	10 000	1.2 M†	FF	12.08 (GPU)	Pretrained Wavegram-CNN14 and hand-crafted features into an MLP
5-A	Sechet (LISN)	[13]	60 000	1.8 M†	FF + 3 × 10 000	0.05 + 9 000	CNN summary network and neural spline flow
6-A	Garofalo (POLIMI)	[14]	–	–	500 × 100 (max) + 1 500	780 (CPU+GPU)	Iterative DDSP resonator bank, 500 candidates
7-A	Heminway	[15]	20 000	0.54 M	FF	0.2 (CPU)	Supervised 1D CNN
8-A	Lipkin (SARG)	[16]	50 000	11.7 M	FF	0.006 (CPU)	Supervised NN (modified ResNet-18)
9-A	Lee (MIT-CCRMA)	[17]	–	–	753 (avg)	190 (CPU+GPU)	Three-stage optimization, LHS, coarse-to-fine
10-A	Park & Yi (Illinois)	[18]	–	–	13 368 (avg)	114.74 (CPU+GPU)	CMA-ES plus ternary refinement stage
11-A	Kim (KAIST)	[19]	5 000	0.91 M	FF	0.087 (GPU)	Dual-branch CNN (log-Mel and dilated waveform)
13-A1	Marttila (QMUL)	[20]	327 680	2.9 M	FF + 100 (DDSP)	300	Supervised AST plus DDSP plate-model refinement
13-A2	Marttila (QMUL)	[20]	327 680	6.5 M	FF + 100 (PSO)	180	Supervised AST, diffusion transformer, PSO refinement
Task B — modal parameter estimation							
2-B	Lu	[21]	600	–	FF + 80	6.60 (CPU)	Per-band count (ExtraTrees), DDSP refinement
3-B	Bittner (CDLEML)	[22]	128	–	FF	157 (CPU)	Isotonic count, Matrix Pencil, state-space gain fit
4-B	Zhang (WHU-IASP)	[23]	10 000	n/r‡	FF	0.47 (CPU+GPU)	DMD and AST fusion, Conv1D for σ, gain, activity
7-B	Heminway	[15]	10 000	0.11 M	30 epochs	70–120 (CPU)	`modalfit` below 1 kHz, 2D CNN above
8-B	Lipkin (SARG)	[24]	–	–	0	165 (CPU)	Hankel DMD, gains from the DTFT
9-B	Franchino (MIT-CCRMA)	[25]	–	–	8 × MP	1 350 (CPU)	Matching pursuit, AR pooling, ridge-LS gain refit
12-B1	Jung	[26]	240	5 058	20 (avg)	373.38	Peak picking, residual expansion, MLP calibration
12-B2	Jung	[26]	240	5 058	20 (avg)	373.43	As 12-B1 plus high-band density fill
13-B1	Marttila (QMUL)	[27]	327 680	3.1 M	FF	0.001	Supervised count-density U-Net
13-B2	Marttila (QMUL)	[27]	327 680	n/r‡	FF	0.002	Supervised complex CNN and transformer

among the shortest in the challenge. For scale, a uniform grid with 8 points per dimension over the six free parameters would contain 262 144 points, fewer than the training set. What the numbers show is that the network has learned to interpolate a *deterministic* simulator inside the region of parameter space on which it was trained. Since the test targets came from the same generator and ranges as the training data, the challenge cannot distinguish this from a genuinely general inverse solver; the untested directions are the interesting ones, and we return to them in Sec. 7.1.

Other supervised methods did not reach the same precision despite a similar strategy (7-A, 1-A, 8-A); whether the difference stems from architecture or from dataset size, which exceeds the next-largest training set fivefold, warrants further research. The winning method also adds a short PSO refinement of 100 iterations on top of the network. PSO alone does not explain the result, since the PSO baseline runs ten times as many evaluations and still ranks 12th, but a local search on a good initial guess is evidently useful; an ablation study would settle the question.

Among the iterative entries, 10-A searches with CMA-ES over at most 30 000 evaluations, 13 368 on average, yet runs faster than several other methods because its synthesised IR is short enough to keep the cost per evaluation low yet long enough to stay informative. Method 9-A uses divide-and-conquer to reach a remarkably low error in 753 iterations, and 5-A a similar one from a supervised first guess plus 30 000 fine-tuning iterations that dominate its runtime, the highest of the challenge. Finally, 13-A1 ranks below 13-A2, and 6-A 11th, because of isolated badly estimated IRs; on the median index they would rank third and fifth, showing that DDSP methods can solve this problem.

Table 3: Final ranking for Task A, sorted by $E\{\text{NMSE}\}$; lower is better. Bold marks the best value in each column. CP: constant predictor. Read $a\text{e}{-b}$ as $a \times 10^{-b}$.

#	Author	$E\{\text{NMSE}\}$	$M\{\text{NMSE}\}$	$E\{\text{SMSE}\}$	$M\{\text{SMSE}\}$
1	Marttila (13-A2)	**1.785e−13**	**3.717e−15**	**2.987e−10**	**8.531e−11**
2	Park & Yi (10-A)	1.933e−11	3.103e−14	6.836e−10	3.738e−10
3	Lee (9-A)	1.007e−4	5.906e−6	0.014	6.096e−4
4	Sechet (5-A)	4.689e−4	2.962e−5	0.384	0.366
5	Marttila (13-A1)	1.133e−3	2.285e−12	8.780e−4	5.024e−8
6	Heminway (7-A)	2.885e−3	0.003	0.487	0.472
7	Guo (4-A)	4.836e−3	0.003	0.616	0.473
8	Lipkin (8-A)	6.476e−3	0.003	0.674	0.515
9	Lu (2-A)	0.010	0.006	0.558	0.550
10	Kim (11-A)	0.013	0.008	0.571	0.486
11	Garofalo (6-A)	0.018	6.062e−6	0.121	0.062
12	Baseline PSO	0.048	0.027	0.386	0.354
13	Yang (1-A)	0.056	0.054	1.137	0.561
14	Baseline CP	0.120	0.121	2.105	1.293

6.2. Task B

The picture emerging from Task B is markedly different. No method reaches machine precision, and even the best stops at $E\{\text{RE}\} \approx 0.33$, about a sixth of the metric's dynamic range. The gap between the top two entries and the rest is wide: 13-B1 and 13-B2, both from QMUL, score 0.328 and 0.337, while the third-ranked 2-B sits at 0.539, and Both QMUL entries share one formulation: rather than picking spectral peaks and estimating triples one by one, they jointly infer a per-bin mode density on a $1/20$-octave log-frequency grid together with the corresponding decay rates and gains, differing only in the backbone—a one-dimensional U-Net for 13-B1, a complex-valued transformer on the real and imaginary parts of the FFT for 13-B2. That two very different architectures score almost identically suggests the formulation and the training set drive the result, not the architecture.

The component-wise breakdown in Table 4 reveals the most important scientific outcome of Task B: no method has yet learned to recover individual modal gains. Method 13-B1 achieves $E\{\mathrm{RE}_\Omega\} = 0.013$ on modal frequencies—essentially perfect—and $E\{\mathrm{RE}_\sigma\} = 0.058$ on decays, but $E\{\mathrm{RE}_b\} = 0.847$ on gains, and this is no peculiarity of the winner: every submission scores $\mathrm{RE}_b > 0.83$. Two mechanisms plausibly contribute. First, as noted by team 9-B, in dense modal regions many modes fall within a fraction of a hertz of each other and their damped-sinusoid atoms become almost linearly dependent, so a global least-squares fit recovers the joint amplitude of a cluster but not the individual gains, regardless of regularisation. Second, RE_b normalises by $b_m^{(\mathrm{ref})}$ and clips at 1, penalising heavily any method that misses small gains while saturating for large errors.

A physical caveat governs how all of these numbers should be read. The half-power bandwidth of a mode is σ_m/π and grows quadratically with frequency through $\sigma_m = \eta_0 + \eta_1\Omega_m^2$, while the mean spacing between neighbouring modes falls with frequency, so the modal overlap $\mu = (\sigma_m/\pi)\,/\,\overline{\Delta f}$ crosses unity somewhere inside the audio band. Across the 16 test plates that crossover lies between 252 Hz and 1434 Hz, with a corpus mean near 1 kHz. Below it, modes are separated by more than their own bandwidths and identification is well posed; above it, neighbouring modes merge and no unique set of triples can be recovered from the response. Only 9.5% of the ground-truth modes lie below 1 kHz, however, so roughly nine tenths of the weight in (17), which averages over all M modes equally, falls in the regime where the inverse problem has no unique solution—and there the half-octave matching threshold rewards predicting the correct modal *density* rather than locating individual poles.

To probe how sensitive the ranking is to the gain metric, we recomputed all scores under two alternatives. A logarithmic per-mode gain error drops the winning RE_b from 0.85 to 0.55 but leaves the ranking essentially unchanged. A genuinely different picture emerges only if per-mode scoring is abandoned and $\log|H(\omega)|$, reconstructed from each submission's triples through (9), is compared on a dense log-frequency grid across [20 Hz, 10 kHz] (Fig. 2 and the right-hand block of Table 4). No method then dominates across the audio band: 7-B and 3-B win where modes are sparse, 9-B is nearly perfect above 1 kHz at the cost of a large low-frequency overshoot, and the aggregate winner is 2-B, mid-field in every band and top by consistency alone. 13-B2, second officially, falls to *last*: its gains are almost uniformly 1.6 orders of magnitude smaller than 13-B1's (median $\log_{10}|b_m|$ near -12.6 against -11.0), a systematic $\sim$30 dB underestimate that the saturating per-mode RE_b cannot see but the spectral metric registers at every bin. Future benchmarks should therefore complement the per-mode formulation with a frequency-domain error on $|H(\omega)|$, closer to what a listener integrates and sensitive to both bulk-gain bias and band-dependent strengths that (16) averages away.

The middle of the field, summarised in Table 2, converges within a factor of two despite resting on very different principles, and all of it remains a factor of two above the top two. This is the sense in which the metric, rather than the physics, favours the density-regression formulation of the QMUL entries: neither of them identifies individual poles, and 2-B fixes its frequencies at initialisation and refines only decay and gain, yet both score well because a correct density suffices above the overlap limit. The sub-band analysis places the genuine identifiers differently, with 7-B and 3-B leading below 1 kHz where identification is well posed and 9-B above it. At the bottom, 8-B applies forward-backward Hankel DMD directly to the scalar IR; the method is elegant, but at the model order used it finds on average only $\tilde{M} \approx 1\,100$ modes against a ground-truth average of $\bar{M} \approx 5\,300$, and the $\Delta M/M$ penalty dominates its score. The baseline sits last on both metrics, confirming that in the dense modal regime a naive spectral-peak strategy is essentially uninformative, and *every* submitted method beats it.

The ranking also ties closely to the estimated mode count: $E\{\mathrm{RE}\}$ and $E\{\Delta M\}$ rise in near-lockstep, the leading methods sitting at $E\{\Delta M\} \approx 130$ against $\bar{M} = 5\,303$ while 8-B misses by a factor of five; the only significant inversion is 3-B, whose Matrix Pencil stage recovers the right count but at the cost of $E\{\mathrm{RE}_\sigma\} = 0.76$. Getting the number of modes right is at least as important as getting each individual mode right.

7. DISCUSSION

No single solution obtains both the best accuracy and the best inference time, and none achieves the goal with a small amount of information: the best performers need either a large dataset or a large number of evaluations. Whether comparable performance is reachable with less, for instance by exploiting a differentiable physical model as a prior, remains open.

DDSP methods were used to refine a first guess (13-A1, 2-B) or as the only solution (6-A). Such methods have identified audio circuits successfully [28] and proved worthwhile here, but did not reach excellence. One explanation worth investigating is that they have a low *output/error information ratio* in this setting: a plate reverb requires an entire IR to be generated to yield one scalar error, and the unit-impulse input conveys no information of its own. Convolving the plate IR with an audio signal and identifying the system online would help, but the modal model of Sec. 2 is not readily adapted online.

7.1. Generalization

The dataset is generated from a linear plate model, deliberately: it admits a closed-form modal decomposition and an unambiguous ground truth. The consequences differ by task. Task A methods estimate the parameters of a specific linear PDE, so retargeting them would require regenerating the training dataset for supervised methods, or rewriting the forward model for DDSP and iterative ones. Task B methods estimate the modal triples of a biquad filterbank, the standard reduced-order representation of a much wider class of linear and time-invariant modal systems, so as long as the target IR admits an approximate modal-sum decomposition in the audible band, the same pipelines apply in principle to other geometries, boundary conditions, or measured plate IRs.

Real plates exhibit weak nonlinear behaviour, both geometric and dissipative, as well as air-loading and coupling effects not captured by (1). If the target came from a real or nonlinear plate, the Task A ground truth in its current parameterisation would no longer exist and the task would need redefining. Task B would remain well posed—a listener-facing modal reverb is a biquad bank whatever the underlying physics—but the QMUL entries would likely need retraining, whereas the classical and hybrid methods (3-B, 7-B, 9-B) would carry over almost unchanged, assuming nothing about the source of the IR beyond its being well approximated by a bank of resonances. That family, currently mid-table, may therefore be the more portable technology, while the QMUL entries trade portability for accuracy on the distribution they were trained on.

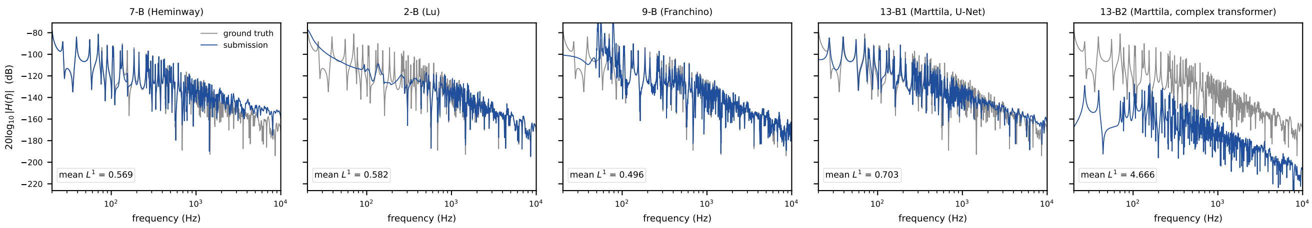

Figure 2: Reconstructed frequency response $|H(f)|$ from five representative Task B submissions on test file 8 ($M = 1\,189$ modes), overlaid on the ground truth (grey). Each panel reports the mean L^1 error on $\log|H(f)|$ over [20 Hz, 10 kHz]. Methods 7-B (Heminway) and 2-B (Lu) track the ground-truth envelope tightly; 9-B (Franchino) is nearly exact above ∼1 kHz but overshoots at low frequencies; 13-B1 (Marttila, U-Net) resolves individual low-frequency peaks best of all; 13-B2 (Marttila, complex transformer) reproduces the correct spectral shape but is uniformly ∼30 dB below the ground truth, an offset invisible to the per-mode metric of Eq. (16).

Table 4: Final ranking for Task B, sorted by the official per-mode metric $E\{\mathrm{RE}\}$ of Eq. (17); lower is better, averaged over the 16 test IRs. The right-hand block is the complementary frequency-domain evaluation: mean L^1 error on $\log|H(f)|$ (natural log units; divide by $\log 10$, multiply by 20 for dB) between the ground truth and the response reconstructed from the submitted triples via (9) on a 4 096-point log-spaced grid. Parenthesised: rank under that metric. Bold marks the best value per column.

		Per-mode metric, Eq. (17)					Mean L^1 error on $\log\|H(f)\|$				
ID	First author	$E\{\mathrm{RE}\}$	$E\{\mathrm{RE}_\Omega\}$	$E\{\mathrm{RE}_\sigma\}$	$E\{\mathrm{RE}_b\}$	$E\{\Delta M\}$	all (rank)	[20, 200] Hz	[0.2, 1] kHz	[1, 3] kHz	[3, 10] kHz
13-B1	Marttila	**0.328**	**0.013**	0.058	0.847	135	0.882 (3)	0.790	0.798	0.752	1.291
13-B2	Marttila	0.337	0.016	**0.023**	0.925	**125**	2.983 (10)	2.779	2.793	3.239	3.396
2-B	Lu	0.539	0.025	0.569	**0.838**	278	**0.625** (1)	0.910	0.687	0.284	0.307
7-B	Heminway	0.643	0.201	0.268	0.868	1738	0.651 (2)	**0.346**	**0.264**	0.878	1.541
3-B	Bittner	0.743	0.249	0.760	0.970	509	1.561 (7)	**0.346**	0.738	2.323	4.286
4-B	Zhang	0.862	0.272	0.499	0.913	2108	1.905 (9)	2.272	1.629	1.395	2.036
9-B	Franchino	0.957	0.324	0.800	0.937	1546	1.646 (8)	3.305	1.551	**0.030**	**0.077**
12-B2	Jung	1.066	0.184	0.530	0.881	2725	1.125 (5)	1.632	0.753	0.694	1.045
12-B1	Jung	1.102	0.349	0.665	0.906	2853	1.051 (4)	1.632	0.753	0.694	0.663
8-B	Cogliati	1.485	0.650	0.935	0.973	4166	1.160 (6)	1.174	0.969	0.827	1.694
–	Baseline	1.968	0.983	0.985	0.985	5242	2.463 (11)	1.581	1.007	4.309	4.412

This asymmetry is a property of the benchmark as much as of the methods, and it is the clearest lesson of this first edition. Because every test target was drawn from the same generator and the same ranges as the data participants were free to synthesise, dense sampling of the training distribution was not merely permissible but optimal, and the evaluation offers no way to tell such a strategy apart from a genuinely general inverse solver. A future edition should make that distinction measurable, and we intend to reserve part of the test set for conditions no participant can sample in advance: parameters outside the published ranges, measurement noise, free rather than simply supported edges, and where possible a physical plate, scored separately from the in-distribution partition.

8. CONCLUSIONS

The 1st DAFx Parameter Estimation Challenge has shown that plate parameter identification from audio data is now tractable. The top-ranked Task A submission combines both available families: a neural network learns the inverse of the forward plate model to extreme precision given sufficient training data, and a short PSO search further improves the estimate. Purely iterative methods such as CMA-ES score on par when the budget allows, their strength being robustness: they assume nothing about parameter distributions and need no training data. Task B remains open: none of the methods recovers individual modal gains, and the frequency-domain evaluation reorders the ranking substantially. Both observations point to the metric as much as to the methods, and future editions should score modal estimates in the frequency domain and weight the resolvable and unresolvable regimes separately.

Several avenues merit investigation. Embedding the plate model as a differentiable layer, via DDSP or neural ODEs, may improve sample efficiency and provide better priors than generic architectures. Estimation error is measured in normalized units, whereas listeners are insensitive to many small parameter variations, so subjective studies of perceptual thresholds could reduce the precision actually required.

9. ACKNOWLEDGMENTS

The authors would like to acknowledge the participants for their enthusiasm and for their kindness in sharing results and details. 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.

10. REFERENCES

[1] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[2] S. Bilbao, K. Arcas, and A. Chaigne, "A physical model for plate reverberation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, 2006, pp. 165–168.

[3] K. Arcas and A. Chaigne, "On the quality of plate reverberation," *Applied Acoustics*, vol. 71, no. 2, pp. 147–156, 2010.

[4] M. Ducceschi and L. Gabrielli, "The 1st DAFx parameter estimation challenge: A benchmark for plate reverb system identification," 2025. [Online]. Available: https://github.com/LOGUNIVPM/1st-DAFx-Challenge/blob/main/DAFxChallengeDetails.pdf

[5] S. Bilbao, "A digital plate reverberation algorithm," *Journal of the Audio Engineering Society*, vol. 55, no. 3, pp. 135–144, March 2007.

[6] J. O. Smith, *Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects*. W3K Publishing, 2010.

[7] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. Chichester, UK: John Wiley & Sons, 2009.

[8] R. Yamamoto, E. Song, and J.-M. Kim, "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6199–6203.

[9] J. Engel, C. Gu, A. Roberts *et al.*, "DDSP: Differentiable digital signal processing," in *International Conference on Learning Representations*, 2020.

[10] J.-C. Yang, "Neural networks for physical parameter estimation of plate reverberation from impulse responses," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[11] M. Lu and J. D. Reiss, "Simulation-based plate-reverb parameter estimation from a single impulse response," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[12] Z. Guo, Y. Zhang, L. Chen, W. Liu, and G. Huang, "Physics-inspired feature fusion for plate parameter estimation from acoustic impulse responses," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[13] D. Sechet, M. Evrard, and M. Kowalski, "Simulation-based inference for plate reverb system identification," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[14] F. Garofalo, A. A. Lillo, A. Ilic Mezza, R. Giampiccolo, and A. Bernardini, "Parameter estimation via differentiable modal plate synthesis," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[15] R. Heminway, "A convolutional approach to plate reverberator parameter estimation," 1st DAFx Parameter Estimation Challenge, Tech. Rep., 2026. [Online]. Available: https://github.com/LOGUNIVPM/1st-DAFx-Challenge/tree/main/TechReports

[16] D. A. Anderson, J. Lipkin, M. Chen, B. R. Thompson, A. Cogliati, and M. C. Heilemann, "A multi-resolution spectrogram approach for estimating the physical parameters of a plate reverb," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[17] J. W. Lee, J. Chowdhury, F. Franchino, M. Rau, and S. Kim, "ALAMODE: Automated learning of acoustical modal parameters via differential evolution," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[18] B. Park, J. Yi, T. Kim, and M. Kim, "Accurate plate reverb parameter estimation using two-stage evolutionary search," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[19] K. Kim and S. Kim, "A dual-branch network for plate reverb physical parameter estimation," 1st DAFx Parameter Estimation Challenge, Tech. Rep., 2026. [Online]. Available: https://github.com/LOGUNIVPM/1st-DAFx-Challenge/tree/main/TechReports

[20] D. Marttila, R. Diaz, P. Tablas De Paula, I. Ibnyahya, and C.-Y. Yu, "Transformer-based plate parameter estimation with differentiable and particle-swarm refinement," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[21] M. Lu and J. D. Reiss, "Band-count dense modal estimation with fixed-frequency differentiable resonator refinement," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[22] M. Bittner and A. Jantsch, "Non-iterative modal parameter estimation for plate reverbs via matrix-pencil-guided state space model initialization," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[23] Y. Zhang, L. Chen, Z. Guo, W. Liu, and G. Huang, "A dual-stream framework combining audio spectrogram transformer and dynamic mode decomposition for plate modal parameter estimation," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[24] J. Lipkin, M. Chen, D. A. Anderson, B. R. Thompson, M. C. Heilemann, and A. Cogliati, "Data-driven modal parameter estimation for plate reverberation for task B via Hankel dynamic mode decomposition," 1st DAFx Parameter Estimation Challenge, Tech. Rep., 2026. [Online]. Available: https://github.com/LOGUNIVPM/1st-DAFx-Challenge/tree/main/TechReports

[25] F. Franchino, J. Chowdhury, J. W. Lee, M. Rau, and S. Kim, "Multi-view subband autoregressive pole harvesting for modal plate identification," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[26] D. Jung, "Peak-residual modal estimation with learned calibration and high-band density correction," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[27] R. Diaz, P. Tablas De Paula, D. Marttila, I. Ibnyahya, and C.-Y. Yu, "Count-density networks for modal plate parameter estimation," in *Proc. 29th Int. Conf. Digital Audio Effects (DAFx26)*, Cambridge, USA, 2026.

[28] F. Esqueda, B. Kuznetsov, and J. D. Parker, "Differentiable white-box virtual analog modeling," in *Proc. 24th Int. Conf. Digital Audio Effects (DAFx)*, 2021, pp. 41–48.