

ACCURATE PLATE REVERB PARAMETER ESTIMATION USING TWO-STAGE EVOLUTIONARY SEARCH

Byunghoo Park* Jayeon Yi* Takyoun Kim Minje Kim

Siebel School of Computing and Data Science
University of Illinois Urbana-Champaign
Urbana, USA

{bhpark3, jayeonyi, tk30, minje}@illinois.edu

ABSTRACT

We describe our submission to Task A of the 1st DAFx parameter estimation challenge. The task is to recover the six physical parameters of a simulated metal-plate reverberator — its dimensions and material properties — from a single impulse response (IR). We treat this as a black-box optimization: candidate parameter sets are fed to the simulator and scored by a loss against the target IR. The method has two stages. The first uses CMA-ES, an evolutionary optimizer, to recover five of the six parameters, comparing IRs under an amplitude-normalized loss. Amplitude normalization makes the search robust but discards the cue to the sixth parameter, the plate’s surface density; a second stage therefore estimates it alone, with a ternary search on the un-normalized loss. As the choice of loss strongly affects the search, we select it beforehand, and analyze why compression in the common multi-scale spectral loss degrades recovery. Finally, we test our method on a validation set of 50 IRs, discuss a pathological failure mode, and ablate to justify having two different stages instead of a unified CMA-ES search.

1. INTRODUCTION

A plate reverberator [1] produces artificial reverberation [2] by picking up the vibration of a driven metal plate; such units are now commonly emulated in software. Given a simulated plate IR, one may wish to recover the physical parameters that generated it, both to learn useful configurations and as a step towards broader creative applications.

The 1st DAFx parameter estimation challenge [3] benchmarks this problem. It consists of two tasks: Task A, “physical parameter identification”, is to infer the physical parameters of the plate from a simulated IR, while Task B, “modal parameter estimation”, is to dissect the modes of vibration. In this work, we describe our submission to Task A.

Existing methods do not transfer readily. Xu and Ding [4] adapt the Newton method to estimate coefficients of linear time-invariant systems, but it has been demonstrated only on simple third-order systems and does not readily scale to a problem of this size. Lee *et al.* [5] optimize a differentiable room reverberator via gradient descent; our preliminary attempts to adapt this approach were unstable and prone to divergence, likely reflecting a rugged,

ill-conditioned loss landscape. We therefore adopt the covariance matrix adaptation evolution strategy (CMA-ES) [6], a derivative-free optimizer suited to exactly such problems. CMA-ES has also shown promise in similar parameter-estimation settings, such as articulatory speech synthesis [7].

2. BACKGROUND

2.1. Plate Reverb Simulator, and Task A of the Challenge

The plate simulator [3] implements the general solution of the damped Kirchhoff–Love equation [1, 8], governed by the set of physical parameters

$$P := \{\rho, E, \nu, L_x, L_y, h, T_0, \eta_0, \eta_1, x_i, y_i, x_o, y_o\}, \quad (1)$$

which are, in order: the volumetric density (ρ), Young’s modulus (E), Poisson’s ratio (ν), the width (L_x), the height (L_y), the thickness (h), the tension applied via connections to the chassis (T_0), decay constants (η_0, η_1), the input location (x_i, y_i), and the output location (x_o, y_o). We refer the reader to the challenge documentation [3] for the full model. For our purposes, a relevant property is that the simulator output is a sum of second-order infinite impulse response (IIR) filters whose coefficients are functions of the physical parameters. We therefore estimate the parameters by optimizing P directly, searching for the parameter set whose rendered IR best matches the target.

Task A of the challenge is to estimate $P_{\text{true}} := \{x_o, y_o, L_y, \mu, T_0/\mu, D/\mu\}$, a derived subset of P associated with a given IR, where $\mu := \rho h$ is the surface density and $D := Eh^3/12(1 - \nu^2)$ is the flexural rigidity. Other parameters ($\nu, L_x, \eta_0, \eta_1, x_i, y_i$) remain fixed. This task design allows for a one-to-one correspondence between the IR and the six parameters, and thus a fair evaluation via normalized mean square error (NMSE):

$$\text{NMSE}(P_{\text{true}}, P_{\text{est}}) = \frac{1}{6} \sum_{i=1}^6 \left(\frac{P_{\text{true},i} - P_{\text{est},i}}{M_i - m_i} \right)^2 \quad (2)$$

where M_i and m_i are the upper and lower bounds of the values each parameter may assume, while P_{true} and P_{est} are the 6-dimensional vectors of true and estimated parameters. The bounds are derived from the legal ranges of the raw quantities $E, \rho, h, L_y, T_0, x_o$, and y_o .

2.2. Covariance Matrix Adaptation Evolution Strategy

Covariance matrix adaptation evolution strategy (CMA-ES) [6] is a derivative-free optimizer for continuous black-box objectives. At each generation, it samples a population of candidate solutions

*These authors contributed equally to this work.

from a multivariate Gaussian, evaluates them, and adapts the distribution: the mean is moved towards the rank-weighted best candidates, while the covariance matrix and a global step size are updated to align the search with the directions in which good candidates have consistently been found [6].

CMA-ES is designed for ill-conditioned, non-separable problems with many local optima (i.e. “multimodal”) [6]. Its covariance matrix learns the local geometry of the objective, providing implicit preconditioning of the search space, while its rank-based selection yields invariance under monotonic transformations of the objective. These properties suit the present problem: the inverse mapping from IR to parameters is also expected to be ill-conditioned and multimodal, the regime CMA-ES is built for. We note that, although the simulator is at least piecewise differentiable, direct gradient descent on it tended to diverge, and proved difficult to stabilize through hyperparameter tuning.

3. METHODOLOGY

Our method estimates the six Task A parameters over two stages: a CMA-ES search over the seven raw physical parameters, followed by a dedicated refinement of the mass density μ .

To meet the runtime requirements set in place by the challenge [3], we truncated all IRs and IR generations to the first 0.25s, while retaining the sample rate of 44.1 kHz. Moreover, all IR syntheses during the CMA-ES searches were run on a Quadro RTX 6000 GPU using a PyTorch port of the plate.

Both stages minimize a common objective: an L_1 distance on STFT magnitudes between the target IR and the one rendered by the plate simulator. The STFT uses a 4096-point Hann window with a hop of 1024. Frames are centered on the signal samples, with reflection padding applied at both boundaries. Contrary to the baseline solution [3], we do not log-compress the magnitudes nor use multiple resolutions, as these were either found to deter convergence or be of less help. See Sec. 4 for an analysis.

3.1. Stage 1: CMA-ES Search

We found CMA-ES searches over a seven-dimensional raw-parameter space of $E, \rho, h, L_y, T_0, x_o$, and y_o to recover all Task-A parameters well, with the exception of μ . Therefore, in Stage 1, we launch multiple *restarts* of said search, and upon the first success compute and fix the five parameters $x_o, y_o, L_y, T_0/\mu$, and D/μ using the outputted seven-parameter estimates.

A single *restart* runs CMA-ES (Sec. 2.2), updating a population of seven-element vectors over successive generations until one of four conditions is met: a budget of 30,000 loss function evaluations, a tolerance of 1×10^{-5} on the range of recent best-loss values, or a per-generation or per-restart wall-clock limit (300 and 600 seconds, respectively). Each candidate is rendered by the simulator and compared to the target. Both IRs are peak-normalized before a loss can be derived; this makes the loss comparable across different IRs and accelerates optimization (Sec. 5).

The restarts are launched sequentially until one succeeds. Every restart draws its initial distribution mean by Latin hypercube sampling (LHS) in a normalized $[-1, 1]^7$ cube, to ensure diversity across restarts. CMA-ES then operates in this normalized space with initial step size $\sigma_0 = 0.6$, mapping affinely onto the physical bounds of $E, \rho, h, L_y, T_0, x_o, y_o$ to render a candidate. To reduce runtime across the entire search all restarts, Optuna [9]

Table 1: Component decomposition from L_1 -STFT to the MSS+SC variant (components defined in text); the challenge MSS baseline is the *log + multi* row, and *log (bounded)* is the variant $\log(1+x)$. Values are geometric-mean NMSE and convergence rate (NMSE < 0.02) over the 50 IRs.

Configuration	Geomean NMSE	Conv. rate
L_1 -STFT (base)	4.9×10^{-14}	1.00
+ multi	4.8×10^{-14}	1.00
+ SC	1.0×10^{-13}	0.98
+ SC + multi	3.2×10^{-14}	1.00
+ log (bounded)	8.9×10^{-4}	0.40
+ log	1.3×10^{-1}	0.00
+ log + multi (MSS)	1.3×10^{-1}	0.02
+ log + SC	7.0×10^{-3}	0.48
MSS+SC (log + SC + multi)	2.7×10^{-4}	0.68

Table 2: Effect of STFT window size on recovery, for the single-resolution uncompressed L_1 -STFT loss. Recovery saturates by 4096 points (bold, our choice); every window still outperforms the compressed losses of Table 1.

Window (pts)	Geomean NMSE	Conv. rate
512	5.9×10^{-12}	0.92
1024	3.7×10^{-12}	0.88
2048	1.8×10^{-13}	0.96
4096	4.9×10^{-14}	1.00
8192	9.2×10^{-14}	0.96

prunes unpromising restarts via a SuccessiveHalvingPruner (reduction factor 3, min resource 15) and selects each restart’s population size in [30, 60], as we had observed large population sizes to be sometimes detrimental to convergence. As the ground truth is unavailable at inference time, the search is configured to halt once any completed restart reaches a loss below the empirically chosen threshold of 0.01, or after 400 unsuccessful restarts.

We chose not to pursue a direct search over the six-dimensional Task-A space, as preliminary experiments showed greatly reduced convergence. A good out-of-bounds handler would be necessary to leverage the complex bounds derived from the seven raw parameters (Sec. 2.1). This question is left for future research.

3.2. Stage 2: Surface-Density Refinement

The peak normalization employed in Stage 1 removes the absolute amplitude, and with it the only cue to μ . Thus, Stage 1 recovers μ poorly. Meanwhile, with the ratios D/μ and T_0/μ and the plate geometry held fixed, we found μ to be monotonic to the overall amplitude of the response.

Therefore we estimate μ with a separate 50-iteration ternary search over μ , holding the other five parameters at their Stage 1 estimates and minimizing the same objective without any IR normalization. The search spans $\mu \in [2.43, 106.15]$, the range induced by the bounds on ρ and h .

4. LOSS ANALYSIS

The selected L_1 -on-STFT-magnitude loss, under our setting, outperforms the challenge’s multi-scale spectral loss (MSS) baseline [10], which is a log-magnitude L_1 loss summed over multiple STFT resolutions. In this section, we analyze this gap.

Two components separate our loss from this baseline: log-magnitude compression (*log*) and multi-scale averaging (*multi*). Because spectral-convergence terms are common in such losses, we additionally test a spectral-convergence term (*SC*) [11] — a normalized magnitude error that emphasizes high-energy regions — as an augmentation; adding it yields the MSS+SC variant [12]. We toggle each component in isolation, holding the STFT framing fixed.

The comparisons are run on a set of 50 generated IRs. Because the candidate losses span very different numerical ranges, the CMA-ES termination tolerance and early-stopping threshold were set relative to each loss’s own range, rather than the fixed values used in the main pipeline (Sec. 3.1), so that all candidates are compared under comparable settings; the selected L_1 -STFT loss was re-run under this setting, so its figures here differ from those of Sec. 5. Table 1 reports the geometric-mean NMSE and the convergence rate (the fraction of IRs recovered to $\text{NMSE} < 0.02$) for each configuration.

Two observations follow. First, *log* compression is responsible for the gap. On its own it reduces convergence to none and raises the geometric-mean NMSE from roughly 10^{-14} to 10^{-1} ; the challenge MSS baseline, which adds *multi* to *log*, inherits this collapse almost entirely, recovering only 2% of IRs. The *SC* and *multi* components are individually harmless, each leaving recovery at the level of the base loss; their only effect is to partly offset the damage from *log*. Adding *SC* to the baseline — the MSS+SC variant — raises convergence from 0.02 to 0.68 (geomean 2.7×10^{-4}), still far short of the uncompressed L_1 -STFT.

Second, the effect is specific to the “unbounded” $\log(x)$ form. Schwär and Müller [13] identify a related deficiency in harmonic estimation and propose the bounded form $\log(1+x)$ together with a full configuration (Smooth MSS). We find both less damaging than plain $\log$ — the bounded form reaches geomean 8.9×10^{-4} and their full configuration geomean 7.9×10^{-11} — yet neither matches the uncompressed loss.

Window size only has a secondary effect. Table 2 varies the single-resolution window from 512 to 8192 points: recovery is weakest at the smallest windows and saturates by 4096, which motivates our choice. The effect is small, however, and bounded — every window recovers far better than any compressed loss, with even the worst (512 points, geomean 5.9×10^{-12}) an order of magnitude better than Smooth MSS. This localizes the failure to log-magnitude compression rather than to spectral resolution.

5. METHOD VALIDATION AND SHORTCOMINGS

To test the final two-stage design, we ran the full pipeline on another set of 50 IRs generated from the plate simulator with known, randomly sampled parameters. The two stages used no more than 4GB of GPU VRAM.

Fig. 1 reports the NMSE of each IR before and after the Stage 2 μ -refinement. 49 out of 50 IRs improve: the median NMSE falls by eleven orders of magnitude, from 1.91×10^{-3} after Stage 1 alone to 1.34×10^{-14} after refinement. Likewise, the L_1 -STFT loss on the unnormalized IRs falls by four or five orders of magni-

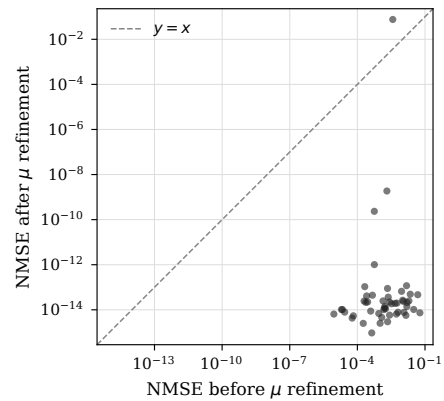


Figure 1: Per-IR NMSE (log-log) before and after Stage 2 μ -refinement, over 50 generated IRs. All points except one (with pathological x_o) fall well below $y = x$, with refined NMSE often on the order of 10^{-14} or lower.

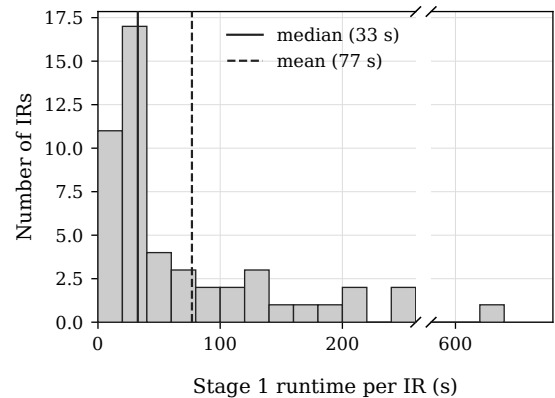


Figure 2: Distribution of per-IR Stage 1 runtime over the 50 validation IRs (linear axis, broken to show the tail). Most IRs finish near the median (32.6 s), while a few hard IRs extend past 200 s and a lone outlier reaches ~ 640 s, pulling the mean (76.9 s) above the median.

tude. Stage 1 takes a median of 32.6 s and a mean of 76.9 s per IR, after which Stage 2 contributes the NMSE improvement above at a negligible 0.62 s per IR on average. Because Stage 2 adjusts only μ , this also confirms that Stage 1 had mostly already recovered the other five parameters (D/μ , T_0/μ , L_y , x_o , y_o) to comparable precision.

The outlier in Fig. 1, with its ground-truth $x_o = 0.999980$ hovering very close to the upper limit of 1.0, represents a pathological corner case. While Stage 1 had estimated x_o up to an impressive $\sim 0.02\%$ relative error ($x_o = 0.999760$), the value of $(1 - x_o)$ for the ground-truth and the prediction differed by approximately one order. Meanwhile, the simulator is known to set the amplitude of each mode proportional to $|\sin(x_o \pi m)|$, where m is an integer associated with the x-axis component of each mode, typically $m \leq 200$. However, $|\sin(x_o \pi m)|$ is almost proportional to $(1 - x_o)m$ when $x_o m$ is close to any integer. Thus, for this particular IR, all mode amplitudes were altogether simultaneously

Table 3: Our two-stage pipeline vs. One-stage joint CMA-ES on the 50 validation IRs. Our pipeline is faster and more accurate. Both fail on the pathological outlier of Fig. 1, with comparable error, but the one-stage search uses far more time.

Search	All 50 IRs		Outlier in Fig. 1
	NMSE median	Time (s) median / mean	NMSE / Time (s)
Proposed	1.3×10^{-14}	33.1 / 76.9	$7.6 \times 10^{-2} / 205$
One-stage	2.0×10^{-14}	36.7 / 167	$1.4 \times 10^{-2} / 3604$

“mispredicted” by approximately one order. This in turn caused the estimations for μ , also an amplitude factor, to drift dramatically. Conversely, when $1 - x_o$ and $1 - y_o$ are not too small, $|\sin(x_o \pi m)|$ and $|\sin(y_o \pi m)|$ are thrown off by different magnitudes, and μ is correctly estimated. Meanwhile, the median estimation error for x_o and y_o were at the order of 10^{-8} , with the 2×10^{-4} error of the outlier being the top-1.

Observing the above, one may wonder if an alternate one-stage CMA-ES on unnormalized IRs would yield better estimations by jointly predicting μ and eliminating error propagation between stages (with the stages themselves). However, after scaling all termination criteria by each IR’s peak amplitude for a fair comparison, we discovered this one-stage search to slightly fall behind in NMSE across the set, while being slower (Table 3). Although the pathological outlier is better estimated by the one-stage search, this estimation is produced only after hitting the restart cap at over an hour, versus a few minutes taken for the two-stage search. We therefore adopt the two-stage pipeline for our final submission.

6. CONCLUSION

We have described our submission to Task A of the 1st DAFx parameter estimation challenge, in which an IR from a plate simulator is inverted back to its six generating physical parameters. We adopted a two-stage evolutionary approach: a CMA-ES search that recovers five of the six target parameters (D/μ , T_0/μ , L_y , x_o , y_o) under an amplitude-normalized loss, followed by a ternary search that refines the surface density μ on the un-normalized loss. We found the choice of loss to be decisive: an uncompressed L_1 -on-STFT-magnitude loss recovers the parameters almost perfectly, whereas the log-magnitude compression in the common multi-scale spectral loss collapses recovery. On a validation set of 50 IRs, our method reaches NMSE values of $\sim 10^{-14}$ in about 80 seconds per IR, and an ablation confirms that the two-stage split enables faster and more accurate estimations compared to a single-stage joint CMA-ES search. We also characterized a pathological failure mode that arises when an output pickup coordinate lies extremely close to the plate boundary. Our solution has been run on the challenge’s evaluation set; the results will be determined by the organizers.

7. REFERENCES

- [1] K. Arcas and A. Chaigne, “On the quality of plate reverberation,” *Applied Acoustics*, vol. 71, no. 2, pp. 147–156, 2010.
- [2] V. Valimaki, 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.
- [3] M. Ducceschi and L. Gabrielli, “The 1st DAFx Parameter Estimation Challenge: A Benchmark for Plate Reverberation System Identification,” University of Bologna and Università Politecnica delle Marche, Tech. Rep., 2025, available at <https://github.com/LOGUNIVPM/1st-DAFx-Challenge>.
- [4] L. Xu and F. Ding, “Parameter estimation for control systems based on impulse responses,” *International Journal of Control, Automation and Systems*, vol. 15, no. 6, pp. 2471–2479, 2017.
- [5] S. Lee, H.-S. Choi, and K. Lee, “Differentiable artificial reverberation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2541–2556, 2022.
- [6] N. Hansen, “The cma evolution strategy: A tutorial,” *arXiv preprint arXiv:1604.00772*, 2016.
- [7] M. Cámara, Z. Xu, Y. Zong, J. L. Blanco, and J. D. Reiss, “Optimization Techniques for a Physical Model of Human Vocalisation,” in *Proceedings of the 26th International Conference on Digital Audio Effects (DAFx23)*, Copenhagen, Denmark, 2023, pp. 29–36.
- [8] A. W. Leissa, *Vibration of plates*. Scientific and Technical Information Division, National Aeronautics and Space Administration, 1969, vol. 160.
- [9] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.
- [10] J. Engel, L. Hantrakul, C. Gu, and A. Roberts, “DDSP: Differentiable digital signal processing,” in *International Conference on Learning Representations (ICLR)*, 2020.
- [11] S. Ö. Arık, H. Jun, and G. Diamos, “Fast spectrogram inversion using multi-head convolutional neural networks,” *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 94–98, 2018.
- [12] 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 *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6199–6203.
- [13] S. J. Schwär and M. Müller, “Multi-scale spectral loss revisited,” *IEEE Signal Processing Letters*, vol. 30, pp. 1712–1716, 2023.