



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/226929/>

Version: Accepted Version

---

**Proceedings Paper:**

Dal Santo, Gloria, De Bortoli, Gian Marco, Prawda, Karolina Anna et al. (2025) FLAMO: An Open-Source Library for Frequency-Domain Differentiable Audio Processing. In: ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE International Conference on Acoustics, Speech and Signal Processing, 06-11 Apr 2025 IEEE, IND.

<https://doi.org/10.1109/ICASSP49660.2025.10888532>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# FLAMO: An Open-Source Library for Frequency-Domain Differentiable Audio Processing

Gloria Dal Santo<sup>1</sup>, Gian Marco De Bortoli<sup>1</sup>, Karolina Prawda<sup>2</sup>, Sebastian J. Schlecht<sup>3</sup>, Vesa Välimäki<sup>1</sup>

<sup>1</sup>*Acoustics Lab, Dpt. of Information and Communications Engineering, Aalto University, Espoo, Finland*

<sup>2</sup>*AudioLab, School of Physics, Engineering, and Technology, University of York, York, United Kingdom*

<sup>3</sup>*Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany*  
gloria.dalsanto@aalto.fi; gian.debortoli@aalto.fi; karolina.prawda@york.ac.uk; sebastian.schlecht@feu.de; vesa.valimaki@aalto.fi

**Abstract**—We present *FLAMO*, a Frequency-sampling Library for Audio-Module Optimization designed to implement and optimize differentiable linear time-invariant audio systems. The library is open-source and built on the frequency-sampling filter design method, allowing for the creation of differentiable modules that can be used stand-alone or within the computation graph of neural networks, simplifying the development of differentiable audio systems. It includes predefined filtering modules and auxiliary classes for constructing, training, and logging the optimized systems, all accessible through an intuitive interface. Practical application of these modules is demonstrated through two case studies: the optimization of an artificial reverberator and an active acoustics system for improved response coloration.

**Index Terms**—Audio systems, gradient methods, machine learning, optimization, reverberation.

## I. INTRODUCTION

As Engel’s differentiable digital signal processing (DDSP) [1] became a popular approach to modeling audio synthesis and processing techniques, there has been increasing interest in interpretable and differentiable audio processors such as filters, equalizers (EQs), and reverberators [2]–[6]. The appeal lies in the ability to run gradient-based optimization of their parameters, embed them in neural networks, or generate their parameters via hyperconditioning [7].

Frequency sampling is a technique used in DDSP, where the system’s response is sampled at discrete complex frequencies. This approach is particularly useful for linear time-invariant (LTI) systems, as the convolution operation allows factorization of their frequency response into a cascade of processing units with multiplied frequency responses. More often, frequency sampling is used to obtain a finite impulse response (FIR) approximation of infinite impulse response (IIR) filters [2], simplifying the implementation compared to time-domain methods, which suffer from vanishing/exploding gradients, high memory costs, and slow training [2]–[4]. Using frequency-sampling, Nercessian [8] realized a parametric EQ using differentiable biquads, later extending it to model audio distortion effects [7]. Similarly, Lee et al. [5] applied this approach to create differentiable artificial reverberators using filtered velvet noise [9] and feedback delay networks (FDN) [10], and used them to train a neural network as a

parameter estimator. The authors used frequency sampling to optimize an FDN for smooth, natural-sounding reverbs [11], [12]. Similarly, the authors optimized an active acoustics (AA) system to achieve a flat feedback loop magnitude response and improve system stability [13].

This paper presents *FLAMO*, a Python library for trainable DDSP modules based on frequency sampling. The library is developed using the PyTorch framework for automatic differentiation [14]. It is freely available online<sup>1</sup>, under a permissive open-source license, and comes with comprehensive documentation<sup>2</sup>. We demonstrate its application in artificial reverberation and AA by replicating the author’s previous work [11]–[13] using the library’s classes and modules.

One issue with frequency sampling is balancing accuracy and computational efficiency. Higher sampling densities can provide more accurate filter response approximations by reducing time-aliasing [15], but they also increase the computational load. Moreover, ensuring the stability of IIR filters can be challenging. To address these challenges, we propose an anti-aliasing method that uses exponential decaying time envelopes and assignable mapping functions to map raw filter parameters to a target distribution. In addition, *FLAMO* provides methods for logging intermediate results, domain flexibility, integration with neural networks, and pre-configured trainer and dataset classes, making it a versatile tool for the DDSP community.

The paper is structured as follows. Sec. II provides background on frequency-domain optimization and the proposed time-aliasing reduction method. Sec. III gives an overview of the library’s core and utility classes. Sec. IV validates the library’s effectiveness. Sec. V concludes the paper.

## II. FREQUENCY-DOMAIN OPTIMIZATION

### A. Frequency sampling method

IIR filters can be found in many audio effects, such as reverbs, EQs, and tone-shaping filters [16], [17]. While differentiable FIR filters can be implemented directly by convolutional layers [2], implementing IIR filters is more challenging due to their recursive nature and stability requirements. To avoid the drawbacks of time-domain optimization [2]–[4], the frequency

The Aalto University School of Electrical Engineering funded the work of the first author.

<sup>1</sup><https://github.com/gdalsanto/flamo>

<sup>2</sup><https://gdalsanto.github.io/flamo>

sampling method can be used to create an FIR approximation of IIR filters by sampling the filter’s frequency response at discrete complex frequencies [7], [8] whose resolution is chosen to minimize time-aliasing distortion [5], [7].

The response of a real-valued filter is sampled on a vector of linearly-spaced frequencies from 0 to  $\pi$  rad/s:

$$\mathbf{z}_M = [e^{j\pi \frac{0}{M-1}}, e^{j\pi \frac{1}{M-1}}, \dots, e^{j\pi \frac{M-2}{M-1}}, e^{j\pi}], \quad (1)$$

where  $M$  is number of frequency bins, and  $j = \sqrt{-1}$ . The frequency response of a single-input and single-output IIR filter  $H$  of order  $N$  can be evaluated over  $\mathbf{z}_M$  as follows

$$H(\mathbf{z}_M) = \frac{\text{DFT}(\mathbf{b})}{\text{DFT}(\mathbf{a})} = \frac{b_0 + b_1 \mathbf{z}_M^{-1} + \dots + b_N \mathbf{z}_M^{-N+1}}{a_0 + a_1 \mathbf{z}_M^{-1} + \dots + a_N \mathbf{z}_M^{-N+1}}, \quad (2)$$

where coefficients  $\mathbf{b} = [b_0, b_1, \dots, b_N]$  and  $\mathbf{a} = [a_0, a_1, \dots, a_N]$  are the feedforward and feedback gains of the filter, respectively. We assume real-time-domain signals such that their discrete Fourier transform (DFT) is described completely by the frequency bins in  $\mathbf{z}_M$ .

The overall frequency response of two cascaded systems, denoted as  $\mathbf{H}_1$  and  $\mathbf{H}_2$ , is obtained by multiplying their frequency response matrices, as long as the input and output dimensions are consistent,

$$\mathbf{H}_{\text{series}}(\mathbf{z}_M) = \mathbf{H}_1(\mathbf{z}_M) \mathbf{H}_2(\mathbf{z}_M), \quad (3)$$

where the use of boldface notation indicates a multiple-input and multiple-output (MIMO) system. Meanwhile, for a recursive system with feedforward path described by  $\mathbf{G}$  and feedback path described by  $\mathbf{F}$ , the resulting frequency response is determined by

$$\mathbf{H}_{\text{recursion}}(\mathbf{z}_M) = (\mathbf{I} - \mathbf{G}(\mathbf{z}_M) \mathbf{F}(\mathbf{z}_M))^{-1} \mathbf{G}(\mathbf{z}_M), \quad (4)$$

where  $\mathbf{I}$  is the identity matrix of the same size as  $\mathbf{G}\mathbf{F}$ .

To determine  $h[n]$  we perform the inverse DFT (IDFT) by using the Hermitian property of real-valued signals on the frequency response  $H(\mathbf{z}_M)$ . However, if the effective duration  $L$  of the impulse response (IR) exceeds the Fourier transform length  $2(M-1)$ ,  $h[n]$  will experience aliasing [2], [5], [15]. This issue is especially significant for lossless systems, where there are frequency points  $z_m = e^{j\pi \frac{m}{M-1}}$  such that  $H(z_m) = \infty$ . For lossy IIR filters, the time-aliasing error asymptotically decreases as the value of  $M$  increases [5], at the cost of higher computational expense. Fig. 1 demonstrates time aliasing in the IR of an FDN used to synthesize reverb with homogeneous decay [18], characterized by a reverberation time  $T_{60}$  of 9 s. To produce the IR, we performed DFT with  $M = 2f_s$  (orange) and  $M = 10f_s$  (black), where  $f_s$  is the sampling frequency. Due to time aliasing, the orange curve exhibits spurious impulses both before and after the actual onset time.

### B. Time-aliasing mitigation

To mitigate time-aliasing, we propose an anti-aliasing method that involves applying an exponentially decaying en-

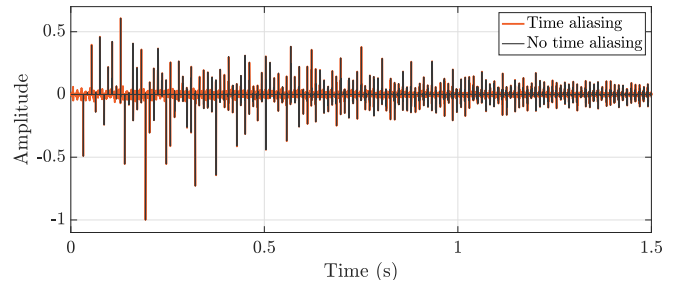


Fig. 1. IR of an FDN decaying with  $T_{60} = 9$  s with (orange) and without (black) time aliasing.

velope, denoted as  $\gamma^n$  for  $n \geq 0$ , before transforming to the frequency domain. The IR of each module is then given by

$$\hat{h}[n] = h[n] \gamma^n, \quad (5)$$

where  $0 < \gamma \leq 1$  is the decay parameter.

Using the DFT pair  $f[n] \gamma^n = f[n] e^{n \ln \gamma}$  and  $F(\omega + j \ln \gamma)$ , where  $f[n]$  is a real function with DFT  $F(\omega)$  and  $\omega$  is the angular frequency in rad/s, we can express the frequency response of (5) as  $\hat{H}(e^{j\omega}) = H(e^{j(\omega + j \ln \gamma)}) = H(e^{j\omega} / \gamma)$ . Thus, introducing an exponential decay in the time domain corresponds to sampling the frequency response outside the unit circle, with a radius of  $1/\gamma$ . The original IR  $h$  can then be recovered from  $\hat{h}$  by compensating for the decay in time-domain, i.e.,

$$h[n] = \hat{h}[n] \gamma^{-n} = \text{IDFT}(\hat{H}(e^{j\omega})) \gamma^{-n}. \quad (6)$$

If samples in the wrapping region during convolution are not sufficiently attenuated by  $\gamma^n$ , they appear at the start of the impulse response. Conversely, a very small  $\gamma^n$  can amplify numerical noise, which depends on machine precision. Thus,  $\gamma$  must be carefully selected to balance attenuation and noise.

The technique can be applied to a series of multiple LTI modules. By consistently applying the same exponential decay to each time-domain filter before performing the DFT, one can effectively manage time-domain aliasing. This method is particularly useful for enhancing the early response of a system. Without such a technique, avoiding aliasing would either necessitate excessive damping of the system or require computing the entire response, both of which are inefficient.

### III. FLAMO LIBRARY STRUCTURE

Fig. 2 provides an overview of the *FLAMO* library’s structure. The primary class is `DSP`, from which the `Gain`, `Filter`, and `Delay` classes inherit attributes and methods. The library’s interface has been simplified by the `Trainer` and `Shell` utility classes, which allow users to set training parameters, initiate optimization, and log intermediate results. Each processing module receives the frequency response of the incoming signal as input and multiplies it by the frequency response of the system that the module represents. This operation is defined using notation based on the Einstein summation convention using the `torch.einsum()` function [19]. Frequency responses are complex-valued tensors, and we utilize

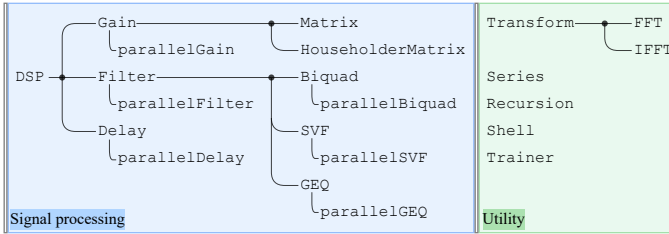


Fig. 2. Inheritance tree of the classes in the *FLAMO* library designed for signal processing (left) and utility (right) purposes.

PyTorch’s ability to backpropagate through real-valued functions of complex tensors via Wirtinger calculus [20]. Safety checks are implemented to ensure coherent tensor shapes and frequency-sampling parameter values.

Each module operates as a MIMO system. In the case of parallel processing, where each channel of the incoming signal passes through a different instance of the same system, element-wise vector multiplication of the frequency responses can be used to reduce the number of computations. For this purpose, each class has a corresponding `parallel*` version, where `*` denotes the original system module. A summary of the shapes of the tensors expected and produced by each class is provided in Table I, from which the `Delay` class has been omitted since it shares the same shapes as the `Filter` class.

#### A. Digital signal processor class

The `DSP` class represents a generic learnable module within the library. It inherits from PyTorch’s `nn.Module` [21], with its primary attribute, `param`, representing the learnable parameters for each instance of the class. These parameters can be mapped from the initial distribution to a user-specified distribution via the `map` argument. By default, the distribution is set to normal at initialization. Some inherited classes built into the library have predefined mappings to ensure system stability or to achieve filter-specific parameterizations.

The `Gain` and `Matrix` child classes implement systems of broadband gains. The latter allows one to choose among a set of pre-built mappings, such as mappings to orthogonal matrices [22]. For FIR filters one can use the `Filter` class.

The library provides subclasses for the most commonly used filter parameterizations, including biquad, state variable filter (SVF) [3], [23], and graphic EQs (GEQs) [17], [24]. The `Biquad` class implements low-pass, high-pass, and band-pass filters using RBJ cookbook formulas [25], which map the cut-off frequency and gain parameters to the `b` and `a` coefficients. For higher-order filters, the `SVF` class allows cascading multiple biquad filters parameterized by SVF parameters. These parameters offer simpler stability conditions [3], [23] and often outperform raw biquad parameters [5]. The `SVF` class includes pre-built mappings for second-order low-pass, high-pass, and band-pass filters, as well as low-shelf, high-shelf, peaking, and notch filters [23]. Alternatively, the class can initialize a generic SVF filter with random mixing coefficients. The `GEQ` class implements GEQs as described in [24] for full and one-third octave intervals.

TABLE I

FREQUENCY RESPONSE SHAPES FOR EACH MODULE, THEIR EXPECTED INPUT, AND THE PRODUCED OUTPUT.  $N_{\text{IN}}$  AND  $N_{\text{OUT}}$  ARE THE INPUT AND OUTPUT DIMENSIONS,  $M$  IS THE NUMBER OF FREQUENCY POINTS, AND  $B$  IS THE BATCH SIZE.

Class	Module $H(z)$	Input	Output
<code>Gain</code>	$N_{\text{out}}N_{\text{in}}$	$BMN_{\text{in}}$	$BMN_{\text{out}}$
<code>parallelGain</code>	$N_{\text{in}}$	$BMN_{\text{in}}$	$BMN_{\text{in}}$
<code>Filter</code>	$MN_{\text{out}}N_{\text{in}}$	$BMN_{\text{in}}$	$BMN_{\text{out}}$
<code>parallelFilter</code>	$MN_{\text{in}}$	$BMN_{\text{in}}$	$BMN_{\text{in}}$

Differentiable delays are implemented directly in the frequency domain via the complex exponential  $z_M^{-m}$ , where  $m$  is the delay in samples. By default, delays are fractional, but integer sample delays can be requested. The learnable parameter of the `Delay` class is the delay length in seconds. We recommend adjusting the `unit` attribute, which controls the time unit, as using subdivisions or multiples of time can improve update effectiveness in certain applications.

#### B. Utility classes

The `Series` class extends PyTorch’s `nn.Sequential` [26] to maintain coherence among chained modules and prevent error-prone nesting. The `Recursion` class implements a recursive closed-loop frequency response (4) having processing modules as feedforward and feedback paths. The `Shell` class is a container designed to interface the user-defined DSP chain with the desired loss function and to provide auxiliary functionalities. It has three main attributes: `core`, `input_layer`, and `output_layer`. The `core` represents the differentiable system to be optimized, while `input_layer` and `output_layer` handle transformations between the system, the dataset input, and the loss function. The `Shell` keeps these components separate, allowing the system to be defined independently of the loss function and dataset. Additionally, the class includes methods to return the system’s time and frequency responses by temporarily modifying the input and output layers.

`Series`, `Recursion`, and `Shell` classes also ensure that the operational flow attributes, e.g.  $M$ , are consistent across all modules, between feedforward and feedback paths, and between `input_layer`, `core`, and `output_layer`, respectively. The `Trainer` class manages the optimization of the modules within a system, it handles key functionalities such as initialization of training parameters, criterion management, training loop, and logging of intermediate states.

## IV. APPLICATIONS

#### A. Active acoustics

AA systems are used to electronically provide controlled variability in the acoustics of closed spaces [27]–[29]. They consist of microphones, loudspeakers, and a DSP unit. The AA system must be well-tuned to blend the artificial sound field with the natural-room sound field and to avoid coloration artifacts. The tuning process requires time and expertise, as many

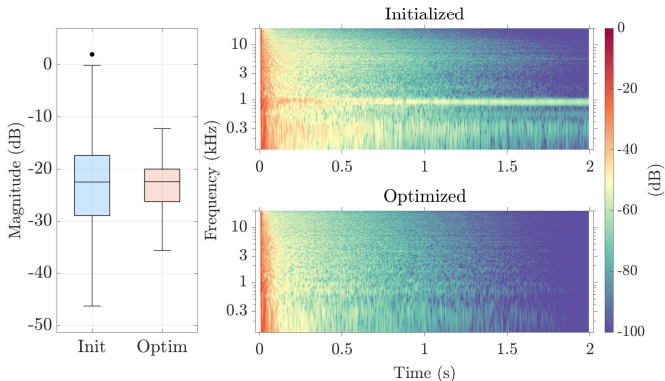


Fig. 3. (Left) Eigenvalue magnitude distribution and (right) spectrogram of an active acoustics system before and after optimization for spectral flatness.

parameters, e.g. transducer gains and filter coefficients, must be carefully chosen. Previously, the authors have demonstrated the applicability of DDSP in the AA scenario to automatize the tuning process to flatten the spectral behavior, which led to a reduced coloration in the full-system IR [13].

We redefine the former design [13] with the *FLAMO* library. The `Filter` class implements the learnable FIR matrix  $U(z)$  and a fixed white-Gaussian-noise-reverb matrix  $R(z)$ . The general system gain  $G$  is built with the `parallelGain` class. The loudspeaker-to-microphone IR matrix  $H_{LM}$  is implemented as a `Filter`-class instance of the measured room IRs. The transfer function matrix describing one feedback-loop iteration is  $F_{MM}(z) = GH_{LM}(z)R(z)U(z)$ . We run optimization to flatten the frequency response of  $F_{MM}(z)$  and improve the system stability. Fig. 3 presents the optimization results for the magnitude eigenvalue distribution and the spectrograms of the full-system IR. In the box plot, the lower and upper edges represent the 25th and 75th percentiles, respectively, highlighting the narrowing of the magnitude eigenvalue distribution from initialization. The coloration artifacts, indicated by long-ringing modes in the spectrograms, are also considerably attenuated. The black dot in Fig. 3 indicates the maximum outlier of the initial and optimized eigenvalue magnitude distribution and it is responsible for the strong resonance at 1 kHz in the top-spectrogram. Further outliers have not been considered as they don't add information regarding system stability or spectral flatness.

### B. Delay-line-based artificial reverberation

The FDN is a recursive system used in reverb synthesis [10], [16], [18] and consists of delay lines  $m$ , a set of input and output gains  $b$  and  $c$ , and a scalar feedback matrix through which the delay outputs are coupled to the delay inputs. The transfer function of the FDN is

$$H(z) = c^T [D_m(z)^{-1} - A]^{-1} b + d \quad (7)$$

where  $A$  is the  $N \times N$  feedback matrix, with  $N$  being the number of delay lines,  $D_m(z)$  is the diagonal delay matrix, and  $d$  is the direct gain. The operator  $(\cdot)^T$  denotes the transpose. A common design choice is to start with a lossless FDN prototype with a smooth response. This can be achieved by

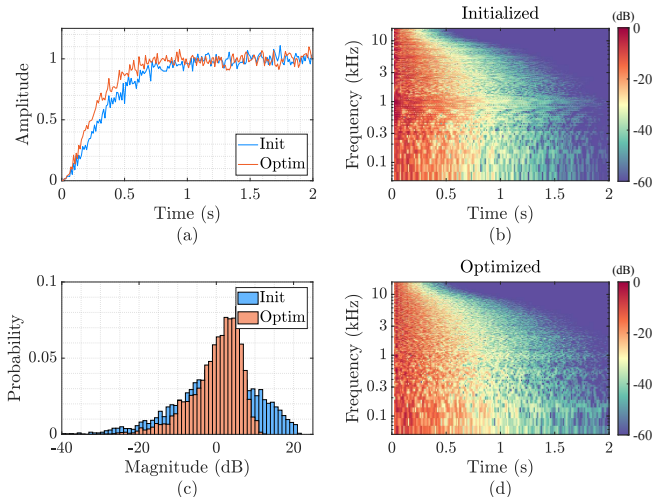


Fig. 4. Effect of the optimization of an FDN of size  $N = 6$  with gain coefficient optimized to improve smoothness. Echo density profile (a) modal excitation distribution (c) and spectrogram of the IR (b) before and (d) after optimization.

using an orthogonal matrix  $A$  [30], [31] optimized for spectral flatness and temporal density [11], [12]. Attenuation filters are then inserted to control frequency-dependent  $T_{60}$  [10], [24], [32], [33]. We implement and optimize a lossless FDN with temporal anti-aliasing using *FLAMO*. This approach contrasts with [11], [12], where the system had to be made lossy to enable backpropagation. Matrix  $A$  is implemented using a learnable `Matrix`-class instance with orthogonal mapping. Input and output gains are learnable `Gain`-class instances. Delays of given lengths are an instance of the `parallelDelay` class. The attenuation filters are then implemented using the class `parallelGEQ`, with the parameter mapping proposed in [24], and inserted in the optimized system.

Improvements in temporal density can be observed from the echo density profile [34], which is representative of the distribution of reflections over time. Fig. 4 (a) shows a faster echo build-up achieved by the optimized FDN. Reduction in coloration is reflected by a narrower distribution of the modal excitation [35] depicted in Fig. 4 (c), where a significant attenuation of the loudest modes can be observed. Fig. 4 (b) and (d) show the spectrogram of the system's response before and after optimization, respectively. The optimization results in more evenly distributed energy across the frequency range, reducing coloration artifacts.

## V. CONCLUSION

The *FLAMO* library for frequency-sampling optimization of differentiable audio modules simplifies the implementation and optimization of differentiable LTI systems, offering an intuitive API, and can be integrated with PyTorch neural networks. We validate the library's effectiveness through artificial and enhanced reverberation scenarios. The source code is available in an online repository<sup>1</sup> and on the Python package index<sup>3</sup>. We plan to continuously expand the *FLAMO* library to support the audio signal processing community.

<sup>3</sup><https://pypi.org/project/flamo>

## REFERENCES

- [1] J. Engel, L. H. Hantrakul, C. Gu, and A. Roberts, "DDSP: Differentiable digital signal processing," in *Proc. Int. Conf. Learning Representations*, 2020.
- [2] B. Hayes, J. Shier, G. Fazekas, A. McPherson, and C. Saitis, "A review of differentiable digital signal processing for music and speech synthesis," *Frontiers in Signal Processing*, vol. 3, 2024.
- [3] B. Kuznetsov, J. D. Parker, and F. Esqueda, "Differentiable IIR filters for machine learning applications," in *Proc. 23rd Int. Conf. Digital Audio Effects*, pp. 297–303, 2020.
- [4] C. J. Steinmetz, N. J. Bryan, and J. D. Reiss, "Style transfer of audio effects with differentiable signal processing," *J. Audio Eng. Soc.*, vol. 70, pp. 708–721, 2022.
- [5] S. Lee, H.-S. Choi, and K. Lee, "Differentiable artificial reverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 2541–2556, 2022.
- [6] S. Lee, M. Martínez-Ramírez, W.-H. Liao, S. Uhlich, G. Fabbro, K. Lee, and Y. Mitsufuji, "Grafx: an open-source library for audio processing graphs in pytorch," *arXiv preprint arXiv:2408.03204*, 2024.
- [7] S. Nercessian, A. Sarroff, and K. J. Werner, "Lightweight and interpretable neural modeling of an audio distortion effect using hyperconditioned differentiable biquads," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 890–894, IEEE, 2021.
- [8] S. Nercessian, "Neural parametric equalizer matching using differentiable biquads," in *Proc. 23rd Int. Conf. Digital Audio Effects*, pp. 265–272, 2020.
- [9] V. Välimäki, B. Holm-Rasmussen, B. Alary, and H.-M. Lehtonen, "Late reverberation synthesis using filtered velvet noise," *Appl. Sci.*, vol. 7, no. 5, p. 483, 2017.
- [10] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 2, pp. 221–224, 1992.
- [11] G. Dal Santo, K. Prawda, S. J. Schlecht, and V. Välimäki, "Differentiable feedback delay network for colorless reverberation," in *Proc. 26th Int. Conf. Digital Audio Effects*, (Copenhagen, Denmark), pp. 244–251, 2023.
- [12] G. Dal Santo, K. Prawda, S. J. Schlecht, and V. Välimäki, "Optimizing tiny colorless feedback delay networks," *arXiv preprint arXiv:2402.11216*, 2024.
- [13] G. M. De Bortoli, G. Dal Santo, K. Prawda, V. Välimäki, T. Lokki, and S. J. Schlecht, "Differentiable active acoustics," in *Proc. 27th Int. Conf. Digital Audio Effects*, (Guildford, UK), 2024.
- [14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.
- [15] J. O. Smith, *Mathematics of the Discrete Fourier Transform (DFT)*. W3K, 2007.
- [16] 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, 2012.
- [17] V. Välimäki and J. D. Reiss, "All about audio equalization: Solutions and frontiers," *Appl. Sci.*, vol. 6, no. 5, p. 129, 2016.
- [18] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Audio Engineering Society Convention 90*, 1991.
- [19] "PyTorch torch.einsum." <https://pytorch.org/docs/stable/generated/torch.einsum.html>. Accessed: 2024-09-8.
- [20] "PyTorch complex numbers." [https://pytorch.org/docs/stable/complex\\_numbers.html](https://pytorch.org/docs/stable/complex_numbers.html). Accessed: 2024-08-23.
- [21] "PyTorch module." <https://pytorch.org/docs/stable/generated/torch.nn.Module.html>. Accessed: 2024-09-8.
- [22] M. Lezcano-Casado and D. Martínez-Rubio, "Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group," in *Proc. Int. Conf. Machine Learning*, pp. 3794–3803, 2019.
- [23] A. Wishnick, "Time-varying filters for musical applications," in *Proc. 17th Int. Conf. Digital Audio Effects*, pp. 69–76, 2014.
- [24] S. J. Schlecht and E. A. Habets, "Accurate reverberation time control in feedback delay networks," in *Proc. 20th Int. Conf. Digital Audio Effects, Edinburgh, UK*, pp. 337–344, 2017.
- [25] R. Bristow-Johnson, "RBJ Audio EQ Cookbook." <https://webaudio.github.io/Audio-EQ-Cookbook/Audio-EQ-Cookbook.txt>, 2005. [Online; accessed 19-August-2024].
- [26] "PyTorch sequential." <https://pytorch.org/docs/stable/generated/torch.nn.Sequential.html>. Accessed: 2024-09-8.
- [27] P. U. Svensson, *On Reverberation Enhancement in Auditoria*. PhD thesis, Chalmers University of Technology, Göteborg, Sweden, 1994.
- [28] M. Poletti, "The stability of single and multichannel sound systems," *Acta Acust. un. Acust.*, vol. 86, no. 1, pp. 163–178, 2000.
- [29] M. Poletti, "Active acoustic systems for the control of room acoustics," *Build. Acoust.*, vol. 18, no. 3–4, pp. 237–258, 2011.
- [30] S. J. Schlecht and E. A. Habets, "Feedback delay networks: Echo density and mixing time," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 2, pp. 374–383, 2016.
- [31] J. Stautner and M. Puckette, "Designing multi-channel reverberators," *Computer Music Journal*, vol. 6, no. 1, pp. 52–65, 1982.
- [32] 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.
- [33] K. Prawda, S. J. Schlecht, and V. Välimäki, "Improved reverberation time control for feedback delay networks," in *Proc. 22nd Int. Conf. Digital Audio Effects*, pp. 1–7, 2019.
- [34] J. S. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *Proceedings of the Audio Engineering Society Convention 121*, Audio Engineering Society, 2006.
- [35] J. Heldmann and S. J. Schlecht, "The role of modal excitation in colorless reverberation," in *Proc. 24th Int. Conf. Digital Audio Effects*, pp. 206–213, 2021.