De Bortoli, Gian; Dal Santo, Gloria; Prawda, Karolina; Lokki, Tapio; Välimäki, Vesa; Schlecht, Sebastian

Differentiable Active Acoustics: Optimizing Stability via Gradient Descent

Published in:
Proceedings of the 27th International Conference on Digital Audio Effects (DAFx24)

Published: 03/09/2024

Document Version
Publisher's PDF, also known as Version of record

DIFFERENTIABLE ACTIVE ACOUSTICS—OPTIMIZING STABILITY VIA GRADIENT DESCENT

Gian Marco De Bortoli *, *Gloria Dal Santo, Karolina Prawda, Tapio Lokki, Vesa Välimäki and Sebastian J. Schlecht* †

Aalto Acoustics Lab
Dept. of Information and Communications Engineering
Aalto University
Espoo, Finland
gian.debortoli@aalto.fi

ABSTRACT

Active acoustics (AA) refers to an electroacoustic system that actively modifies the acoustics of a room. For common use cases, the number of transducers—loudspeakers and microphones—involved in the system is large, resulting in a large number of system parameters. To optimally blend the response of the system into the natural acoustics of the room, the parameters require careful tuning, which is a time-consuming process performed by an expert. In this paper, we present a differentiable AA framework, which allows multi-objective optimization without impairing architecture flexibility. The system is implemented in PyTorch to be easily translated into a machine-learning pipeline, thus automating the tuning process. The objective of the pipeline is to optimize the digital signal processor (DSP) component to evenly distribute the energy in the feedback loop across frequencies. We investigate the effectiveness of DSPs composed of finite impulse response filters, which are unconstrained during the optimization. We study the effect of multiple filter orders, number of transducers, and loss functions on the performance. Different loss functions behave similarly for systems with few transducers and low-order filters. Increasing the number of transducers and the order of the filters improves results and accentuates the difference in the performance of the loss functions.

1. INTRODUCTION

Active acoustics (AA) systems include sound reinforcement and reverberation enhancement systems [1]. Usually, they comprise several microphones and loudspeakers distributed in a closed space and a digital signal processor (DSP). Feedback is an AA system's inherent component: the signal produced by a sound source is picked up by the microphones, processed in the DSP, played back in the room by the loudspeakers, and picked up again by the microphones.

We can divide AA systems into two categories [1, 2, 3]: *inline* systems that suppress the feedback with the use of directional microphones placed close to the sound source and *non-inline*, or *regenerative*, systems, that use feedback as their working principle. Inline systems work far from the stability limit, and hence they use long impulse responses (IRs) to produce artificial reverberation. Regenerative systems, on the other hand, do not suppress feedback and can generate long reverberation by working close to the stability limit allowing the feedback to generate many repetitions of the input signal. Thus, they can also employ filters with short IRs. Usually, commercial systems use a hybrid approach comprising both an inline and a regenerative component [2, 4, 5].

Due to the feedback nature of AA systems, stability is one of the crucial considerations in system design and use [1, 2, 6, 7]. An unstable feedback loop results in the signal amplitude increasing at each loop iteration, up to the point of transducer saturation [1]. To avoid such a state, the gain that can be safely applied to the system is constrained by the gain before instability (GBI) [1, 3]. Approaching the GBI, however, may still result in audible artifacts in the enhanced sound, such as strong coloration in the form of long-ringing tones and modulation [1, 8]. The ringing tones theoretically coincide with the frequencies at which the feedback loop's signal amplification is stronger [6, 7]. Therefore, the design and implementation of AA systems benefit from having the energy of the feedback loop evenly distributed across frequencies [8, 9]. An AA system with an identical stability threshold for all frequencies can work closer to the GBI without coloration affecting the feedback loop. Upon reaching the GBI, such a system would become unstable at all frequencies at once.

In the literature, numerous techniques aim to maximize the AA systems' performance in terms of control over reverberation time (RT) and/or gain, simultaneously maintaining their stability and minimizing artifacts. The state of the art includes transducer positioning and directivity investigation [10, 11], equalization of transducer gains [12, 13, 14], adaptive feedback cancellation [15, 16, 17], spectrum decorrelation techniques [8, 18, 19, 20], and time-varying reverberators [3, 21, 22]. The success in providing high GBI and good-quality sound is strongly method-dependent [23]. In recent years, research explored geometric and perceptually-motivated approaches [5, 24, 25] to assess the quality of AA systems. However, they all share a major drawback: they require fine-tuning [3, 14], which is time-inefficient and demands expert knowledge, especially in systems with a high number of channels and complex DSPs.

A way to avoid the laborious manual tuning of AA system parameters is through automation. In this paper, we propose a PyTorch formulation of AA that allows for automatic differentiation of the DSP. The differentiable DSP (DDSP) can be optimized towards a target in the same fashion as a machine learning pipeline [26, 27]. We restrict the optimization design to regenerative AA systems with short DSP finite impulse response (FIR) filters, and we test the framework using several setups with different

* Corresponding author, e-mail: gian.debortoli@aalto.fi

† Also a member of Media Lab, Dept. of Art and Media, Aalto University, Espoo, Finland

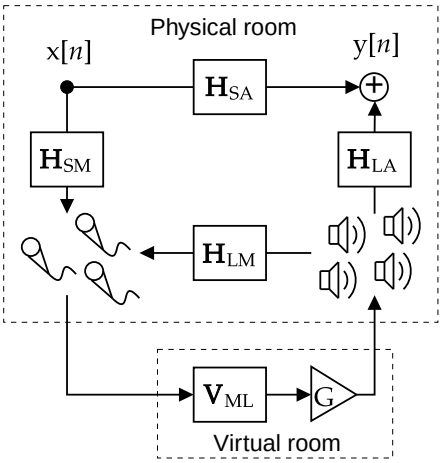


Figure 1: Block diagram representing an AA system in a real-world scenario. The signal routing is depicted by solid arrows with annotations of the involved IRs.

filter orders, loss functions, and transducer configurations. In the end, we test the application of this approach to an inline system.

The paper is organized as follows: Section 2 offers background information on the stability of AA systems; Section 3 describes the proposed framework and the optimization algorithm; Section 4 shows the results of comparing system stability between non-optimized and optimized DSP; and Section 5 concludes the article.

2. PROBLEM DEFINITION

Figure 1 shows the signal flow in an AA system's feedback loop. The sound field is generated in the *physical room*, where the transducers, i.e., n_{M} microphones and n_{L} loudspeakers, are positioned. The *virtual room* is the DSP, which digitally enhances the room acoustics. At time sample n, any sound $\mathbf{x}[n]$ produced in the physical room is picked up by the microphones, while the signal $\mathbf{y}[n]$ received at any position in the audience is the superimposition of the contributions from the physical room and the AA system. In the physical room, $\mathrm{H}_{\mathrm{SA}}[n] \in \mathbb{R}$, $\mathbf{H}_{\mathrm{SM}}[n] \in \mathbb{R}^{n_{\mathrm{M}}}$, and $\mathbf{H}_{\mathrm{LA}}[n] \in \mathbb{R}^{n_{\mathrm{L}}}$ are the room IRs (RIRs) between the sound source and audience position, the sound source and the system's microphones, the systems' loudspeakers and audience position, respectively. The RIRs $\mathbf{H}_{\mathrm{LM}}[n] \in \mathbb{R}^{n_{\mathrm{M}} \times n_{\mathrm{L}}}$ from the loudspeakers to the microphones are the system's feedback paths. In AA systems, any linear and time-invariant DSP results in a matrix of IRs, $\mathbf{V}_{\mathrm{ML}}[n] \in \mathbb{R}^{n_{\mathrm{L}} \times n_{\mathrm{M}}}$, from every microphone to every loudspeaker, and an amplification gain G. In this work, $\mathbf{V}_{\mathrm{ML}}[n]$ does not contain any internal feedback path, and therefore it is essentially a matrix of FIR filters.

The DSP transfer functions (TFs) from microphones to loudspeakers are $\mathbf{V}_{\mathrm{ML}}(z) = \mathcal{Z}\{\mathbf{V}_{\mathrm{ML}}[n]\}$ and the physical room transfer functions (RTFs) are $\mathbf{H}_{\mathrm{LM}}(z) = \mathcal{Z}\{\mathbf{H}_{\mathrm{LM}}[n]\}$. $\mathcal{Z}$ denotes the z-transform and $z = \sigma e^{\jmath\omega}$ is a complex number where e is Euler's number, $\jmath$ is the imaginary unit, and σ and ω are the radius and the phase, respectively, of z on the complex plane. An AA system's stability is analyzed along the unit circle, and hence, from now on, we consider $\sigma = 1$, and we discuss stability with respect to the discrete normalized angular frequency $\omega_k \in [0, \pi]$ for $k = 0, \ldots, K-1$.

An AA system's feedback loop iteration is determined by the product of the feedforward TFs and the feedback RTFs:

$$\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k}) = \mathrm{G}\, \mathbf{H}_{\mathrm{LM}}(e^{\jmath\omega_k})\mathbf{V}_{\mathrm{ML}}(e^{\jmath\omega_k}), \tag{1}$$

where $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ contains the TFs from any microphone to any microphone. The amplification gain G is a real scalar multiplier, and thus, for this work, we conveniently choose $\mathrm{G} = 1$.

Typically, all elements of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ are non-zero, meaning that all the channels in the system are coupled. An equivalent system with decoupled channels—i.e. eigenchannels—can be obtained by applying Eigen-decomposition to $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ [28]:

$$\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k}) = \mathbf{Q}(e^{\jmath\omega_k})\mathbf{\Lambda}(e^{\jmath\omega_k})\mathbf{Q}^{-1}(e^{\jmath\omega_k}), \tag{2}$$

where $\mathbf{Q}(e^{\jmath\omega_k})$ is the matrix of the system's eigenvectors, and $\mathbf{\Lambda}(e^{\jmath\omega_k})$ is a diagonal matrix containing the system's eigenvalues $\{\lambda_i(e^{\jmath\omega_k})\}$, for $i = 1, 2, \ldots, n_{\mathrm{M}}$. By applying Eq. (2) to each frequency ω_k we obtain the evolution of the eigenvalues over frequency. We refer to the complete collection of all system's eigenvalues, across both frequency and eigenchannels, with the term *eigenvalue set* and to the collection of the magnitude values of the eigenvalue set with the term *eigenvalue magnitude distribution*.

According to Nyquist's stability criterion [6, 7], the system is stable if all of the eigenvalues in the eigenvalue set have a real part lower than 0 dB or a non-zero imaginary part. A more stringent but safer and simpler approach is to consider a system stable if and only if the whole eigenvalue magnitude distribution is below 0 dB [20, 28]. Thus, we can determine the AA system stability by analyzing only the eigenvalue magnitude distribution. However, we know that for each frequency ω_k, the eigenvalue with the largest magnitude is the most likely to invalidate the stability condition. Therefore we can simplify the stability analysis by considering, for each frequency ω_k, only the eigenvalue with the largest magnitude across channels:

$$\lambda_{\max}(e^{\jmath\omega_k}) = \max_i \{|\lambda_i(e^{\jmath\omega_k})|\}. \tag{3}$$

Fulfilling the stability condition alone, however, does not warrant a colorless feedback loop. By applying Eq. (3) for each frequency ω_k we obtain the maximum eigenvalue curve. Strong irregularities in such a curve lead to perceivable ringing tones well below the stability limit due to some frequencies being more amplified than others by the feedback loop. A flat $\lambda_{\max}(e^{\jmath\omega_k})$ curve grants homogeneous decay for all frequencies looping in the system. In the case of a flat $\lambda_{\max}(e^{\jmath\omega_k})$ curve, a target RT-frequency profile can be then achieved by introducing an equalizer, but this goes beyond the scope of this work.

The maximum eigenvalue curve, though, is not representative of the whole system's eigenvalue set. A flat maximum eigenvalue curve implies that at least for one eigenchannel all frequencies contribute to the enhanced sound field. If, however, the eigenvalue magnitude distribution is broad, then most of the feedback loop energy is concentrated in one or few eigenchannels. On the other hand, magnitudes of the system's eigenvalues close to each other at a frequency ω_k provide a good distribution of the energy between the eigenchannels at that frequency. But, if this last condition is not preserved across frequencies, then not all of the spectrum contributes to the enhanced sound field. For this reason, the optimization should target an eigenvalue magnitude distribution that is narrow across both eigenchannels and frequencies.

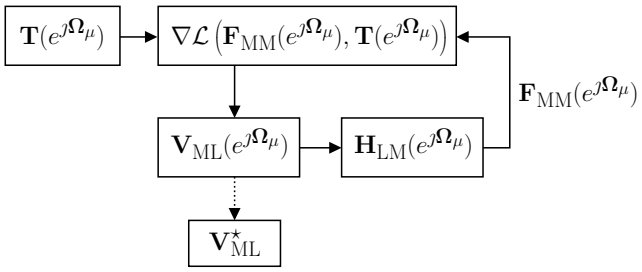


Figure 2: DDSP training pipeline to optimize the virtual room $\mathbf{V}^\star_{\mathrm{ML}}$. The parameters of the DSP of the differentiable AA system are updated through backpropagation of the gradient of the loss function.

Flattening the feedback loop magnitude responses may improve the eigenvalue magnitude distribution. An optimization algorithm requires less computational complexity if it targets flat $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ magnitude curves instead of a narrow eigenvalue magnitude distribution since eigendecomposition is not required. However, due to the way the elements of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ are combined in Eq. (2), such an algorithm would not account for how the phase responses in $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ affect the eigenvalue set. The result of the optimization would be a less optimal eigenvalue magnitude distribution. To our knowledge, there is no previous study that considers the difference between optimizing the magnitude responses of the feedback loop matrix or the eigenvalues concerning the generated enhanced reverberation in the physical room. For this reason, all three levels of analysis—$\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ magnitude responses, maximum eigenvalue curve, and eigenvalue magnitude distribution—are considered in this work.

3. PROPOSED METHOD

This section introduces an optimization-based approach to improve the feedback loop's energy distribution across frequencies. To achieve this goal, we defined the system's DSP as a set of trainable FIR filters, following a DDSP framework wherein sparse frequency sampling is employed [29]. In this framework, filter coefficients are optimized to minimize a loss function via stochastic gradient descent. This approach differs from conventional black-box machine learning techniques since the trainable parameters possess physical interpretations. Specifically, the DSP filters are responsible for the artificial sound enhancement injected in the physical room and directly affect the stability of the system.

3.1. Differentiable active acoustics

The diagram of the training pipeline for the proposed architecture is presented in Fig. 2. $\mathbf{V}_{\mathrm{ML}}(e^{\jmath\omega_k})$ is a matrix of $n_{\mathrm{L}} \times n_{\mathrm{M}}$ learnable FIR filters. For a given $\mathbf{H}_{\mathrm{LM}}(e^{\jmath\omega_k})$, the optimization framework estimates the coefficients of the optimized DSP $\mathbf{V}^\star_{\mathrm{ML}}$ FIR filters. The loss $\mathcal{L}$ of the feedback loop matrix $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ with respect to a target $\mathbf{T}(e^{\jmath\omega_k})$ is minimized using stochastic gradient descent, denoted with the gradient operator.

Before the training, we store the RIRs $\mathbf{H}_{\mathrm{LM}}[n]$, initialize the coefficients of $\mathbf{V}_{\mathrm{ML}}[n]$ by drawing from the random uniform distribution $\mathcal{U}(-1, 1)$, and define the dataset as K discrete frequency points sampled uniformly in the interval $[0, \pi - \frac{\pi}{K}]$,

$$\mathbf{\Omega}_K = \left\{\pi\frac{0}{K}, \ldots, \pi\frac{K-1}{K}\right\} \tag{4}$$

The dataset size K is chosen to ensure oversampling. In this study, we used $K = 480\,000$ with a sampling rate of $f_{\mathrm{s}} = 48\,\mathrm{kHz}$. At each training step, a random subset of μ frequency points is extracted from $\mathbf{\Omega}_k$ to form a batch. Consequently, for each batch $\mathbf{\Omega}_\mu$, $\mathbf{V}_{\mathrm{ML}}(e^{\jmath\mathbf{\Omega}_\mu})$ and $\mathbf{H}_{\mathrm{LM}}(e^{\jmath\mathbf{\Omega}_\mu})$ are computed using a non-uniform discrete Fourier transform and combined, according to Eq. (1), to obtain $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu})$. 90% of the dataset was used for the training, and the remaining 10% was used for the validation.

We observed empirically that losses converged after 10 epochs with a batch size of $\mu = 2400$. We employed an Adam optimizer [30] with learning rate $\eta = 10^{-3}$. The learnable FIRs were unconstrained, and thus each FIR was independent and each sample in each FIR was free to vary within the real numbers set.

3.2. Loss functions

The model was trained on six different configurations employing the mean squared error (MSE) loss. The loss minimization was computed from the magnitude of either $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu})$ or the system's eigenvalues in correspondence with the batch's frequency points.

The peaks in the magnitude of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ and in the eigenvalue magnitude distribution are the most dangerous for the system's stability and coloration. With this motivation, in this work, we consider an MSE variant, where the loss exponent depends on the sign of the magnitude difference between the target and either $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu})$, $\lambda_{\max}(e^{\jmath\mathbf{\Omega}_\mu})$, or $\{\lambda_i(e^{\jmath\mathbf{\Omega}_\mu})\}$. This variant, which we will refer to as *Mean Asymmetric Error* (MAsE), was introduced in [26] to attenuate masker tones in an artificial reverberator. Specifically, considering two generic tensors $\mathbf{A}$ and $\mathbf{B}$—both with dimensions d_1, d_2, and d_3—the loss functions are:

$$\mathcal{L}\Big(\mathbf{A}, \mathbf{B}\Big) = \frac{1}{d_1\, d_2\, d_3} \sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\sum_{k=1}^{d_3} (A_{ijk} - B_{ijk})^p. \tag{5}$$

For MSE, the exponent is $p = 2$. For MAsE, it is adjusted as follows:

$$p = \begin{cases} 2 & \text{for } \big(A_{ijk} - B_{ijk}\big) \leqslant 0\,, \\ 4 & \text{for } \big(A_{ijk} - B_{ijk}\big) > 0\,. \end{cases} \tag{6}$$

The six considered configurations were:

- MSE for the magnitude of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu})$:

$$\text{MSE-Magn} = \mathrm{MSE}\Big(\big|\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu})\big|, \mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu}))\Big)$$

- MAsE for the magnitude of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu})$:

$$\text{MAsE-Magn} = \mathrm{MAsE}\Big(\big|\mathbf{F}_{\mathrm{MM}}(e^{\jmath\mathbf{\Omega}_\mu}))\big|, \mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu}))\Big)$$

- MSE for the magnitude of the maximum eigenvalue curve:

$$\text{MSE-EVmax} = \mathrm{MSE}\Big(\big|\lambda_{\max}(e^{\jmath\mathbf{\Omega}_\mu})\big|, \mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu})\Big)$$

- MAsE for the magnitude of the maximum eigenvalue curve:

$$\text{MAsE-EVmax} = \mathrm{MAsE}\Big(\big|\lambda_{\max}(e^{\jmath\mathbf{\Omega}_\mu})\big|, \mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu})\Big)$$

- MSE for the eigenvalue magnitude distribution:

$$\text{MSE-allEVs} = \text{MSE}\Big(\big|\{\lambda_j(e^{\jmath\mathbf{\Omega}_\mu})\}\big|, \mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu})\Big)$$

- MAsE for the eigenvalue magnitude distribution:

$$\text{MAsE-allEVs} = \text{MAsE}\Big(\big|\{\lambda_j(e^{\jmath\mathbf{\Omega}_\mu})\}\big|, \mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu})\Big)\ ,$$

where the absolute value symbol $|\cdot|$ represents the matrix elementwise absolute value. Depending on the specific loss, the dimensions d_1, d_2, and d_3 in (5) were $(\mu, n_\text{M}, n_\text{M})$ for MSE-Magn and MAsE-Magn, $(\mu, 1, 1)$ for MSE-EVmax and MAsE-EVmax, and $(\mu, n_\text{M}, 1)$ for MSE-allEVs and MAsE-allEVs.

The presented pipeline does not apply particular constraints on the choice of the DSP and the target, which can be defined based on different strategies and loss functions. The framework can be applied to both regenerative and inline systems. In this work, we first consider regenerative AA systems. Regenerative systems usually employ filters with short IRs, which can be obtained with low-order FIR filters. This leads to a simple implementation with a small number of learnable parameters and a fast training process. In the end, we optimize the low-order FIR filters in an inline AA system comprising a fixed artificial reverberator. Furthermore, we aim to improve stability, and we do not consider perceptual metrics. Thus, as a target, we use a frequency-independent matrix of ones, i.e., $\mathbf{T}(e^{\jmath\mathbf{\Omega}_\mu}) = \mathbf{1}\ \forall\ \mathbf{\Omega}_\mu$. This typically leads to a loop TF that is too bright, since natural RIRs tend to have lowpass characteristics. Such a sound design aspect can be easily amended by post-filtering with an equalizer with a smooth frequency response.

4. RESULTS

In this section, we evaluate the stability improvement provided by the presented framework. We assess the performance of the optimization algorithm in terms of the flatness of the feedback loop magnitude responses and the system's eigenvalues distribution. Since the magnitude responses of $\mathbf{F}_\text{MM}(e^{\jmath\omega_k})$ and the eigenvalue magnitude distribution are correlated, we compare the results on both metrics for all the loss functions listed in Sec. 3.2.

4.1. Analysis setup

For the simulations, we used RIRs measured in a room equipped with a multi-input multi-output channel system comprising four microphones and 13 loudspeakers. The room fulfills the ITU-R BR.1116 standard requirements with a volume of 103 m^3 and RT of 0.3 s over a wide frequency range (100–8000 Hz). Four Behringer ECM8000 microphones are fixed on the ceiling. There are nine Genelec 8260A and four Genelec 8340A loudspeakers in a 9.0.4 setup.

During the measurements, we used a 2-s-long exponential sine sweep, that was played once for each loudspeaker and recorded simultaneously by all microphones. The procedure was repeated for each loudspeaker, obtaining a set of 52 recordings, which were then convolved with the inverse sweep to obtain the IRs [31].

To test the dependence of the algorithm on transducer number, we considered two transducer setups in the proposed framework analysis. The first setup comprised only two microphones and two loudspeakers, and this setup is referred to as *small system* in the remainder of this paper. The second setup, dubbed *full system*, included all system transducers.

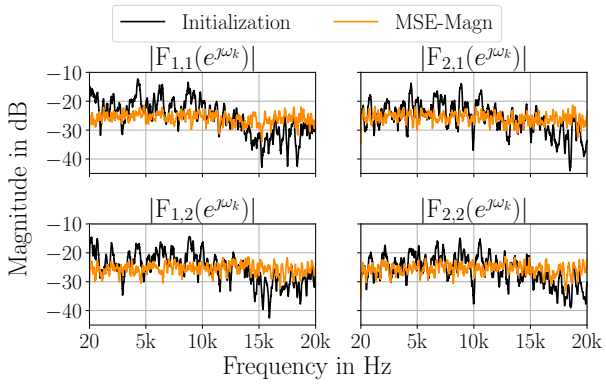


Figure 3: Comparison of the $\mathbf{F}_\text{MM}(e^{\jmath\omega_k})$ magnitude responses for the *small system* with DSP of order 100 between the non-trained DSP (black) and the DSP trained using MSE-Magn loss function (orange).

To test the algorithm's performance in relation to the length of the DSP's IRs, we considered FIR filters of orders 100 and 1000. The order of the filters remains consistent across all the DSP matrix elements. We compare non-trained and trained FIR filters, where the initial (non-trained) FIR coefficients are randomized non-integer values between -1 and 1. We always normalized the DSP IR matrix before and after training:

$$\big(\mathbf{V}_\text{ML}[n]\big)_\text{normalized} = \frac{\mathbf{V}_\text{ML}[n]}{\|\mathbf{V}_\text{ML}\|_\text{F}}, \tag{7}$$

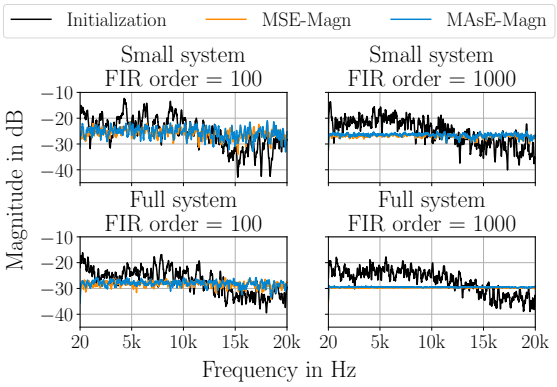
where $\|\mathbf{V}_\text{ML}\|_\text{F} = \sum_{i,j}\sum_k \mathbf{V}^2_{\text{ML}_{i,j}}[k]$ is the Frobenius norm of $\mathbf{V}_\text{ML}[n]$. The normalization employed in Eq. (7) makes the comparison fair since both the initialized and the optimized DSPs do not affect the average energy of the signal along the feedforward path of the feedback loop. Thus, the matrix $\mathbf{F}_\text{MM}$ has the same energy in both the initialized and the optimized case, and we can easily compare the difference in their respective GBI values.

4.2. Feedback loop flatness

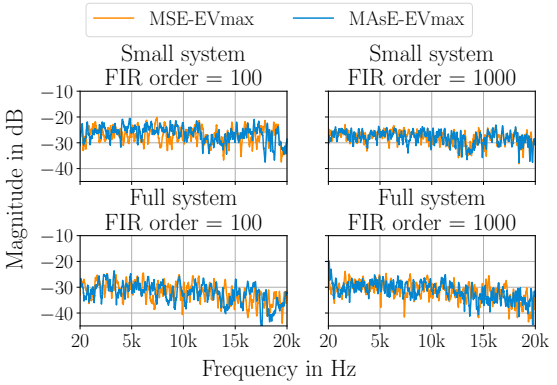
To assess the algorithm's ability to flatten the feedback loop TFs, we compared the magnitude responses of the $\mathbf{F}_\text{MM}(e^{\jmath\omega_k})$ elements before and after the training. $\mathbf{F}_\text{MM}(e^{\jmath\omega_k})$ was computed in the frequency domain as in Eq. (1), where $\mathbf{H}_\text{LM}(e^{\jmath\omega_k})$ and $\mathbf{V}_\text{ML}(e^{\jmath\omega_k})$ were obtained with a DFT of 48k frequency points in the range $[0, \pi]$ rad. For better visualization, the curves in Figs. 3 and 4 were smoothed through an average pooling with a kernel size of 256 samples.

Figure 3 shows the comparison between a DSP of non-trained, randomized FIR filters (in black) and a trained DSP (in orange) for the *small system* setup. Subplot titles indicate which TFs are illustrated in the respective panes. The training was conducted using MSE-Magn to maximize the curve flattening. The trained DSP produces much flatter magnitude responses than the non-trained DSP. Additionally, the optimized magnitude responses display similar values across all loop paths, meaning that the energy is uniformly distributed among the elements of $\mathbf{F}_\text{MM}(e^{\jmath\omega_k})$.

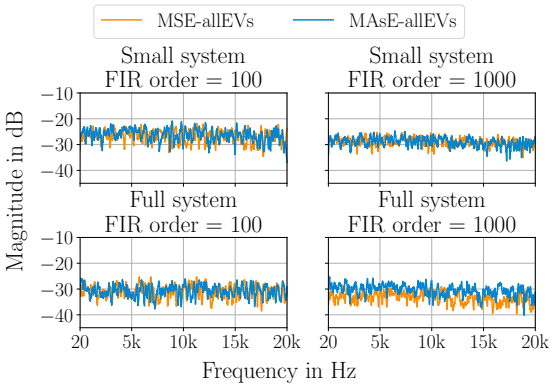
Figure 4 compares the magnitude responses for the $(1, 1)$ element of $\mathbf{F}_\text{MM}(e^{\jmath\omega_k})$ (*cf.* top-left pane of Fig. 3) after training the DSP with all proposed losses. The results are presented for both



(a) Optimization obtained using MSE-Magn (orange) and MAsE-Magn (blue) loss functions. The reference (black) is obtained with the non-trained DSP.



(b) Optimization obtained using MSE-maxEV (orange) and MAsE-maxEV (blue) loss functions.



(c) Optimization obtained using MSE-allEVs (orange) and MAsE-allEVs (blue) loss functions.

Figure 4: Comparison of the magnitude response of element (1,1) of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ after optimization. In each pane, the top row shows results for the *small system*, the bottom row for the *full system*, the left column for FIRs of order 100, and the right column for FIRs of order 1000.

the *small system* and the *full system* and FIRs of orders 100 and 1000. All magnitude responses in Fig. 4a are relatively flat and exhibit uniform energy distribution across the analyzed frequencies. The optimization algorithm performance increases with the growing filter order and number of transducers, but the results are similar regardless of the loss used during training.

Table 1: Mean and standard deviation values in dB of the magnitude response of element (1,1) of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ of the *small system*. The best values in each column are marked in bold font.

Condition	Small system			
	order 100		order 1000	
	Mean	Std. dev.	Mean	Std. dev.
Initialization	-26.63	7.076	-26.74	7.369
MSE-Magn	-26.49	**4.687**	-27.16	**2.762**
MAsE-Magn	-26.30	4.920	-26.85	2.786
MSE-EVmax	**-27.88**	5.651	-29.26	5.532
MAsE-EVmax	-27.53	5.547	-29.37	5.432
MSE-allEVs	-27.59	5.437	**-29.72**	5.333
MAsE-allEVs	-27.41	5.445	-29.53	5.275

Table 2: Mean and standard deviation values in dB of the magnitude response of element (1,1) of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ of the *full system*. The best values are marked in bold font.

Condition	Full system			
	order 100		order 1000	
	Mean	Std. dev.	Mean	Std. dev.
Initialization	-28.24	6.776	-28.38	6.792
MSE-Magn	-28.81	**3.147**	-29.70	**0.558**
MAsE-Magn	-28.47	3.214	-29.52	0.778
MSE-EVmax	-33.01	5.818	-33.56	5.995
MAsE-EVmax	**-33.34**	5.865	-33.28	6.037
MSE-allEVs	-31.72	5.618	**-34.15**	5.471
MAsE-allEVs	-31.58	5.539	-31.35	5.411

Figures 4b and 4c show the evaluation of the remaining losses: MSE-EVmax, MAsE-EVmax, MSE-allEVs, and MAsE-allEVs. The resulting magnitude curves are not as flat as those obtained with MSE-Magn and MAsE-Magn. In general, neither the filter order nor the number of transducers seem to affect the results significantly. For the *full system* in Fig. 4b, the frequencies above 15 kHz have a visibly lower magnitude than the rest. Such behaviour does not appear, however, in Fig. 4c.

In Figs. 4b and 4c the optimization results are similar regardless of the loss function used during training. However, for *full system* with the 1000-order FIR in the bottom-right pane of Fig. 4c the average magnitude obtained with MAsE-allEVs is almost 3 dB higher than the respective value for MSE-allEVs. The same behavior was discovered for all 16 elements of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$. The explanation for such a result is discussed in Sec. 4.3.

The statistics on magnitude flatness are gathered in Tables 1 and 2 for the *small system* and *full system*, respectively. The standard deviations further confirm that the optimization successfully flattens the magnitude responses of the system. The MSE-Magn loss performs best in this task, especially for the FIR order 1000. The system optimized on eigenvalue-based losses generally displays higher standard deviations than when magnitude-based losses are used. This is, however, expected in a magnitude-oriented task.

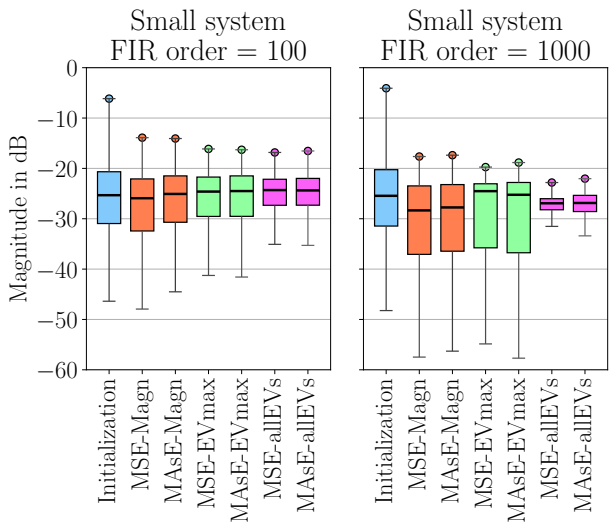


Figure 5: Eigenvalue magnitude distribution for *small system* configuration.

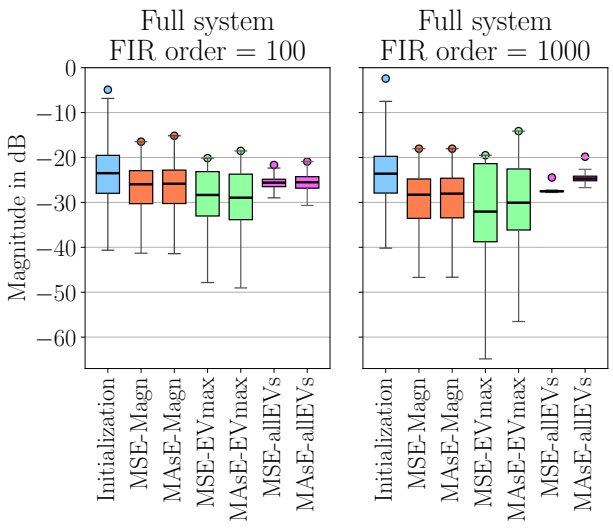


Figure 6: Eigenvalue magnitude distribution for the *full system*.

The mean values of the magnitude in Tables 1 and 2 are lower when the loss is computed from the system eigenvalues, with EVmax-based losses being the most beneficial for the low FIR order and MSE-allEVs for the high FIR order. This may allow more amplification of the reverberation produced with the AA system before reaching coloration or instability.

4.3. System eigenvalues

In this section, we assess the ability of the optimization algorithm to narrow the eigenvalue magnitude distribution. The magnitude computations were conducted as described in Sec. 4.2. The eigenvalues of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ for each frequency bin were then obtained through Eq. (2).

The results are presented in Figs. 5 and 6. The convention used for the boxplots represents the median with the central mark, while the whiskers correspond to the minimum between the extreme value and 1.5 times the interquartile range. In both figures, the maximum eigenvalue for each condition is shown and represented as a colored dot.

Figure 5 shows the *small system*'s eigenvalue magnitude distribution obtained with the DSP of orders 100 (left) and 1000 (right). The comparison is between the non-trained DSP (in blue) and the DSP trained with all the loss functions described in Sec. 3.2: MSE-Magn and MAsE-Magn (in orange), MSE-EVmax and MAsE-EVmax (in green), MSE-allEVs and MAsE-allEVs (in purple).

MSE-Magn and MAsE-Magn lowered the magnitude of the eigenvalue set but changed the shape of the eigenvalue magnitude distribution only marginally: the maximum—neglecting outliers—and the median are both closer to the third quartile, but the rest of the distribution remains unchanged. Increasing the FIRs order did not provide any substantial difference for the eigenvalues above the median, but broadened the distribution of the eigenvalues below the median. MSE-EVmax and MAsE-EVmax provided similar results to MSE-Magn and MAsE-Magn. In the case of low FIR order the obtained eigenvalues magnitude distribution is slightly narrower, whereas in the case of high FIR order the only visible difference is that the median is closer to the third quartile. MSE-allEVs and MAsE-allEVs narrowed the eigenvalue magnitude distribution considerably. For FIRs of order 100, the difference between the losses targeting the maximum eigenvalue curve and the losses targeting the whole eigenvalue set is subtle. MSE-allEVs and MAsE-allEVs provide a slightly better distribution than MSE-EVmax and MAsE-EVmax For FIRs of order 1000, the effect of narrowing the distribution is much more prominent when all the system's eigenvalues are considered. In this case, MSE-allEVs provided the best eigenvalues distribution.

Figure 6 shows the eigenvalue distribution comparison for the case of the *full system* configuration. In contrast to the *small system*, MSE-Magn and MAsE-Magn produced a visibly narrower eigenvalue magnitude distribution with respect to the non-trained DSP, especially in the case of the low FIR order. For MSE-Magn and MAsE-Magn, increasing the order of the FIRs broadened the distribution but lowered the magnitude values. MSE-Magn and MAsE-Magn also provided narrower eigenvalue magnitude distributions than MSE-EVmax and MAsE-EVmax. This is especially true for FIR of order 1000.

Figure 6 shows considerable differences between MSE-EVmax, MAsE-EVmax, MSE-allEVs, and MAsE-allEVs. With FIRs of order 100, MSE-EVmax and MAsE-EVmax decreased the eigenvalues' magnitude values, but the distribution was slightly wider than when MSE-Magn and MAsE-Magn were used. The distributions of MSE-allEVs and MAsE-allEVs are, instead, much narrower for short FIRs. With FIRs of order 1000, MSE-EVmax and MAsE-EVmax broadened the eigenvalue magnitude distribution with respect to initialization, with the only good outcome being that the eigenvalue set has lower magnitude values and the upper whisker is closer to the upper quartile. In the case of MSE-allEVs and MAsE-allEVs, the distributions converged to a very narrow magnitude interval, and these are the only two cases in which a loss function provided an outlier that does not coincide with the upper whisker limit. Again, MSE-allEVs produced the best results. The right pane of Fig. 6 shows that almost all of the eigenvalue magnitude distribution obtained via the MSE-allEVs loss is lower in magnitude than the distribution obtained via the MAsE-allEVs

loss. This explains the difference in the average magnitude shown in the bottom-right pane of Fig. 4c.

Figures 5 and 6 show that all loss functions have reduced the eigenvalues' magnitudes with respect to the non-trained DSP cases. This reflects the flattening of the magnitude responses of $\mathbf{F}_{\mathrm{MM}}(e^{\jmath\omega_k})$ described in Sec. 4.2, and the corresponding better distribution of the feedback loop's energy across frequencies. This is positive because—assuming that the stability condition is met, and recalling that all the DSPs were normalized—the gain in the processing path is further away from the GBI for the trained DSP than the non-trained DSP. Thus, the trained system can introduce more energy into the room than the non-trained one before generating sound coloration or reaching the instability limit. The lowest and the highest top outlier magnitude decrements obtained were 7.75 dB for the *small system* and the FIR of order 100 with MSE-Magn loss and 22.0 dB for the *full system* and the FIR of order 1000 with the MSE-allEVs loss, respectively.

4.4. Inline system

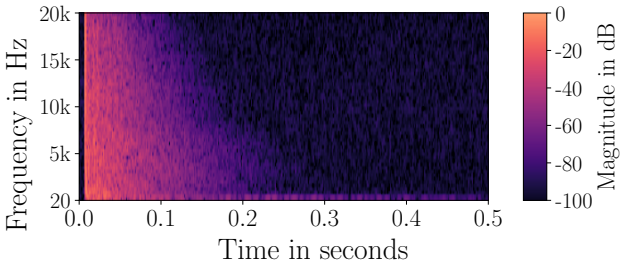
In this section, we consider the optimization of short FIR filters inside the DSP of an inline system. The DSP also includes a fixed artificial reverberator designed to be exponentially decaying white Gaussian noise with a RT of 1 second. The role of the artificial reverberator is to lengthen the sound energy decay, whereas the role of the FIR filters is to improve system stability. The stability analysis and the PyTorch pipeline remain unaltered, although the definition of the DSP changes:

$$\mathbf{V}_{\mathrm{ML}}(e^{\jmath\omega_k}) = \mathbf{R}_{\mathrm{LL}}(e^{\jmath\omega_k})\mathbf{U}_{\mathrm{ML}}(e^{\jmath\omega_k}), \tag{8}$$

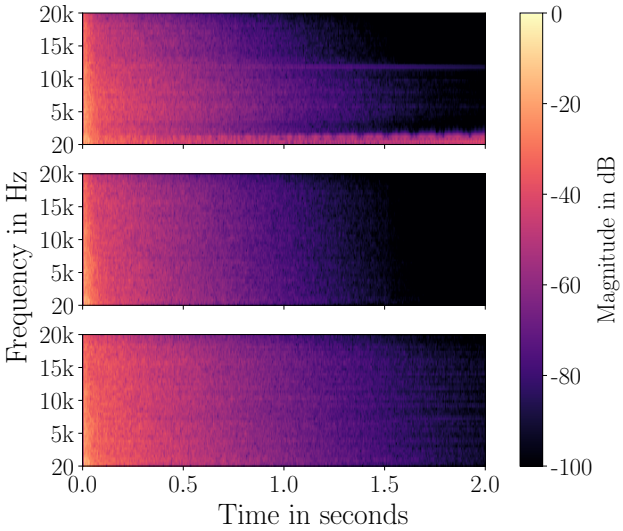
where $\mathbf{U}_{\mathrm{ML}}(e^{\jmath\omega_k})$ is the matrix of the FIR filters and $\mathbf{R}_{\mathrm{LL}}(e^{\jmath\omega_k})$ is the fixed diagonal matrix of artificial reverberators. The order of the learnable FIR filters is 100 and the DSP is optimized through the MSE-allEVs loss function. We simulated the full scenario depicted in Fig. 1. To obtain the RIRs $\mathrm{H}_{\mathrm{SA}}[n]$, $\mathbf{H}_{\mathrm{SM}}[n]$, and $\mathbf{H}_{\mathrm{LA}}[n]$ we included an additional microphone—Behringer type ECM8000—, to simulate a listening position in the audience, and one of the systems' loudspeakers served as the sound source.

Figure 7 shows the spectrogram of the signal $y[n]$ at the audience microphone when the source speaker is fed with an impulse at time sample $n = 0$ for different system conditions. Fig. 7a is the signal received when the AA system is turned off, thus corresponding to $\mathrm{H}_{\mathrm{SA}}[n]$, showing the acoustical properties of the physical room. The RT is 0.3 seconds, *cf.* Sec. 4.1. In Fig. 7b, the top pane is the response obtained when the AA system with a non-trained DSP is turned on and the gain G set just above the GBI. Instability below 500 Hz and a strong ringing tone at 12 kHz are visible. The center pane is the response obtained by keeping G fixed but employing the trained DSP. The optimization successfully removed instability and the long-ringing tone and distributed the energy more evenly across the entire frequency range, reducing coloration. The bottom pane is the response when the AA system with the trained DSP is turned on with the gain G incremented by +6 dB. The system is still stable, showing no discrepancy between the decay times at different frequencies. Audio examples and configuration details are available online [1].

[1] http://research.spa.aalto.fi/publications/papers/dafx24-diff-aa/



(a) AA system turned off.



(b) Non-trained DSP (top), trained DSP (center) and trained DSP with an increased gain value (bottom).

Figure 7: Spectrograms of the signal $y[n]$ registered at the audience position for multiple AA system configurations.

5. CONCLUSIONS

The presented work introduced the use of a DDSP framework in the AA context. The proposed optimization retains the AA system's flexibility and allows for automatic tuning of its parameters to obtain a flat magnitude response of the feedback loop and a narrow eigenvalue magnitude distribution. Specifically, the method focuses on adjusting the coefficients of the FIR filters in the DSP matrix.

We evaluated the proposed framework using multiple transducer setups, filter orders, and loss functions. Results show that increasing the filter order and the number of transducers improves the optimization performance. In our experiments, we obtained flat magnitude responses from the feedback loop TFs with a standard deviation < 1 dB. In this task, using a mean squared error between the feedback loop magnitude and a target performed the best of all the analyzed loss functions.

In terms of the system's eigenvalues, we also obtained narrow distributions with a standard deviation < 0.5 dB, neglecting outliers. This time, minimizing the mean squared error of all of the eigenvalues proved to be the best loss function.

This work demonstrated the effectiveness of DDSP in the AA scenario, providing a starting point for future improvements.

6. ACKNOWLEDGMENTS

The work of the first author was funded by the Academy of Finland, project N° 357391. The work of the second author was funded by the Aalto University School of Electrical Engineering.

7. REFERENCES

[1] P. U. Svensson, *On Reverberation Enhancement in Auditoria*, Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, Nov. 1994.

[2] M. Poletti, "Active acoustic systems for the control of room acoustics," *Build. Acoust.*, vol. 18, no. 3–4, pp. 237–258, 2011.

[3] S. J. Schlecht and E. A. Habets, "Reverberation enhancement systems with time-varying mixing matrices," in *Proc. 59th Int. Audio Eng. Soc. Conf.: Sound Reinforcement Engineering and Technology*, 2015.

[4] M. Poletti, "The control of early and late energy using the variable room acoustics system," in *Proc. Acoustics*, 2006, pp. 20–22.

[5] P. Coleman, N. Epain, S. Venkatesh, and F. Roskam, "Exploring perceptual annoyance and colouration assessment in active acoustic environments," in *Proc. Audio Eng. Soc. Int. Conf. on Acoustics & Sound Reinforcement*, 2024.

[6] H. Nyquist, "Regeneration theory," *Bell System Technical Journal*, vol. 11, no. 1, pp. 126–147, 1932.

[7] R. V. Waterhouse, "Theory of howlback in reverberant rooms," *J. Acoust. Soc. Am.*, vol. 37, no. 5, pp. 921–923, 1965.

[8] M. R. Schroeder, "Improvement of acoustic-feedback stability by frequency shifting," *J. Acoust. Soc. Am.*, vol. 36, no. 9, pp. 1718–1724, 1964.

[9] C. P. Boner and C. R. Boner, "Behavior of sound system response immediately below feedback," *J. Audio Eng. Soc.*, vol. 14, no. 3, pp. 200–203, 1966.

[10] P. U. Svensson, "Influence of electroacoustic parameters on the performance of reverberation enhancement systems," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 162–171, 1993.

[11] W. K. Connor, "Experimental investigation of sound-system-room feedback," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 27–32, 1973.

[12] W. K. Connor, "Theoretical and practical considerations in the equalization of sound systems," *J. Audio Eng. Soc.*, vol. 15, no. 2, pp. 194–198, 1967.

[13] C. P. Boner and C. R. Boner, "A procedure for controlling room-ring modes and feedback modes in sound systems with narrow-band filters," in *Proc. Audio Eng. Soc. 16th Conv.*, 1964.

[14] C. P. Boner and C. R. Boner, "Minimizing feedback in sound systems and room-ring modes with passive networks," *J. Acoust. Soc. Am.*, vol. 37, no. 1, pp. 131–135, 1965.

[15] P. Gil-Cacho, T. Van Waterschoot, M. Moonen, and S. H. Jensen, "Regularized adaptive notch filters for acoustic howling suppression," in *Proc. 17th European Signal Process. Conf.*, 2009, pp. 2574–2578.

[16] J. S. Abel, E. F. Callery, and E. K. Canfield-Dafilou, "A feedback canceling reverberator," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Aveiro, Portugal, Sep. 2018, pp. 100–106.

[17] T. Van Waterschoot and M. Moonen, "Adaptive feedback cancellation for audio applications," *Signal Processing*, vol. 89, no. 11, pp. 2185–2201, 2009.

[18] U. P. Svensson, "Computer simulations of periodically time-varying filters for acoustic feedback control," *J. Audio Eng. Soc.*, vol. 43, no. 9, pp. 667–677, 1995.

[19] J. L. Nielsen and U. P. Svensson, "Performance of some linear time-varying systems in control of acoustic feedback," *J. Acoust. Soc. Am.*, vol. 106, no. 1, pp. 240–254, 1999.

[20] S. J. Schlecht and E. A. Habets, "The stability of multichannel sound systems with time-varying mixing matrices," *J. Acoust. Soc. Am.*, vol. 140, no. 1, pp. 601–609, 2016.

[21] M. Poletti, "A unitary reverberator for reduced colouration in assisted reverberation systems," in *Proc. INTER-NOISE and NOISE-CON Congr. Conf.*, 1995, vol. 5, pp. 1223–1232.

[22] T. Lokki and J. Hiipakka, "A time-variant reverberation algorithm for reverberation enhancement systems," in *Proc. COST G-6 Conf. Digital Audio Effects (DAFX-01)*, Limerick, Ireland, 2001, pp. 28–32.

[23] T. Van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, 2010.

[24] F. Kaiser, C. Frischmann, V. Werner, and T. Rohde, "Room acoustics evaluation of active acoustics systems – Results from measurements," in *Proc. Int. Symp. Room Acoustics*, Sep. 2019.

[25] V. Werner, S. Neeten, and F. Kaiser, "Evaluation of a geometric approach to active acoustics," in *Proc. Inst. Acoust.*, 2023, vol. 45, Pt. 2.

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

[27] B. Kuznetsov, J. D. Parker, and F. Esqueda, "Differentiable IIR filters for machine learning applications," in *Proc. Int. Conf. Digital Audio Effects (eDAFx-20)*, 2020, pp. 297–303.

[28] M. Poletti, "The stability of single and multichannel sound systems," *Acta Acust. un. Acust.*, vol. 86, no. 1, pp. 163–178, 2000.

[29] G. Dal Santo, K. Prawda, S. J. Schlecht, and V. Välimäki, "Feedback delay network optimization," *arXiv preprint arXiv:2402.11216*, 2024.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[31] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. Audio Eng. Soc. Conv. 108*, 2000.