MODELING THE IMPULSE RESPONSE OF HIGHER-ORDER MICROPHONE ARRAYS USING DIFFERENTIABLE FEEDBACK DELAY NETWORKS

Riccardo Giampiccolo [1,*], *Alessandro Ilic Mezza* [1], *Mirco Pezzoli* [1], *Shoichi Koyama* [2], *Alberto Bernardini* [1], *and Fabio Antonacci* [1,†]

[1] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
[2] National Institute of Informatics, Tokyo, Japan
riccardo.giampiccolo@polimi.it | alessandroilic.mezza@polimi.it | mirco.pezzoli@polimi.it
shoichi.koyama@ieee.org | alberto.bernardini@polimi.it | fabio.antonacci@polimi.it

ABSTRACT

Recently, differentiable multiple-input multiple-output Feedback Delay Networks (FDNs) have been proposed for modeling target multichannel room impulse responses by optimizing their parameters according to perceptually-driven time-domain descriptors. However, in spatial audio applications, frequency-domain characteristics and inter-channel differences are crucial for accurately replicating a given soundfield. In this article, targeting the modeling of the response of higher-order microphone arrays, we improve on the methodology by optimizing the FDN parameters using a novel spatially-informed loss function, demonstrating its superior performance over previous approaches and paving the way toward the use of differentiable FDNs in spatial audio applications such as soundfield reconstruction and rendering.

1. INTRODUCTION

The modeling of Room Impulse Responses (RIRs) plays a key role in a wide range of spatial audio applications, from immersive AR/VR to realistic teleconferencing and spatialized music reproduction systems [1]. Higher-Order Microphone (HOM) arrays have gained increasing popularity in these scenarios due to their ability to capture spatial soundfields with high fidelity [2]. However, accurately modeling their impulse responses to be used in applications of interest remains a challenging task. Indeed, conventional methods [3], such as Spatial Impulse Response Rendering (SIRR) [4] and its extension, Higher-Order-SIRR (HO-SIRR) [5], often rely on the spherical harmonic (SH) decomposition in the time-frequency domain, which, being computationally demanding, poses significant limitations in scenarios for which efficiency is of paramount importance [5].

Lately, differentiable digital signal processing has opened new frontiers in data-driven acoustic modeling, offering promising alternatives to traditional approaches [6–9]. In particular, Feedback Delay Networks (FDNs), first introduced by Gerzon in the 1970s and generalized to the multichannel case by Stautner and Puckette [10], have been recently incorporated into differentiable learning frameworks to automatically optimize their parameters, making them a powerful tool for modeling reverberant spaces [7,8,11]. However, while differentiable FDNs have been successfully employed for single-input single-output (SISO) scenarios, e.g., matching the time-frequency energy decay of target RIRs [8, 12], their application to multiple-output FDNs remains largely unexplored. For instance, in [13], the FDN parameters are optimized only to match the energy decay in the time domain, without conditioning the training on frequency-domain or spatial descriptors, thus hindering its ability to model the spatial cues of the impinging soundfield.

In this work, we consider the scenario of a single source impinging on the HOM. We propose a novel loss function that allows us to optimize the parameters of differentiable single-input multiple-output (SIMO) FDNs in order to recreate a target soundfield, matching the energy distribution both in space and time-frequency domains. Although different FDN prototypes with attenuation, tone, or directional filters exist [12,14–16], we consider the original "general" FDN [13, 17] as to demonstrate that it already owns the potentiality to model target soundfields. We introduce a novel loss term accounting for the inter-channel level difference [18] of the first-order Ambisonic representations of the signals. Also, we propose novel terms for improving the energy matching, ultimately regularizing the peak amplitude and improving the inter-channel correlation [19]. We evaluate the proposed approach by modeling the response of a HOM from the HOMULA-RIR corpus [20], outperforming state-of-the-art methodologies for the optimization of differentiable FDNs. In addition to the evaluation through objective metrics, we conduct a perceptual test on spatial quality. The results support our analysis and further confirm the effectiveness of our method, highlighting its potential for spatial audio applications such as soundfield reconstruction and real-time spatial auralization.

2. DIFFERENTIABLE MIMO FEEDBACK DELAY NETWORKS

Let us consider the multiple-input multiple-output (MIMO) FDN prototype shown in Fig. 1, which was shown to be suitable for optimization in [13]. With N being the number of delay lines, $\mathbf{u}[n] \in \mathbb{R}^I$ the vector of input signals, and $\mathbf{y}[n] \in \mathbb{R}^J$ the vector of output signals, the FDN can be described by means of the following equations

$$\begin{aligned} \mathbf{y}[n-\boldsymbol{\mu}] &= \mathbf{G}\left(\mathbf{C}\,\mathbf{s}[n] + \mathbf{D}\,\mathbf{u}[n]\right), \\ \mathbf{s}[n+\boldsymbol{m}] &= \mathbf{A}\,\mathbf{s}[n] + \mathbf{B}\,\mathbf{u}[n], \end{aligned} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the feedback matrix, $\mathbf{B} \in \mathbb{R}^{N\times I}$ is the input gain matrix, $\mathbf{C} \in \mathbb{R}^{J\times N}$ is the output gain matrix, $\mathbf{D} \in \mathbb{R}^{J\times I}$

* Corresponding author.

† This work was partially supported by the European Union – Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP D43C22003080001, partnership on "Telecommunications of the Future" (PE00000001 – program "RESTART")

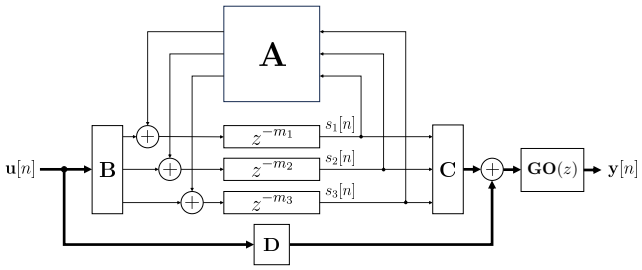


Figure 1: *MIMO FDN with three delay lines ($N = 3$).*

is the direct gain matrix, and $\mathbf{s}[n] \in \mathbb{R}^N$ is the output of the delay lines at time index n. $\mathbf{G} \in \mathbb{R}^{J\times J}$ is a diagonal matrix containing real scaling parameters and $\boldsymbol{\mu} := [\mu_1, \ldots, \mu_J]^{\mathrm{T}}$ is a vector containing the line-of-sight delays, which, in our work, we consider fractional [8, 13]. Referring to Fig. 1, it follows that $\mathbf{O}(z) = \mathrm{diag}([z^{-\mu_1}, \ldots, z^{-\mu_J}])$. Finally, denoting the lengths of the delay lines in samples as $\boldsymbol{m} = [m_1, ..., m_N]^{\mathrm{T}}$, we obtain $\mathbf{s}[n + \boldsymbol{m}] := [s_1[n + m_1], ..., s_N[n + m_N]]^{\mathrm{T}}$.

Unitary matrices, such as Hadamard or Householder matrices, are commonly used as prototypes for the feedback matrix $\mathbf{A}$ [21]. Being *unilossless*, in fact, they guarantee stability regardless of the delays in the FDN [22]. Then, with the aim of introducing losses, the feedback matrix is multiplied by a diagonal matrix containing scalar values designed to achieve a specific reverberation time [7].

Although prototypes that include tone correction and attenuation filters are common in the literature [12, 14, 15], in this article, we focus on FDNs characterized by frequency-independent parameters, i.e., the entries of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are scalars. As shown in [8, 12, 13], it follows that the stability of the system can be easily enforced at training time by applying proper reparameterization.

2.1. Differentiable Implementation

In this work, we focus on SIMO FDNs with the purpose of modeling the response of a Higher-Order Microphone (HOM) to a single impulse. This implies that matrices $\mathbf{B}$ and $\mathbf{D}$ are turned into vectors $\mathbf{b} \in \mathbb{R}^N$ and $\mathbf{d} \in \mathbb{R}^J$, respectively. Then, we take into account a particular implementation that allows us to learn $\mathbf{A} \in \mathbb{R}^{N\times N}$ and the nonnegative $\mathbf{b} \in \mathbb{R}^N_{\geq 0}$, $\mathbf{C} \in \mathbb{R}^{J\times N}_{\geq 0}$, $\boldsymbol{m} \in \mathbb{R}^N_{\geq 0}$, and $\mathbf{d} \in \mathbb{R}^J_{\geq 0}$ via standard backpropagation using gradient-based optimization methods [8, 13].

Considering as input the Kronecker delta $\delta[n]$, we aim indeed at minimizing at each time instant n a loss function $\mathcal{L}$ between the target J-channel RIR $\mathbf{h}[n] \in \mathbb{R}^J$ and the output of the differentiable SIMO FDN $\hat{\mathbf{h}}[n] \in \mathbb{R}^J$. At each epoch, the FDN parameters θ undergo an optimization step involving the gradient $\nabla\mathcal{L}_\theta$ computed via reverse-mode automatic differentiation [8, 23]. In particular, our differentiable implementation features:

Differentiable Delay Lines

Reference [8] introduced a method for implementing differentiable delay lines in the frequency domain. Specifically, the buffered signal is first zero-padded and transformed using the Fast Fourier Transform (FFT). Then, the resulting discrete spectrum is multiplied by a conjugate symmetric fractional delay filter response. Finally, the delayed signal is reconstructed in the time domain by performing an Inverse FFT. For a more detailed discussion of the implementation, we refer readers to [8].

Trainable Feedback Matrix

Rather than enforcing unilosslessness constraints on the feedback matrix itself, as in [13], we consider an unconstrained learnable matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ and define the lossy feedback matrix as $\mathbf{A} = \mathbf{U}\boldsymbol{\Gamma}$. Here, $\mathbf{U}$ is an orthogonal matrix, while $\boldsymbol{\Gamma}$ is a learnable diagonal attenuation matrix. We define $\mathbf{U} = \exp(\mathbf{W}_{\mathrm{Tr}} - \mathbf{W}_{\mathrm{Tr}}^{\mathrm{T}})$, where $\mathbf{W}_{\mathrm{Tr}}$ corresponds to the upper triangular portion of $\mathbf{W}$. Since exponentiation of skew-symmetric matrices ensures orthogonality, it follows that $\mathbf{U}$ is orthogonal by construction [7, 11]. Consequently, $\mathbf{U}$ is unilossless regardless of the values assigned to $\mathbf{W}$.

Differentiable Reparameterization

The trainable parameters are learned in an unconstrained fashion and then mapped onto the FDN parameters through differentiable functions. We parameterize $\boldsymbol{\Gamma} = \mathrm{diag}([g(\tilde{\gamma}_1), ..., g(\tilde{\gamma}_N)])$, where $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_N$ are unconstrained scalars, and $g(\cdot)$ is the logistic function. This transformation ensures that the attenuation coefficients γ_n remain within the interval $(0, 1)$. Additionally, we take the absolute value of the entries of $\mathbf{b}$, $\mathbf{C}$, $\mathbf{d}$, and $\boldsymbol{m}$ [8,24], such that, e.g., we have $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_J]$ with columns $\mathbf{c}_j = [|\tilde{c}_{j,1}|, ..., |\tilde{c}_{j,N}|]^{\mathrm{T}}$, where the symbol $\tilde{\cdot}$ indicates that the parameter is learned without constraints.

3. METHODOLOGY

In this section, we first review the learning objectives proposed in the literature for the optimization of FDNs given a target RIR. We then introduce a novel loss function designed to optimize SIMO FDNs so as to replicate spatial characteristics of target soundfields.

3.1. Related Work

Following [12], we can distinguish between two kinds of loss functions: frequency independent (FI) [8, 13] and frequency dependent (FD) [12]. The former one focuses solely on the time-domain characteristics of the FDN impulse response (IR), while the latter includes frequency-domain descriptors in the optimization objectives.

First proposed in [8] for optimizing SISO FDNs and then generalized to the MIMO case in [13], the FI loss entails an error on the energy decay curve (EDC), regularized by an error on the echo density profile (EDP). Indeed, named $\mathbf{h}[n] := [h_1[n], \ldots, h_J[n]]^{\mathrm{T}}$ the L_h-sample J-channel target RIR at time instant n, the FI loss function is defined as

$$\mathcal{L}_{\mathrm{FI}} = \mathcal{L}_{\mathrm{EDC}} + \lambda_0 \mathcal{L}_{\mathrm{EDP}}\,, \tag{2}$$

where $\lambda_0 \in \mathbb{R}_{>0}$ is a positive real-valued hyperparameter and

$$\mathcal{L}_{\mathrm{EDC}} = \frac{\sum_n \|\boldsymbol{e}[n] - \hat{\boldsymbol{e}}[n]\|_2^2}{\sum_n \|\boldsymbol{e}[n]\|_2^2}\,, \tag{3}$$

is the normalized L^2-loss between the multichannel EDC $\boldsymbol{e}[n] := [\varepsilon_1[n], \ldots, \varepsilon_J[n]]^{\mathrm{T}}$ of the target RIR and the EDC $\hat{\boldsymbol{e}}[n] := [\hat{\varepsilon}_1[h], \ldots, \hat{\varepsilon}_J[n]]^{\mathrm{T}}$ of the FDN IR, $\hat{\mathbf{h}}[n]$. The channel EDC is computed via Schroeder's backward integration as

$$\varepsilon_j[n] = \sum_{\tau=n}^{L_h} h_j^2[\tau] \tag{4}$$

and equivalently for $\hat{\varepsilon}_j[n]$.

Moreover, in (2), the EDP loss term is defined through the Soft EDP function [8], a differentiable approximation of the well-known normalized echo density profile [25],

$$\mathcal{L}_{\text{EDP}} = \frac{1}{L_h} \sum_n \|\boldsymbol{p}[n] - \hat{\boldsymbol{p}}[n]\|_2^2 \,, \tag{5}$$

where $\boldsymbol{p}[n] := [\eta_1[n], \ldots, \eta_J[n]]^{\mathrm{T}}$ is the target multichannel Soft EDP and $\hat{\boldsymbol{p}}[n] := [\hat{\eta}_1[h], \ldots, \hat{\eta}_J[n]]^{\mathrm{T}}$ is the Soft EDP of the predicted IR. In particular, $\eta_j[n]$ for a generic channel j is derived following the approach in [25], with the key modification that the nondifferentiable indicator function $\mathbb{1}\{\cdot\}$ is approximated by a scaled sigmoid function $g_\kappa(x) = g(\kappa x)$, $\kappa \gg 1$ [8].

Contrary to $\mathcal{L}_{\text{FI}}$, the FD loss function—proposed in [12] for the SISO case—has never been extended to account for multiple-output FDNs. Thus, we here generalize said learning objective to multichannel RIRs. In particular, along with the terms reported in (2), the FD loss function includes an additional error term based on the mel-scale energy decay relief (EDR). Namely,

$$\mathcal{L}_{\text{FD}} = \lambda_1 \mathcal{L}_{\text{EDC}} + \lambda_2 \mathcal{L}_{\text{EDR}} + \lambda_3 \mathcal{L}_{\text{EDP}}, \tag{6}$$

with $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}_{>0}$ and

$$\mathcal{L}_{\text{EDR}} = \frac{\sum_k \sum_m \|\boldsymbol{r}[k,m] - \hat{\boldsymbol{r}}[k,m]\|_1}{\sum_k \sum_m \|\boldsymbol{r}[k,m]\|_1} \tag{7}$$

being the normalized L^1-loss between the mel-scale EDR of the target multichannel RIR measured in dB

$$\boldsymbol{r}[k,m] = [\mathcal{R}^{\text{dB}}_{\text{mel},1}[k,m], \ldots, \mathcal{R}^{\text{dB}}_{\text{mel},J}[k,m]]^{\mathrm{T}} \tag{8}$$

and the mel-scale EDR of the predicted IR

$$\hat{\boldsymbol{r}}[k,m] = [\hat{\mathcal{R}}^{\text{dB}}_{\text{mel},1}[k,m], \ldots, \hat{\mathcal{R}}^{\text{dB}}_{\text{mel},J}[k,m]]^{\mathrm{T}} \,. \tag{9}$$

In (8), the mel-scale EDR for channel j is computed via backward integration of $|\mathcal{H}_{\text{mel},j}[k,m]|^2$, i.e., the mel-spectrogram of $h_j[n]$. Specifically, $\mathcal{R}^{\text{dB}}_{\text{mel},j}[k,m]$ is defined as

$$\mathcal{R}^{\text{dB}}_{\text{mel},j}[k,m] = 10 \log_{10} \sum_{\tau=m}^{M} |\mathcal{H}_{\text{mel},j}[k,\tau]|^2 \,. \tag{10}$$

Finally, an analogous equation can be derived for $\hat{\mathcal{R}}^{\text{dB}}_{\text{mel},j}[k,m]]^{\mathrm{T}}$.

3.2. Proposed Method

The learning objectives presented in the previous subsection are mainly focused on matching the sound energy decay both in the time and frequency domains. However, they do not account for the distribution of the energy in space, which is essential in spatial audio applications. Therefore, we present a novel training objective that allows us to optimize the SIMO FDN parameters as to match the spatial characteristics of the target multichannel RIR.

Planar Inter-Channel Level Difference Loss

Inspired by [18], we propose to compute an error on the Inter-Channel Level Difference (ICLD) defined starting from the First-Order Ambisonic (FOA) signals. Proposed by Gerzon [26, 27] and lately further improved by Fellgett [28], Ambisonics can be thought of as a three-dimensional extension of the mid/side stereo technique, encoding a given soundfield into a set of spherical harmonic (SH) coefficients. In particular, given the set of azimuth $\theta \in [-\pi, \pi]$ and elevation $\phi \in [-\pi/2, \pi/2]$ angles describing the orientation of the HOM capsules, the FOA representation of the multichannel RIR $\mathbf{h}[n]$ is a signal $\mathbf{s}[n]$ consisting of four channels: the W channel, corresponding to the sound pressure acquired by an omnidirectional microphone, and the X, Y and Z channels, corresponding to the sound pressures acquired by a figure-of-eight microphone oriented along the three spatial axes. In this scenario, we define the ICLD as

$$\mathcal{D}_{XY} = \frac{1}{K} \sum_{k=0}^{K-1} \left(|S_X[k]| - |S_Y[k]| \right)^2 \,, \tag{11}$$

where S_X and S_Y are the Discrete Fourier Transforms (DFTs) of X- and Y-channel, respectively, whereas K is the total number of frequency bins. Then, we define the ICLD loss term as

$$\mathcal{L}_{\text{ICLD}} = \frac{(\mathcal{D}_{XY} - \hat{\mathcal{D}}_{XY})^2}{\mathcal{D}_{XY}^2} \,, \tag{12}$$

where $\hat{\mathcal{D}}_{XY}$ is the ICLD computed on the predicted IR. Here, we take into account only the planar inter-channel level difference as, from a perceptual standpoint, the spatial resolution of the human auditory system is significantly higher in the azimuthal direction, with minimum audible angle thresholds for changes in elevation typically being two to four times greater than those measured in the horizontal plane when broadband stimuli are used [29].

Multi-Resolution EDR Loss

We combine $\mathcal{L}_{\text{ICLD}}$ with the loss terms involving the energy decay outlined in Section 3.1. In particular, to better match the coloration of the target RIR and mitigate the typical comb-like artifacts characterizing FDN IRs [7, 11], we propose to consider a loss function involving the multi-resolution (MR) mel-scale EDR, defined as

$$\mathcal{L}_{\text{MR-EDR}} = \frac{1}{R} \sum_r \mathcal{L}^{(r)}_{\text{EDR}} \,, \tag{13}$$

where $r = 1, ..., R$ is the index of a particular *resolution*, i.e., the combination of FFT length, window length, and hop size chosen for computing the mel-spectrograms in (10).

Time-Domain Envelope Loss

Now, if we consider to add all the learning objectives presented so far, we would obtain a composite loss function with four terms (4T). The minimization of said loss could be impaired since the presence of an additional objective may introduce conflicting gradients, leading to convergence difficulties and suboptimal solutions. We propose, thus, to substitute the $\mathcal{L}_{\text{EDC}}$ and $\mathcal{L}_{\text{EDP}}$ terms with a single loss term involving the envelope of the IR energies. Indeed, we argue that such a signal comprises information on both time-domain energy and peak amplitudes, being thus a valid substitute of the original terms. We define this novel loss function as

$$\mathcal{L}_{\text{ENV}} = \frac{\sum_n \|\mathbf{h}_{\text{f}}^2[n] - \hat{\mathbf{h}}_{\text{f}}^2[n]\|_2^2}{\sum_n \|\mathbf{h}_{\text{f}}^2[n]\|_2^2} \,, \tag{14}$$

with $\mathbf{h}_{\text{f}}^2[n] := [h_{\text{f},1}^2[n], \ldots, h_{\text{f},J}^2[n]]^{\mathrm{T}}$ and $\hat{\mathbf{h}}_{\text{f}}^2[n] := [\hat{h}_{\text{f},1}^2[n], \ldots, \hat{h}_{\text{f},J}^2[n]]^{\mathrm{T}}$, where the energies of the multichannel RIR and the predicted IR, respectively, are processed to extract the time-domain

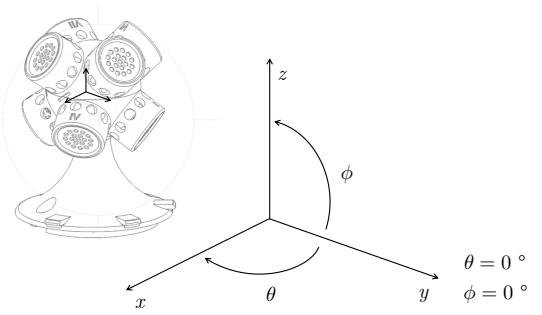


Figure 2: *Spherical coordinate system considered in this study.*

envelope

$$h^2_{\mathrm{f},j}[n] = \sum_{\ell=0}^{15} \bar{\beta}_\ell \max\{h^2_j[n-\ell],\ \bar{\alpha} h^2_j[n-\ell-1]\}, \qquad (15)$$

where $\bar{\beta}_\ell$ are the taps of a (fixed) linear-phase FIR lowpass filter, and $\bar{\alpha} = 0.96$ controls the slope of the envelope and allows for a certain amount of leniency in the penalization of less prominent IR taps. By doing so, we maintain the improvement given by the well-established practice of combining linear-scale L^2-losses to log-scale L^1-losses that has been found beneficial in many audio signal processing applications [12]. Indeed, as shown in [12], a loss term computed on a linear scale places the focus on the first portion of the IRs, while a loss term computed on a dB scale emphasizes errors in the reverberation tail due to the logarithmic compression, improving the overall optimization.

Composite Loss Function

Finally, with the aim of optimizing both energy and inter-channel differences, we define the the following composite loss function

$$\mathcal{L}_{\text{SP}} = \lambda_4 \mathcal{L}_{\text{ICLD}} + \lambda_5 \mathcal{L}_{\text{MR-EDR}} + \lambda_6 \mathcal{L}_{\text{ENV}}\,, \qquad (16)$$

with λ_4, λ_5, and λ_6 being positive real-valued hyperparameters.

4. EVALUATION

We evaluate the proposed methodology considering the response of a HOM from the HOMULA-RIR dataset [20] as the target. Such a corpus contains multichannel RIRs acquired in 25 positions inside the "Schiavoni" seminar room of Politecnico di Milano. In particular, the authors employed the Voyage Audio Spatial Mic, which is a 2nd-order Ambisonic microphone able to record the entire soundfield with 8 capsules mounted together in a near-coincident array. Fig. 2 shows the spherical coordinate convention considered in this work. Among the source signals, we select `S1` and, as HOM, we select `R1-HOM5` [20].

As far as the training procedure is concerned, we first normalize the target multichannel RIR as to have unitary norm, and we store such a value in matrix $\mathbf{G}$ such that $\mathbf{G} = g\mathbf{I}_J$, with $\mathbf{I}_J$ being an $J \times J$ identity matrix. According to (1), this matrix will be later used to re-scale the output of the SIMO FDN.

As in [13], with the purpose of ensuring the first sample of the jth channel to always contain the direct path, we remove the first μ_j samples. These values are stored in vector $\boldsymbol{\mu}$ so that they can be reintroduced at inference time through matrix $\mathbf{O}(z)$, as shown in Fig. 1. To maintain consistency, we apply zero-padding so that each channel has the same number of samples, denoted as L_x. This value thus corresponds to the number of samples of the IR in which the direct arrives first, and it is computed as $L_x = L_j - \mu_{\min}$, where $\mu_{\min} = \min\{\mu_j\}_{j=1}^J$ and L_j denotes the original length of the jth channel in samples. Then, we compute the reverberation times $T_{60,j}$ for each channel j using `pyroomacoustics` [30] and define $T_{60}^{\max} = \max\{T_{60,j}\}_{j=1}^J$. As a final step, we trim all channels to the length $L_h = \lceil T_{60}^{\max} \cdot f_{\text{s}} \rceil$, where f_{s} is the sampling frequency. Indeed, any information beyond $T_{60}^{\max}$ could negatively affect the training process as it could introduce numerical instabilities, potentially leading to unwanted distortions in the reverberant tails [8].

Table 1: *Diffuseness ψ and direction of arrival error (DOAE) ϑ_e for the FDNs optimized according to the three loss functions.*

	ψ	$\lvert\Delta\psi\rvert$	ϑ_{e}
$\mathcal{L}_{\text{FI}}$	0.86	0.15	8.4°
$\mathcal{L}_{\text{FD}}$	0.85	0.14	17.7°
$\mathcal{L}_{\text{SP}}$	0.64	0.07	13.1°

As baselines, we consider the two loss functions presented in Sec. 3.1, namely the frequency-independent loss function $\mathcal{L}_{\text{FI}}$ (2) and the frequency-dependent loss function $\mathcal{L}_{\text{FD}}$ (6), In particular, as for $\mathcal{L}_{\text{FD}}$, the mel-scale EDR in (10) is computed by filtering the 512-bin magnitude STFT of $h_j[n]$ with 64 triangular mel filters, while the STFT is computed using a 320-sample Hann window (20 ms) with hop size of 160 samples (10 ms). The EDP term in (2) and (6) is computed taking into account a Hann window of 20 ms. We vary κ_n, i.e., the parameter governing the sigmoid scaling of the Soft EDP, in a linear fashion by increasing it progressively from 10^2 to 10^5 over the range $n = 0, \ldots, L_h - 1$.

Finally, the multi-resolution EDR in (16) is computed taking into account $R = 3$ sets of resolutions, namely $[1024, 2048, 512]$, $[120, 240, 50]$, and $[600, 1200, 240]$ for the FFT length, hop size, and window length (in samples), respectively.

4.1. Parameter Initialization

We implement the differentiable SIMO FDN with $J = 8$ and $N = 24$ in Python using PyTorch. In particular, we train for a total of 650 epochs and we employ a single Adam optimizer with learning rate of 0.1. Then, we initialize the FDN in the same fashion for both proposed method and baselines. We optimize $\mathbf{W}$, $\mathbf{b}$, $\mathbf{C}$, and the delays $\boldsymbol{m}$, but we do not optimize $\mathbf{d}$, which is, instead, kept fixed and initialized directly according to the amplitude of the target direct-path samples.

Then, for all $\imath\jmath$, we initialize $\tilde{\mathbf{b}}_{\imath\jmath}^{(0)} \sim \mathcal{N}(0, 1/N)$ and $\tilde{\mathbf{C}}_{\imath\jmath}^{(0)} \sim \mathcal{N}(0, 1/N^2)$. We initialize $\tilde{\mathbf{W}}^{(0)}$ such that $\tilde{\mathbf{W}}_{\imath\jmath}^{(0)} \sim \mathcal{N}(0, 1/N)$ and $\tilde{\mathbf{\Gamma}}^{(0)}$ by having

$$\tilde{\gamma}_i^{(0)} = 10^{\frac{-3}{T_{60}^{\max} f_{\text{s}}}} \qquad (17)$$

as to better condition the energy decay in the time domain [17]. Then, we set $\tilde{\boldsymbol{m}}^{(0)}$ so that $\tilde{m}_i^{(0)} = \xi \tilde{m}_i^\star$ with $\tilde{m}_i^\star \sim \text{Beta}(\alpha, \beta)$, for $i = 1, ..., N$, with $\alpha \geq 1$ and $\beta > \alpha$. Moreover, we set $\xi = 1024$ as in [8, 13], together with $\alpha = 1.1$ and $\beta = 6$ such that a maximum possible delay of 64 ms and a mean value of about 10 ms are ensured.

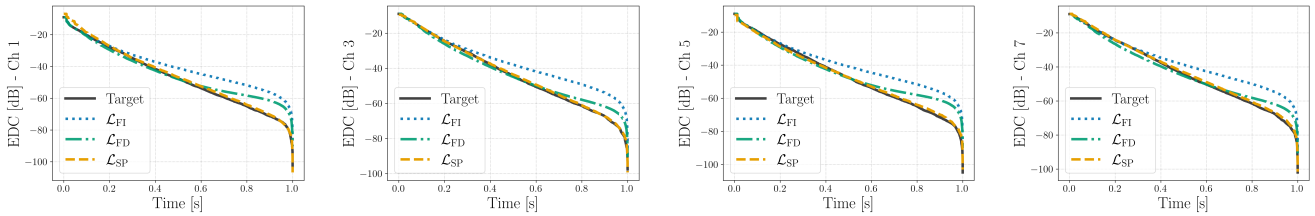
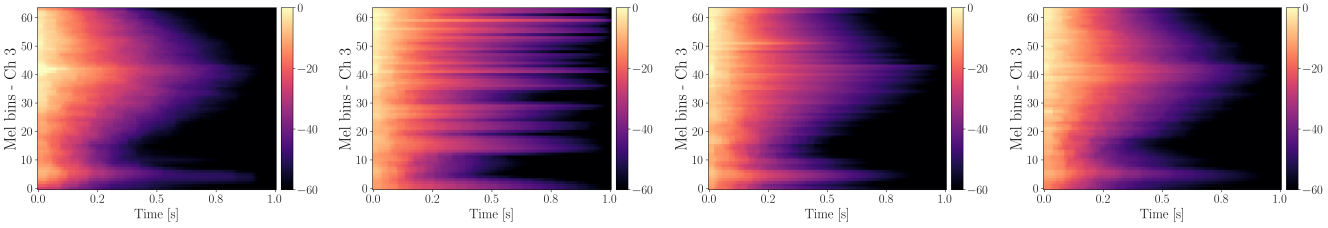
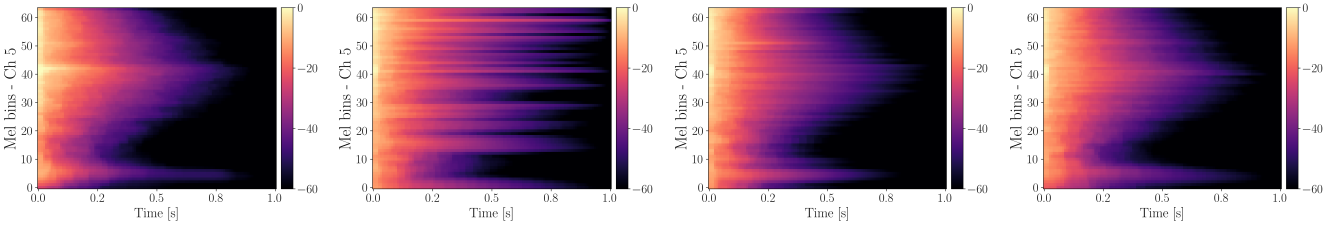


Figure 3: *Comparison among the EDCs with respect to the target.*



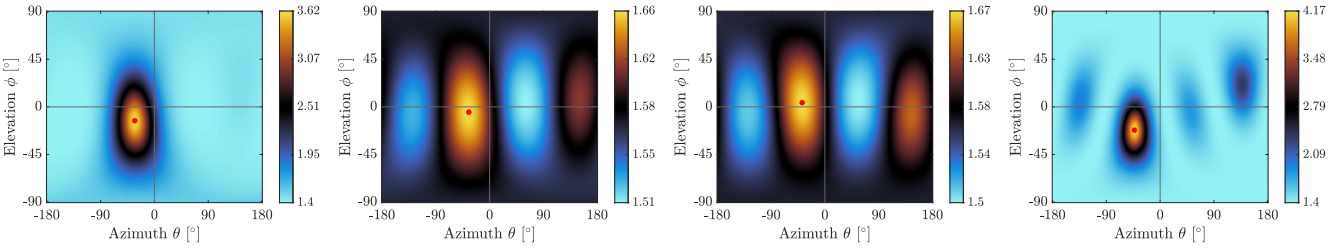
(a) Target RIR. (b) FDN trained with $\mathcal{L}_{\text{FI}}$. (c) FDN trained with $\mathcal{L}_{\text{FD}}$. (d) FDN trained with $\mathcal{L}_{\text{SP}}$.

Figure 4: *Mel-scale EDR (dB) of the SIMO FDN IRs. Channel 3.*



(a) Target RIR. (b) FDN trained with $\mathcal{L}_{\text{FI}}$. (c) FDN trained with $\mathcal{L}_{\text{FD}}$. (d) FDN trained with $\mathcal{L}_{\text{SP}}$.

Figure 5: *Mel-scale EDR (dB) of the SIMO FDN IRs. Channel 5.*



(a) Target RIR. (b) FDN trained with $\mathcal{L}_{\text{FI}}$. (c) FDN trained with $\mathcal{L}_{\text{FD}}$. (d) FDN trained with $\mathcal{L}_{\text{SP}}$.

Figure 6: *Pseudo-spectrum computed using MUSIC in the SH domain.*

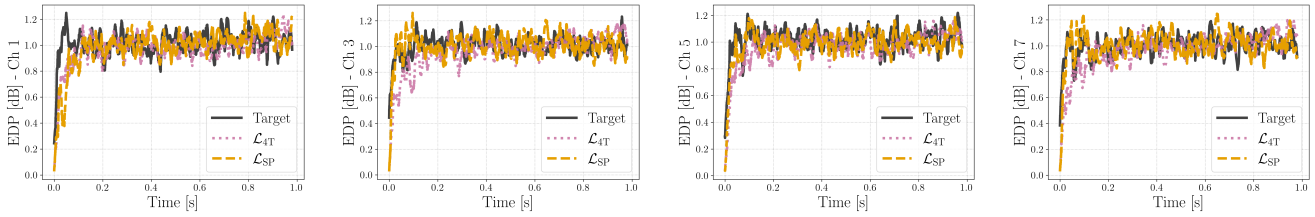


Figure 7: *Comparison with respect to the target between the EDP obtained optimizing the FDN with $\mathcal{L}_{SP}$ and with $\mathcal{L}_{4T}$.*

4.2. Results and Discussion

Fig. 3 shows the energy decay curves, in dB, of the FDN IRs trained using $\mathcal{L}_{\text{FI}}$, $\mathcal{L}_{\text{FD}}$, and $\mathcal{L}_{\text{SP}}$. In particular, the results of $\mathcal{L}_{\text{FI}}$ are shown with a dotted blue line, those of $\mathcal{L}_{\text{FD}}$ with a dash-dotted green line, while, instead, the dashed orange and the solid black lines represent the results of $\mathcal{L}_{\text{SP}}$ and the target EDC, respectively. Moreover, due to space constraints, we show only the results related to odd channels 1, 3, 5, and 7, i.e., capsules spanning different directions, although analogous results are obtained for the remaining ones. On the one hand, we can clearly see that the proposed method nicely follows the target decay, while, on the other hand, the blue and green lines part from it way before the $T_{60}^{\max} = 0.91$ s, with the blue line, i.e., $\mathcal{L}_{\text{FI}}$, being the worst as it is only able to match the target decay for just over 0.2 seconds. Fig. 3 reveals, thus, that the differentiable SIMO FDN trained using $\mathcal{L}_{\text{SP}}$ tend to model better the energy decay in the time domain with respect to the baselines, although it does not minimize directly an EDC loss term. Indeed, we argue that the joint optimization of the RIR energy envelope and the multi-resolution EDR better condition the FDN behavior. Such a trend is also visible looking at Figs. 4 and 5, which present the mel-scale EDRs in dB related to channels 3 and 5. In particular, Figs. 4b and 5b show that, having no frequency-dependent term, $\mathcal{L}_{\text{FI}}$ is not able to condition the training as to model the target coloration, leading, as a consequence, to the typical comb-like artifacts that characterize FDNs [7]. Figs. 4c and 5c reveal that $\mathcal{L}_{\text{FD}}$ is able to model the energy decay in the time-frequency domain, in line with the results obtained for SISO FDNs in [12]. However, it is possible to recognize a further improvement in the coloration-matching of the results showed in Figs. 4d and 5d, given that they present even lower comb-like artifacts, proving the MR-EDR term of $\mathcal{L}_{\text{SP}}$ suitable to obtain even more natural frequency decays. Finally, Fig. 6 shows the pseudo-spectra of the FDN IRs, obtained using the MUSIC method in the SH domain [31], as well as the DOA, which is represented by a red dot. The colorbar of each subfigure is tuned to allow us to appreciate the details of the energy distribution; a single colorbar shared among all the subplots would impair the visualization and thus the analysis of the results. Fig. 6b and Fig. 6c, indeed, depict pseudo-spectra taking values between 1.5 and 1.7. Conversely, the pseudo-spectrum associated to the proposed method in Fig. 6d showcases a dynamic range similar to that of the target (cf. Fig. 6a). This argument is supported also by the different energy distribution in space. Indeed, the baselines (Figs. 6b and 6c) present a more diffused soundfield and a wrong directivity, whereas, thanks to the ICLD loss term, the proposed method is able to optimize the SIMO FDN as to model the spatial distribution more accurately. The remaining discrepancy in spatial energy distribution could be attributed to unmatched differences in the spectra.

As to further evaluate the performance of the three learning objectives, we report in Table 1 further results concerning spatial metrics. In particular, we compute the *diffuseness* ψ, i.e., a metric characterizing the directional power variation, starting from the covariance of the SH signals [32] and employing the Spherical-Array-Processing library by A. Politis [33]. The target RIR features a diffuseness $\psi = 0.71$, showing that the soundfield characterizing the seminar room is, indeed, rather diffused. While all three losses achieve similar ψ, with values between 0.64 and 0.86, the proposed loss $\mathcal{L}_{\text{SP}}$ outperforms the others in absolute diffuseness error, obtaining $|\Delta\psi| = 0.07$ against 0.14 and 0.15 of $\mathcal{L}_{\text{FD}}$ and $\mathcal{L}_{\text{FI}}$, respectively. This suggests that $\mathcal{L}_{\text{SP}}$ provides better control over spatial reproduction accuracy, despite its slightly lower overall diffuseness. Moreover, we compute the DOA error (DOAE) considering [34]

$$\vartheta_{\text{e}} = \arccos(\hat{\mathbf{x}}^{\text{T}}\mathbf{x})\,, \tag{18}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the target and estimated cartesian DOA vectors, respectively. The proposed method turns out to improve the DOA estimation by $4.6°$ compared to $\mathcal{L}_{\text{FD}}$, whereas $\mathcal{L}_{\text{FI}}$, with $\vartheta_{\text{e}} = 8.4°$, is the methodology with the lowest DOAE. Though, it is also worth pointing out that MUSIC is sensitive to errors when applied to diffuse soundfields [33], such as that produced by $\mathcal{L}_{\text{FI}}$ ($\psi = 0.86$). In conclusion, we can state that the proposed learning objective is able to capture the spatial characteristics of the acquired soundfield, providing a fair source localization but, most importantly, a coherent spatial energy distribution.

4.3. Evaluating Envelope Loss Against EDC and EDP Losses

In Sec. 3.2, we argued in favor of replacing the $\mathcal{L}_{\text{EDC}}$ and $\mathcal{L}_{\text{EDP}}$ loss terms with $\mathcal{L}_{\text{ENV}}$ as to better condition the optimization of $\mathcal{L}_{\text{SP}}$. In this subsection, with the aim of evaluating the performance of the ENV loss against EDC and EDP losses, we train an FDN with the following composite loss function

$$\mathcal{L}_{\text{4T}} = \lambda_7\mathcal{L}_{\text{EDC}} + \lambda_8\mathcal{L}_{\text{MR-EDR}} + \lambda_9\mathcal{L}_{\text{EDP}} + \lambda_{10}\mathcal{L}_{\text{ICLD}}\,, \tag{19}$$

where λ_7, λ_8, λ_9, $\lambda_{10} \in \mathbb{R}_{>0}$, and we compare the results with those obtained through $\mathcal{L}_{\text{SP}}$ (16). It follows that (19) is obtained from (16) by substituting $\mathcal{L}_{\text{ENV}}$ with a linear combination of $\mathcal{L}_{\text{EDC}}$ and $\mathcal{L}_{\text{EDP}}$.

Fig. 7 reports the EDP of the two FDN IRs (solid orange curve for $\mathcal{L}_{\text{SP}}$ and dotted pink curve for $\mathcal{L}_{\text{4T}}$) as well as the target EDP (solid black curve). Hyperparameters are set to balance the loss terms during early training. Especially for channels 3, 5, and 7, the $\mathcal{L}_{\text{SP}}$ curve proves to better follow the target black line although the EDP term is not among those constituting the loss itself. The $\mathcal{L}_{\text{4T}}$ curve, instead, shows a slightly lower echo density in the first part of the IR. This can be attributed to the complex nature of the Soft EDP function, which could be subjected to gradient vanishing [8]. Finally, when comparing the EDC of the two realizations, we obtain similar results as evidenced by the T_{60} values averaged over channels, which read 0.892 s for $\mathcal{L}_{\text{SP}}$ and 0.897 s for $\mathcal{L}_{\text{4T}}$, respectively. We can thus state that the proposed $\mathcal{L}_{\text{ENV}}$ term is able to substitute the original two terms, improving, at the same time, the matching of echo density and training stability, while maintaining comparable accuracy on the energy decay in the time domain.

4.4. Perceptual Evaluation

With the purpose of evaluating the performance of different learning objectives as far as spatial quality (SQ) is concerned, we conducted a perceptual listening test online using the webMUSHRA framework [35]. SQ accounts for various spatial attributes, including depth, width, spatial distribution, reverberation, envelopment, and immersion. A total of 13 experienced participants (10 male, 3 female, average age 29.9 years) took part in the study, all of whom used different consumer-grade headphones to conduct the test. Listeners were asked to rate the SQ of different auralizations against a target. Before the assessment, they were given examples of low and high SQ as to define precisely the extent of the perceptual scale.

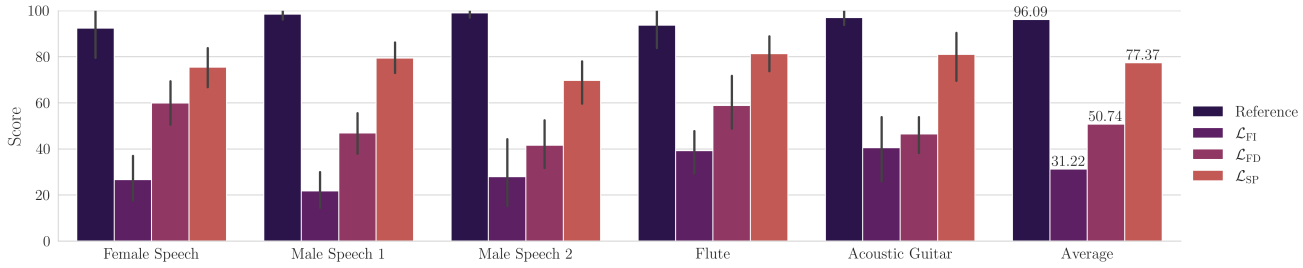


Figure 8: *Results of the perceptual test showing the spatial quality of the FDN IRs.*

The test stimuli[1] consisted of three speech signals and two music signals to ensure a comprehensive assessment across different audio content. For each stimulus, participants were presented four binaural auralizations: the renderings obtained from the IRs of the differentiable SIMO FDNs trained with $\mathcal{L}_{\text{FI}}$, $\mathcal{L}_{\text{FD}}$, and $\mathcal{L}_{\text{SP}}$, along with the target reference. Participants rated the SQ of each rendering against the target using a continuous scale ranging from 0 to 100. The test lasted less than 10 minutes, and none of the listeners reported experiencing hearing fatigue, ensuring reliable and comfortable conditions for the perceptual evaluation.

Fig. 8 shows the results of the test with the 95% confidence intervals and the relative average scores (rightmost bar charts). In all the charts, the Reference stimulus presents a variance different from zero meaning that the assessors were not always able to distinguish it from the other tracks. For instance, concerning the "Flute" case, participants deemed the Reference very close to the result of $\mathcal{L}_{\text{SP}}$. Overall, the proposed loss function obtained the best scores, 77.4 on average against 50.7 and 31.2 of $\mathcal{L}_{\text{FD}}$ and $\mathcal{L}_{\text{FI}}$, respectively, proving its superior ability to condition the differentiable SIMO FDN as to yield spatially coherent impulse responses.

5. CONCLUSIONS

In this article, we presented a novel learning objective specifically designed to optimize differentiable SIMO FDNs as to model the response of Higher-Order Microphone (HOM) arrays. We proposed to incorporate in the loss function a term related to the inter-channel level difference, computed starting from the FFTs of the first-order Ambisonic representation. Also, besides the multi-resolution EDR objective, we introduced a term related to the energy envelope, which we demonstrated to overcome state-of-the-art methods in matching the time-domain energy decay. We tested the proposed methodology to match a multichannel RIR from the HOMULA-RIR dataset, pointing out its superiority in modeling the soundfield both in space and time-frequency domains, a result corroborated also by a perceptual evaluation of the spatial quality.

Future work may concern the study and implementation of novel spatially-informed learning objectives, e.g., to improve DOA estimation, as well as the application of differentiable SIMO FDNs in soundfield reconstruction scenarios, paving the way toward the use of said filters for the efficient rendering of soundfields in spatial audio applications.

[1] Available at https://polimi-ispl.github.io/hom-dfdn/.

6. ACKNOWLEDGMENTS

The authors wish to thank Paolo Ostan and Dr. Raffaele Malvermi for the valuable and insightful discussions on optimization for spatial audio, which greatly contributed to the development of this work.

7. REFERENCES

[1] Vesa Välimäki, Julian D. Parker, Lauri Savioja, Julius O. Smith, and Jonathan S. Abel, "Fifty Years of Artificial Reverberation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[2] Franz Zotter and Matthias Frank, "Higher-Order Ambisonic Microphones and the Wave Equation (Linear, Lossless)," in *Ambisonics*, vol. 19, pp. 131–152. Cham, 2019.

[3] Alan Pawlak, Hyunkook Lee, Aki Mäkivirta, and Thomas Lund, "Spatial Analysis and Synthesis Methods: Subjective and Objective Evaluations Using Various Microphone Arrays in the Auralization of a Critical Listening Room," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 32, pp. 3986–4001, 2024.

[4] Juha Merimaa and Ville Pulkki, "Spatial Impulse Response Rendering I: Analysis and Synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, Dec. 2005.

[5] Leo McCormack, Ville Pulkki, Archontis Politis, Oliver Scheuregger, and Marton Marschall, "Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution," *J. Audio Eng. Soc.*, vol. 68, no. 5, pp. 338–354, June 2020.

[6] Sungho Lee, Hyeong-Seok Choi, and Kyogu Lee, "Differentiable Artificial Reverberation," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 30, pp. 2541–2556, 2022.

[7] Gloria Dal Santo, Karolina Prawda, Sebastian Schlecht, and Vesa Välimäki, "Differentiable Feedback Delay Network For Colorless Reverberation," in *Proc. 26th Int. Conf. Digit. Audio Effects (DAFx23)*, Copenhagen, Denmark, 2023, pp. 244–251.

[8] Alessandro Ilic Mezza, Riccardo Giampiccolo, Enzo De Sena, and Alberto Bernardini, "Data-Driven Room Acoustic Modeling Via Differentiable Feedback Delay Networks With Learnable Delay Lines," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1 (51), pp. 1–20, 2024.

[9] Alessandro Ilic Mezza, Riccardo Giampiccolo, Enzo De Sena, and Alberto Bernardini, "Differentiable Scattering Delay Networks for Artificial Reverberation," in *Proc. 28th Int. Conf. Digit. Audio Effects (DAFx25)*, Ancona, Italy, Sept. 2025.

[10] John Stautner and Miller Puckette, "Designing Multi-Channel Reverberators," *Comput. Music J.*, vol. 6, no. 1, pp. 52, 1982.

[11] Gloria Dal Santo, Karolina Prawda, Sebastian J. Schlecht, and Vesa Välimäki, "Optimizing Tiny Colorless Feedback Delay Networks," *EURASIP J. Audio Speech Music Process.*, vol. 2025, no. 1, pp. 13, Mar. 2025.

[12] Alessandro Ilic Mezza, Riccardo Giampiccolo, and Alberto Bernardini, "Modeling the Frequency-Dependent Sound Energy Decay of Acoustic Environments With Differentiable Feedback Delay Networks," in *Proc. 27th Int. Conf. Digit. Audio Effects (DAFx24)*, Guildford, UK, 2024, pp. 238–245.

[13] Riccardo Giampiccolo, Alessandro Ilic Mezza, and Alberto Bernardini, "Differentiable MIMO Feedback Delay Networks for Multichannel Room Impulse Response Modeling," in *Proc. 27th Int. Conf. Digit. Audio Effects (DAFx24)*, Guildford, UK, 2024, pp. 278–285.

[14] Vesa Välimäki, Karolina Prawda, and Sebastian J. Schlecht, "Two-Stage Attenuation Filter for Artificial Reverberation," *IEEE Signal Process. Lett.*, vol. 31, pp. 391–395, 2024.

[15] Huseyin Hacihabiboglu, Enzo De Sena, and Zoran Cvetkovic, "Frequency-Domain Scattering Delay Networks for Simulating Room Acoustics in Virtual Environments," in *Proc. Int. Conf. Signal Image Technol. Internet-Based Syst.*, Dijon, France, 2011, pp. 180–187.

[16] Alary Benoit, Archontis Politis, Sebastian Schlecht, and Vesa Välimäki, "Directional Feedback Delay Network," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 752–762, Oct. 2019.

[17] Jean-Marc Jot and Antoine Chaigne, "Digital Delay Networks for Designing Artificial Reverberators," in *Proc. 90th Audio Eng. Soc. Conv.*, Paris, France, 1991.

[18] Wootaek Lim and Juhan Nam, "Enhancing Spatial Audio Generation with Source Separation and Channel Panning Loss," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Seoul, Korea, Apr. 2024, pp. 8321–8325.

[19] Sebastian J. Schlecht, Jon Fagerström, and Vesa Välimäki, "Decorrelation in Feedback Delay Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 3478–3487, 2023.

[20] Federico Miotello, Mirco Pezzoli, Luca Comanducci, Alberto Bernardini, Fabio Antonacci, and Augusto Sarti, "HOMULA-RIR: A Room Impulse Response Dataset for Teleconferencing and Spatial Audio Applications Acquired Through Higher-Order Microphones and Uniform Linear Microphone Arrays," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech Signal Process. Workshops (ICASSPW)*, Seoul, Korea, 2024.

[21] Sebastian J. Schlecht, "FDNTB: The Feedback Delay Network Toolbox," in *Proc. 23rd Int. Conf. Digit. Audio Effects (DAFx20)*, Vienna, Austria, 2020, pp. 211–218.

[22] Hequn Bai, Gael Richard, and Laurent Daudet, "Late Reverberation Synthesis: From Radiance Transfer to Feedback Delay Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2260–2271, 2015.

[23] Atılım Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind, "Automatic Differentiation in Machine Learning: a Survey," *J. Mach. Learn. Res.*, vol. 153, pp. 1–43, 2018.

[24] Alessandro Ilic Mezza, Riccardo Giampiccolo, and Alberto Bernardini, "Data-Driven Parameter Estimation of Lumped-Element Models via Automatic Differentiation," *IEEE Access*, vol. 11, pp. 143601–143615, 2023.

[25] Jonathan S. Abel and Patty Huang, "A Simple, Robust Measure of Reverberation Echo Density," in *Proc. 121st Audio Eng. Soc. Conv.*, San Francisco, CA, USA, 2006.

[26] Michael A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10, Feb. 1973.

[27] Michael A. Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound," in *Proc.50th Audio Eng. Soc. Conv.*, London, UK, March 1975.

[28] P. B. Fellgett, "Ambisonic Reproduction of Directionality in Surround-Sound Systems," *Nature*, vol. 252, no. 5484, pp. 534–538, Dec. 1974.

[29] Jonathan Benjamin Alexis Thorpe, *Human Sound Localisation Cues and Their Relation to Morphology*, Ph.D. thesis, University of York, 2009.

[30] Robin Scheibler, Eric Bezzam, and Ivan Dokmanic, "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Apr. 2018, pp. 351–355.

[31] Xuan Li, Shefeng Yan, Xiaochuan Ma, and Chaohuan Hou, "Spherical Harmonics MUSIC versus Conventional MUSIC," *Appl. Acoust.*, vol. 72, no. 9, pp. 646–652, Sept. 2011.

[32] Bradford N. Gover, James G. Ryan, and Michael R. Stinson, "Microphone Array Measurement System for Analysis of Directional and Spatial Variations of Sound Fields," *J. Acoust. Soc. Am.*, vol. 112, no. 5 Pt 1, pp. 1980–1991, Nov. 2002.

[33] Archontis Politis, *Microphone Array Processing for Parametric Spatial Audio Techniques*, Ph.D. thesis, Aalto University, 2016.

[34] Maximo Cobos, Mirco Pezzoli, Fabio Antonacci, and Augusto Sarti, "Acoustic Source Localization in the Spherical Harmonics Domain Exploiting Low-Rank Approximations," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, June 2023, pp. 1–5.

[35] Michael Schoeffler, Sarah Bartoschek, Fabian Robert Stöter, Marlene Roess, Susanne Westphal, Bernd Edler, and Jürgen Herre, "webMUSHRA - A Comprehensive Framework for Web-Based Listening Tests," *J. Open Res. Softw.*, vol. 6, no. 1, pp. 8, Feb. 2018.