Audio Engineering Society

Convention Express Paper 35

Presented at the 153rd Convention
2022 October

A Method for matching room impulse responses with feedback delay networks

Ilias Ibnyahya[1] and Joshua D. Reiss[1]

[1] *Centre for Digital Music, Queen Mary University of London, London, U.K.*

Correspondence should be addressed to Ilias Ibnyahya (`i.ibnyahya@qmul.ac.uk`)

ABSTRACT

Recorded room impulse responses enable accurate and high-quality artificial reverberation. Used in combination with convolution, they can be computationally expensive and inflexible, providing little control to the user. On the other hand, reverberation algorithms are parametric which enable user control. However, they can lack realism and can be challenging to configure. To address these limitations, we introduce a multi-stage approach to optimize the coefficients of a Feedback Delay Network (FDN) reverberator to match a target room impulse response, thus enabling parametric control. In the first stage, we configure some FDN parameters by extracting features from the target impulse response. Then, we use a genetic algorithm to fit the remaining parameters to match the desired impulse response using a Mel-frequency cepstrum coefficients (MFCCs) cost function. We evaluate our approach across a dataset of impulse responses and conducted a subjective listening test. Our results indicate that the combination of the FDN with a short truncation of the target impulse response enables a better approximation, however, there are still differences with respect to the overall spectrum and the clarity factor in some more challenging cases.

1 Introduction

Recorded room impulse response offers an exhaustive representation of the sound characteristics of a reverberant space at a given point. It is often used with convolution to apply accurate reverberation to a digital audio signal. However, impulse response convolution with an audio signal doesn't provide any parametric control over the reverberation characteristics to the user. Also, the computing resources required to convolve an impulse response can be greater than the resources required for delay-based algorithmic reverberation [1].

Although several efficient architectures have been established to implement reverberation, good-sounding parametric reverberation is challenging [2, 3].The Feedback delay network (FDN) for reverberation was introduced by Stautner and Puckette in 1982 [4], and then developed further by Jot [5]. He proposed an analysis-synthesis method allowing the control of the FDN reverberation time as a function of the frequency [6]. However, other parameters such as matrix coefficients, delay lengths, and gains are not covered in this analytical approach.

To counter those limitations, recent approaches used machine learning to find the vector of FDN coefficients that provides the lowest error between a target room impulse response (RIR) and the FDN output. Coggin et al. [7] used a genetic algorithm in combination with a Yule-Walker optimization to determine the values of the absorption filters in the FDN. Other approaches in filter optimization showed better results compared to Yule-Walker method when applied to IIR filters using the least-squares method [8] and deep learning [9]. Shen et al. [10] described a data-driven approach to classify room parameters with FDN coefficients using Support-Vector-Machines. Their approach also used a genetic algorithm to generate the coefficients of the FDN. Previous approaches used signal envelope error [7, 11] and a weighted sum of reverberation time and clarity factor as their cost function [10]. This could be improved with a more perceptually relevant metric such as MFCCs.

Previous works [7, 10, 12] used a truncation of the target impulse response to generate the early reflections and then used the FDN for the late reverberation synthesis. This hybrid approach was initially introduced by Primavera et al. [13]. The length of the truncation is usually around 80 *ms* and doesn't take into account the target impulse response characteristics such as its reverberation time.

Genetic algorithm is often used in FDN optimization [7, 10, 11], as IIR filters are challenging to optimize with differentiable machine learning techniques. Lee et al. [14] developed a method to replace the IIR with FIR approximations, making artificial reverberators differentiable. This new approach opens the scope to more differentiable optimization on FDN. Previous works didn't assess the quality of their design against an impulse response dataset.

We present a model that truncates the early reflections based on the early decay time of the target impulse response. Our genetic algorithm computes the cost function with MFCCs. We use an automatic equalization method that provides lower error compared to the Yule-Walker approach to control energy decay. We compare the performances of the hybrid and FDN-only approach against a dataset of 82 room impulse responses by assessing the quality of our design with objective measurements and with a listening test.

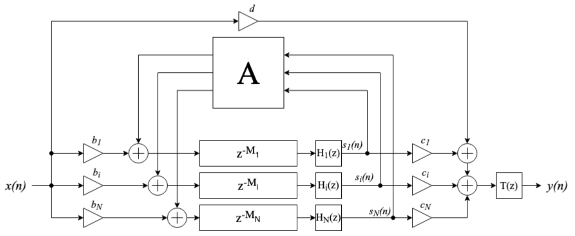

Fig. 1: A feedback delay network (FDN) artificial reverberator of order $N = 3$. Each delay line Z^{-M_i} of length M samples has an absorption equalizer $H_i(Z)$, feeding back into a matrix A, where i is the delay line number. The output is corrected with a parametric tone equalizer $T(z)$. b_i, c_i, d represent the input, output and direct signal gains.

2 Methods

2.1 Dataset and model

We built our target impulse response dataset using OpenAIRLib [15]. We extracted all the audio files from the website and kept all the mono measures, resulting in a dataset of 82 impulse responses. We normalized and removed the initial silence of all the files. Furthermore, we provide the code to generate the dataset along with the model used to match those impulse responses [1].

2.2 FDN Implementation

We used the MATLAB FDN toolbox [16] that provides an implementation of an FDN reverberation. This toolbox handles the synthesis of impulse response generation using delay state-space filter matrices, which reduce the processing time compared to sample-based implementation. The structure of the FDN used is shown in Figure 1.

The parameters detailed in Figure 1 are all the parameters we want to optimize to get an impulse response perceived as the target impulse response. Some parameters were calculated directly from the target analysis, such as the tone correction filter $T(z)$ and the attenuation filters $H_i(z)$.

[1] https://github.com/ilias-audio/MatchReverb

2.3 Absorption filter

The FDN absorption filters $H_i(z)$ were implemented as a 10-band biquad filter. To compute the absorption filters of the FDN, we used a modified version of the analysis-synthesis method developed by Jot [6] to match the $RT_{60}(f)$ of the target reverberation. We filtered the impulse response signal with an octave-band filter bank $G_{octave}(f_c,n)$ to obtain one filtered signal per biquad. We operated a squared time-reversed integration of the impulse response for each band resulting in an energy decay curve $EDC_{filter}(f_c,n)$, where f_c is the center frequency of the octave-band filter and n the sample at a given time, which is given by:

$$EDC_{Filter}(f_c,n) = \sum_{n}^{N} G_{octave}(f_c,n)^2 \qquad (1)$$

where n is the current sample and N is the total number of samples. Some of the impulse responses in our dataset had a high level of static noise. We had to revise our measurement of RT_{60} with the following measurement:

$$RT_{60} = 3 \cdot RT_{15} \qquad (2)$$

Where RT_{15} is measured between $n[-5dB]$ and $n[-20dB]$ points. We apply this measurement to each of the octave-band filters of the $EDC_{filter}(f_c,n)$ giving us an $RT_{60}(f)$. This provides a more robust representation of the attenuation at all frequencies, given in Equation 2.

$$H_i(f) = \frac{-60 \cdot M_i}{RT_{60}(f) \cdot f_s} \qquad (3)$$

Then, we transform the desired reverberation time to a transfer function using Equation 3. Where M is the length in samples of a given delay line i, f_s is the sample rate, and f is the current frequency. From this desired frequency response $H_i(f)$ we then use the accurate equalizer developed by Valimaki et al.[8] using the least-squares method to get the corresponding cascaded second-order biquad filters coefficients. We used the implementation given in the FDN Toolbox [16] for that step. This process allowed us to have an initialization of the attenuation filter parameters that was already the best result.

2.4 Tone correction filter

Since the $RT_{60}(f)$ of an impulse response is relative to its initial energy, we needed to constrain the initial energy, otherwise, the FDN would behave as a decaying white noise. To do so, we applied a static parametric equalizer to shape the initial energy of the impulse response. The tone correction filter $T(z)$ was implemented following a similar method discussed in Section 2.3 for the absorption filter. Here, we look at the initial $EDC_{Filter}(f_c,n)$ at $n = 0$ and generate a frequency response that we then turned into biquad coefficient values using the accurate equalizer algorithm [8].

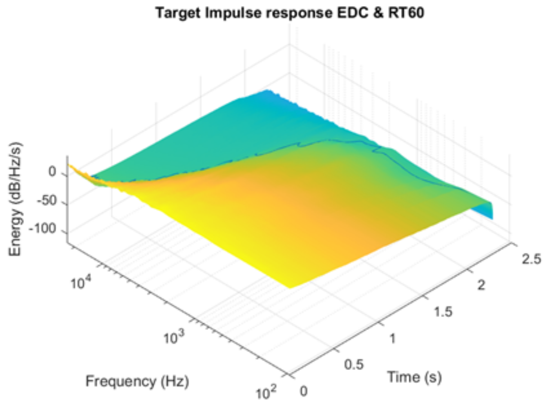

Fig. 2: Example of an octave-band filtered EDC with the blue line representing the $RT_{60}(f)$ calculated from it.

2.5 Optimization algorithm

2.5.1 Arbitrary parameters

In the previous sections, we covered the analysis-synthesis method that we used to determine the parameters for the absorption filters $H_i(z)$ and the global tone correction filter $T(z)$. To our knowledge, there is not yet a clear method to determine the gains, delay lengths, and feedback matrix values of the FDN to match an impulse response. We used a fixed FDN order of $N = 16$ and a random uni-lossless feedback matrix since it provides the best efficiency [1].

2.5.2 Genetic algorithm

The main parameters remaining to optimize were the input, output, direct gains and the delay lengths (respectively b_i, c_i, d and M_i). To find the best fitting values we used a genetic algorithm following the optimization process, detailed in Figure 3. We constrained the search

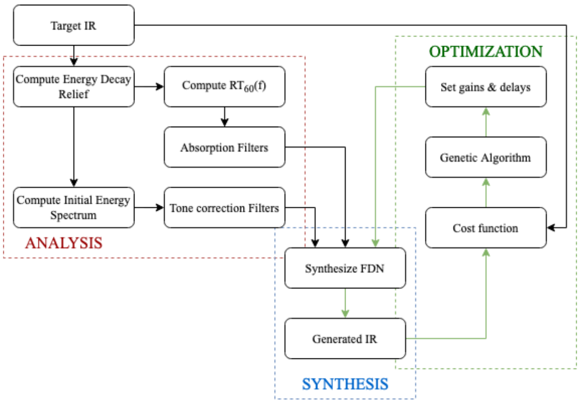

Fig. 3: Block diagram of our method to match a target impulse response. We first analyze and extract features from the target impulse response. Then, we synthesize an impulse response with the FDN and compare it with the target using a cost function. The cost is then used to define the fitness of the impulse response for the Genetic Algorithm.

for an optimal value of linear gain between [-1 : 1] and [0.0002 : 0.25] seconds for the delay lengths. We tried different combinations of meta-parameters for the number of generations and the number of impulse responses per generation. Furthermore, we noticed that the genetic algorithm quickly reaches a bias level in the optimization process, not showing any improvement only after a few generations. The impulse responses have been generated using 50 impulse responses per generation over 5 generations. This process took around 15 minutes for a 5 seconds impulse response with a sampling rate $f_s = 48\ kHz$ using a consumer laptop.

2.5.3 Cost function

To compare the generated impulse response with the target impulse response, we used Mel-frequency cepstral coefficients (MFCC). MFCCs offer a comprehensive representation of the different aspects of the impulse response. The first coefficient vector of the MFCC represents the global signal energy, which is similar to the energy decay curve. This value is used to check the global decaying behavior of the impulse response. The other coefficients of the ceptrum (typically 12 coefficients) are equally spaced on the mel scale, which approximates the human auditory system's response

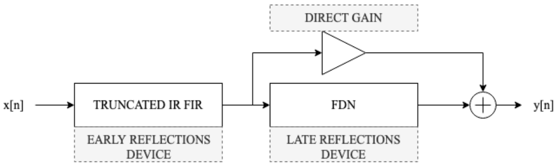

Fig. 4: Block diagram of the hybrid reverberator used in our design for a mono channel. We feed a truncation of the target IR into an FIR filter and then use the FDN to generate the late reflections of the reverberation.

more closely than the linearly-spaced frequency bands used in the normal spectrum. The algorithm aims to minimize the cost function

$$C(M_{tar}, M_{gen}) = \frac{1}{KN} \sum_{i=1}^{K} \sum_{j=1}^{N} \left| M_{tar}(i,j) - M_{gen}(i,j) \right| \tag{4}$$

where K is the number of MFCC, N is the total number of bins, M_{tar} are the target impulse response MFCCs, M_{gen} are the generated impulse response MFCCs, i and j are the array index representing frequency bins and time frames. The main goal of this optimization process is to set the gains to match the energy decay relief and avoid any audible ringing due to symmetric delay lengths.

2.5.4 Hybrid impulse response

During our development, we noticed that the early reflections and echo density build-up were poorly matched with the FDN. This was particularly noticeable when applied to percussive tracks. For this reason, we decided to implement an early reflections device that would help the echo density build-up. To demonstrate that it was a relevant approach, we decided to use a similar method to the one used by Primavera et al. [13]. We truncated the target impulse response at $t = EDT$ and created an FIR filter. On our dataset, the average was $EDT = 0.04\ s$, resulting on average in a 1920 tap FIR filter at $f_s = 48\ kHz$. As the initial power spectrum is already embedded in the FIR filter, we had to modify our optimization method as per Figure 4. We bypassed the tone correction filter $T(z)$.

2.5.5 Listening test

We conducted a multiple stimuli with hidden reference and anchor listening test, where participants have been

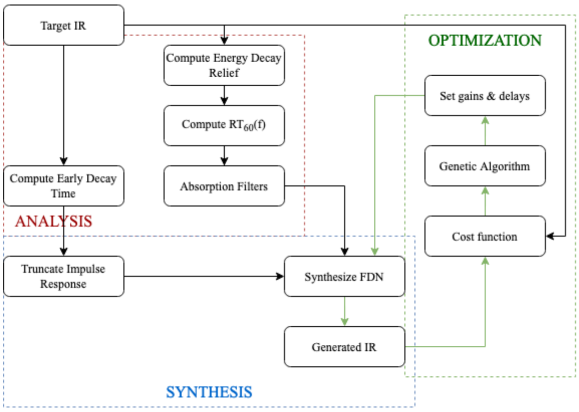

Fig. 5: Variation of the flow diagram in Figure 3 used to generate matched hybrid impulse responses. The truncated impulse response provides the initial energy of the spectrum, removing the necessity of a tone correction filters.

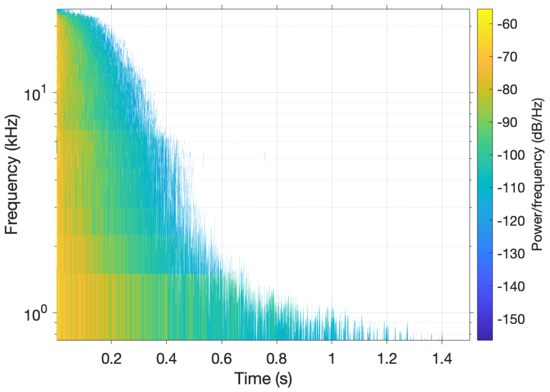

Fig. 6: Spectrogram of a target room impulse response from the OpenAIR dataset. $RT_{60} = 0.8\ s$

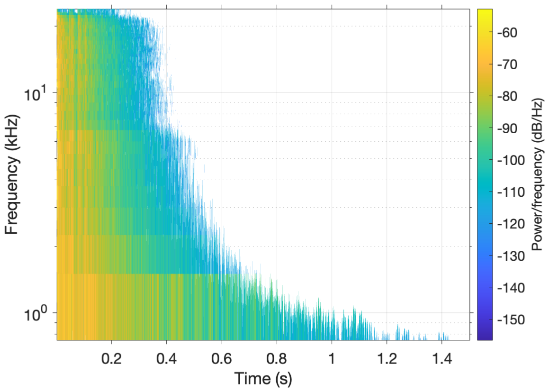

Fig. 7: Spectrogram of the generated impulse response matching Figure 6 using the hybrid approach.

asked to rate 4 reverberated tracks (hidden reference, Anchor, Hybrid reverberation, FDN only reverberation), in 3 different musical contexts (percussion, legato clean electric guitar, speech). We used two target reverberations, one with a short and one with a longer reverberation time. (respectively, $RT_{60} = 0.3\ s$ and $RT_{60} = 0.8\ s$). The dry and wet signals were mixed together at an equal level for the short reverberation. For the long reverberation, the wet signal was $-6\ dB$ lower than the dry signal. For the results, we normalized the answers given by participants based on their highest answers and removed participants that failed to recognize the hidden reference twice or more.

3 Results

3.1 Objective evaluation

The spectrogram of the generated impulse response is similar to the target impulse response, but show differences in the very high frequencies damping (Figures 6 & 7).

Our method produces a low level of error across several metrics such as early decay time, RT_{60}, Bass ratio (Figures 8 & 9) with a mean median value around $\varepsilon = 0.1\ s$. We observed a higher median value around $\varepsilon = 6\ dB$ per octave band in clarity and spectrum matching over the entire dataset (Figure 8). In most metrics, the hybrid reverberation reduces the error compared to the FDN-only impulse responses.

3.2 Listening Test

We had 18 valid participants in our listening test. In both listening tests (Tables 1 & 2), the anchor has the highest score. Participants struggled to distinguish the anchor and the reference, resulting in both being rated very high. The low-pass filtered impulse response was mixed with the dry signal resulting in not cutting the hole high-frequency content of the signal.

The hybrid reverberation has consistently a better rating than the FDN only (Tables 1 & 2). In some cases, such

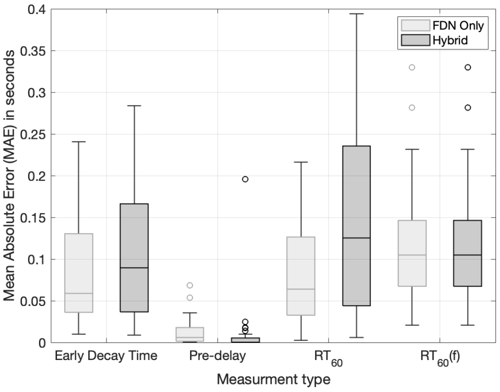

Fig. 8: Absolute error distribution in seconds on the 82 reverberations to match. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the 'o' marker symbol.

as speech and drums, the hybrid reverberation matches the median score and distribution of the anchor. Large room reverberation on drums has the lowest score overall due to its short decay, exposing more reverberation sound.

Audio	FDN Only	Hybrid	Anchor
Speech (%)	17 ± 5	42 ± 14	91 ± 12
Guitar (%)	33 ± 8	50 ± 14	99 ± 19
Drums (%)	20 ± 2	60 ± 18	60 ± 18

Table 1: Small room reverberation convolved with different audio tracks. The table shows the median and one standard deviation of the ratings given by participants.

4 Discussion

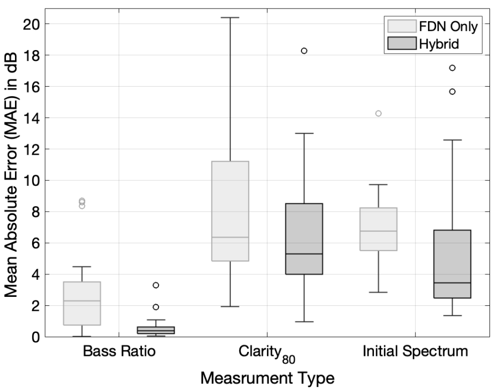

Fig. 9: Shows the absolute error distribution in dB when attempting to match 82 reverberations. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the 'o' marker symbol.

Audio	FDN Only	Hybrid	Anchor
Speech (%)	26 ± 7	45 ± 11	42 ± 12
Guitar (%)	34 ± 8	40 ± 9	90 ± 17
Drums (%)	6 ± 1	16 ± 4	28 ± 7

Table 2: Large room reverberation convolved with different audio tracks. The table shows the median and one standard deviation of the ratings given by participants.

This study shows that matching a target impulse response with an FDN and a direct convolution of the early reflections is better than with an FDN only. We observed a lower error in both objective and subjective tests. Listeners can distinguish between the hybrid model and the reference reverberation, especially on percussive audio tracks. Using MFCCs as a cost function for the genetic algorithm optimization consistently provides a low error across several reverberation metrics. Though our model doesn't match accurately the spectrum and clarity factor, with an average error of 6dB per octave band in both metrics (Figure 8). Many listeners didn't rate the reference at 100, but still gave it the highest score. This is because our test was unlabeled, which led to incorrect ratings. We addressed that by normalizing the listening test data based on the maximum value of each test for each participant. This study proves that MFCCs are suitable as a cost function for impulse response analysis. Our findings show that clarity and spectrum matching have a greater impact on the perception of a reverberation compared to the error in reverberation time.

5 Conclusion

We presented a reproducible method to match impulse responses with an FDN, allowing parametric control of synthesized room impulse response. We evaluated the performances of our model using an FDN and a hybrid model with a truncation of the target impulse response as an early reflection device. Furthermore, we demonstrated the importance of the first milliseconds of early reflections when matching a target impulse response. Additionally, we demonstrate the relevance of MFCCs as a cost function for impulse response matching. Our model shows a low level of error overall, but listeners can hear the difference between our model and the target impulse response. This can be explained by the errors we get on the initial spectrum and on the clarity factor.

References

[1] Schlecht, S. J. and Habets, E. A., "On Lossless Feedback Delay Networks," *IEEE Transactions on Signal Processing*, 65(6), pp. 1554–1564, 2017, ISSN 1053587X, doi:10.1109/TSP.2016.2637323.

[2] Välimäki, V., Parker, J. D., Savioja, L., Smith, J. O., and Abel, J. S., "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech and Language Processing*, 20(5), pp. 1421–1448, 2012, ISSN 15587916, doi:10.1109/TASL.2012.2189567.

[3] Välimäki, V., Parker, J., Savioja, L., Smith, J. O., and Abel, J., "More Than 50 Years of Artificial Reverberation," *Proc. AES 60th International conference*, pp. K–1, 2016.

[4] Stautner, J. and Puckette, M., "Designing Multi-Channel Reverberators," *Computer Music Journal*, 6(1), 1982, ISSN 01489267, doi:10.2307/3680358.

[5] Jot, J. M., "An analysis/synthesis approach to real-time artificial reverberation," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 2, 1992, ISSN 15206149, doi:10.1109/ICASSP.1992.226080.

[6] Jot, J.-M., Cerveau, L., and Warusfel, O., "Analysis and synthesis of room reverberation based on a statistical time-frequency model," *The 103rd AES Convention*, 4629, 1997.

[7] Coggin, J. and Pirkle, W., "Automatic design of feedback delay network reverb parameters for impulse response matching," in *141st Audio Engineering Society International Convention, AES*, 2016.

[8] Valimaki, V. and Liski, J., "Accurate cascade graphic equalizer," *IEEE Signal Processing Letters*, 24(2), 2017, ISSN 10709908, doi:10.1109/LSP.2016.2645280.

[9] Colonel, J. T., Steinmetz, C. J., Michelen, M., and Reiss, J. D., "Direct Design of Biquad Filter Cascades with Deep Learning by Sampling Random Polynomials," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3104–3108, 2022, ISSN 2379-190X, doi:10.1109/ICASSP43922.2022.9747660.

[10] Shen, J. and Duraiswami, R., "Data-driven feedback delay network construction for real-Time virtual room acoustics," in *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 2020, ISSN 21531633, doi:10.1145/3411109.3411145.

[11] Chemistruck, M., Marcolini, K., and Pirkle, W., "Generating matrix coefficients for feedback delay networks using genetic algorithm," in *133rd Audio Engineering Society Convention 2012, AES 2012*, volume 1, pp. 588–593, 2012, ISBN 9781622766031.

[12] Xia, R., Li, J., Primavera, A., Cecchi, S., Suzuki, Y., and Yan, Y., "A hybrid approach for reverberation simulation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E98A(10), 2015, ISSN 17451337, doi:10.1587/transfun.E98.A.2101.

[13] Primavera, A., Cecchi, S., Li, J., and Piazza, F., "Objective and subjective investigation on a novel method for digital reverberator parameters estimation," *IEEE/ACM Transactions on Audio Speech and Language Processing*, 22(2), pp. 441–452, 2014, ISSN 23299290, doi:10.1109/TASLP.2013.2295925.

[14] Lee, S., Choi, H.-S., and Lee, K., "Differentiable Artificial Reverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, pp. 2541–2556, 2022, ISSN 2329-9304, doi:10.1109/TASLP.2022.3193298.

[15] Shelley, S., Foteinou, A., and Murphy, D., "OpenAIR: An Online Auralization Resource with Applications for Game Audio Development: Proceedings of the 41st Int. Conference of the AES, Audio for Games, London, UK, February 2011," 2011.

[16] Schlecht, S. J. et al., "FDNTB: The feedback delay network toolbox," in *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx-20)*, pp. 211–218, 2020.