

VELVET-NOISE FEEDBACK DELAY NETWORK

Jon Fagerström¹, Benoit Alary¹, Sebastian J. Schlecht^{1,2} and Vesa Välimäki¹*

¹Acoustics Lab, Dept. of Signal Processing and Acoustics

²Media Lab, Dept. of Media

Aalto University

Espoo, Finland

jon.fagerstrom@aalto.fi

ABSTRACT

Artificial reverberation is an audio effect used to simulate the acoustics of a space while controlling its aesthetics, particularly on sounds recorded in a dry studio environment. Delay-based methods are a family of artificial reverberators using recirculating delay lines to create this effect. The feedback delay network is a popular delay-based reverberator providing a comprehensive framework for parametric reverberation by formalizing the recirculation of a set of interconnected delay lines. However, one known limitation of this algorithm is the initial slow build-up of echoes, which can sound unrealistic, and overcoming this problem often requires adding more delay lines to the network. In this paper, we study the effect of adding velvet-noise filters, which have random sparse coefficients, at the input and output branches of the reverberator. The goal is to increase the echo density while minimizing the spectral coloration. We compare different variations of velvet-noise filtering and show their benefits. We demonstrate that with velvet noise, the echo density of a conventional feedback delay network can be exceeded using half the number of delay lines and saving over 50% of computing operations in a practical configuration using low-order attenuation filters.

1. INTRODUCTION

Artificial reverberation algorithms have been developed for almost 60 years, starting from the first algorithm by Schroeder [1]. For many years, the feedback delay network (FDN) has been one of the most popular methods to create artificial reverberation [2, 3]. This paper proposes a novel FDN structure for increasing its echo density.

The idea of interconnecting multiple allpass filters through a matrix, which is the underlying principle of the FDN, was first introduced by Gerzon [4]. Stautner and Puckette further developed the idea of a recirculating network of delays [5], and Jot and Chaigne later extended it to the formal design that we now know as the FDN [2]. The analysis and improvement of the FDN remain active areas of research today [6, 7, 8, 9, 10], and very recently, the FDN was extended into the spatial domain [11, 12, 13].

The number and lengths of the delay lines are among the main questions when designing an FDN reverberator. The modal density of the synthetic response is positively correlated with the total

length of all the delay lines [2]. On the other hand, shorter delays help build up the echo density faster, but can lead to metallic timbre in the impulse response due to a lower modal density [14]. Increasing the number of delay lines improves the echo density [2] but also increases the number of arithmetic operations required per sample.

In practical applications, low-order attenuation filters are usually used to ensure a faster decay at high frequencies, which mimics the acoustics of a room. For precise control of the reverberation time of different frequency bands, high-order attenuation filters within the FDN are necessary [15, 16, 17, 18]. However, increasing the complexity of the attenuation filter greatly impacts the computational cost per delay line. Thus, the number of delay lines in the system must be minimized while retaining a sufficiently high echo and modal density.

Traditionally allpass filters have been used to increase the echo density of a reverberator [1, 19]. However, smearing problems have been reported for transient sounds [19]. Other ways to improve the FDN include introducing time-varying elements in the structure, such as modulated delay lines [20], allpass filters [3], or a time-varying feedback matrix [21, 22, 23]. Time-varying delay lines lead to imprecise control of the decay time, whereas an FDN with time-varying allpass filters is not guaranteed to be stable [23]. Since a time-varying feedback matrix is less likely to cause artifacts in the reverberation sound, this method has been found to improve the sound quality of the reverberation tail [23]. Another approach is to introduce short delays in the feedback matrix, so that each matrix element consists of a gain and a delay [24]. Also, separate early reflection modules using finite impulse response (FIR) filters for FDNs have been suggested [25]. However, the magnitude spectrum of these filters should be designed to minimize undesirable spectral coloration.

In this paper, we propose a novel reverberator structure with improved echo density, consisting of a conventional FDN structure with sparse velvet-noise filters, called the Velvet-noise Feedback Delay Network (VFDN). Velvet noise has been previously applied in audio processing to model the reverberation [26, 27, 28, 29] and to design computationally efficient decorrelation filters [30, 31]. The proposed method allows for a delay network using fewer but longer delay lines, while retaining a suitable echo density buildup. Using fewer delays reduces the computational cost and allows for more accurate attenuation filters, whereas the longer delays ensure a suitable modal density is retained.

The rest of this paper is organized as follows. Sec. 2 presents the previous ideas used in this study, including the basics of FDN, echo-density estimation, and velvet noise. Sec. 3 introduces the novel reverberation structure and discusses different ways of applying velvet-noise filtering. Sec. 4 analyzes the results of this

* This research has been funded by the Nordic Sound and Music Computing Network—NordicSMC (NordForsk project no. 86892) and by the Academy of Finland (ICHO project, grant no. 296390).

Copyright: © 2020 Jon Fagerström et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

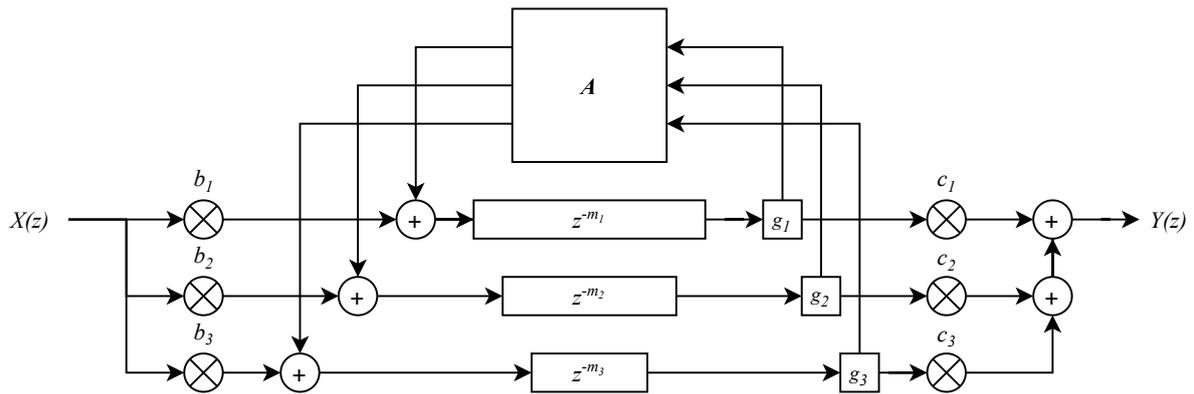


Figure 1: Block diagram of a conventional FDN of size $N = 3$.

work, and Sec. 5 conclude the paper. Spectral coloration caused by velvet-noise filters is studied analytically in the Appendix.

2. BACKGROUND

This section presents the basic FDN structure, the echo-density measure used in this study, and velvet-noise decorrelators.

2.1. Feedback Delay Network

The FDN consists of a set of recirculating delay lines interconnected through a feedback matrix \mathbf{A} (Fig. 1), that defines the recirculating gains for each connection [2]. By ensuring that this matrix is orthogonal, a lossless prototype is obtained, which redistributes the output energy of one delay line to the input of all delay lines. The lossless nature of this system permits the parametric control of the decay rate by using a target reverberation time T_{60} , corresponding to the time it takes to reach 60 dB of attenuation in the decay.

The output sample $y(n)$ of the recursive system, for an input $x(n)$, is formulated as

$$y(n) = \sum_{i=1}^N c_i g_i s_i(n), \quad (1)$$

$$s_i(n + m_i) = \sum_{j=1}^N A_{ij} g_j s_j(n) + b_i x(n), \quad (2)$$

where b_i and c_i are the input and output coefficients, respectively, A_{ij} is the feedback matrix element, g_i is the attenuation gain, and s_i are the output states of each delay line. The T_{60} specification is used to compute the appropriate g_i values used to attenuate the output of each delay line based on its length m_i .

The transfer function of the FDN is

$$H(z) = \frac{Y(z)}{X(z)} = \mathbf{c}^\top (\mathbf{D}_m(z)^{-1} - \mathbf{A})^{-1} \mathbf{b}, \quad (3)$$

where \mathbf{b} and \mathbf{c} are vectors containing the input and output gains, $\mathbf{D}_m(z) = \text{diag}(G_1(z)z^{-m_1}, G_2(z)z^{-m_2}, \dots, G_N(z)z^{-m_N})$, and \mathbf{A} is the feedback matrix. In practice, T_{60} is often specified at various frequencies, such as at octave bands, and then each gain g_i must be replaced with an attenuation filter $G_i(z)$, which can be a graphic equalizer [16, 17, 18].

2.2. Estimating the Echo Density

The echo density of an FDN impulse response is a measure of the number of echoes over time. For a lossless prototype FDN, a straightforward way to estimate it is the empirical echo density [32], which is computed by counting the amount of impulses per time frame at the output of the FDN.

Recently, Tukuljac *et al.* proposed another method for estimating the echo density of an impulse response [33]. This method, called the sorted density (SD) measure, was developed for complex acoustic scenes where the empirical echo density did not work reliably. The steps for computing the SD estimate are briefly described below.

First, the direct sound in the impulse response is removed. Then, the impulse response is converted to an echogram, i.e., $e(n) = h(n)^2$ and normalized to factor out the energy decay. This normalized echogram is then analyzed with a sliding window by computing the SD within each window. Finally, this density is normalized with the expected value of Gaussian noise so that an SD of 1 indicates the expected density of Gaussian noise. Thus, the SD measure yields an echo density measure that increases until it reaches a plateau close to 1.

2.3. Velvet-Noise Decorrelators

A velvet-noise sequence (VNS) is a pseudo-random signal, comparable to white noise, using as few non-zero values as possible [26, 34]. By taking advantage of the sparsity of the signal, computing its time-domain convolution with another signal becomes very efficient [28, 29]. Conceptually, the first step in generating velvet noise is to create a sequence of evenly spaced impulses at a selected density [26]. The sign and location of each impulse are then randomized, but impulses still remain within a given interval, having a range dictated by the desired impulse density. Figs 2(a) and 3(a) show an example of a VNS and its magnitude spectrum, respectively.

For a given density ρ and sampling rate f_s , the average spacing between two neighboring impulses in a VNS is

$$T_d = f_s / \rho, \quad (4)$$

which is called the grid size [34]. The total number of impulses in a VNS of length L_s (in samples) is

$$M = L_s T_d. \quad (5)$$

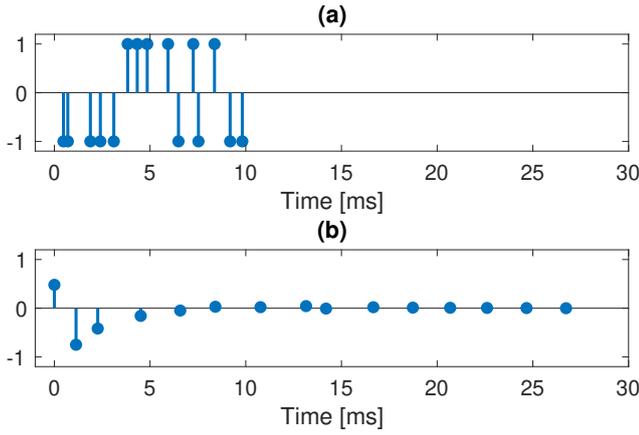


Figure 2: Examples of (a) a non-decaying (VN15) and (b) an optimally decaying velvet-noise (OVN15) sequence ($M = 15$).

The sign of each impulse is

$$s(m) = 2 \text{round}(r_1[m]) - 1, \quad (6)$$

where $m = 0, 1, 2, \dots, M - 1$ is the impulse index, the round function is the rounding operation, and $r_1(m)$ is a random number between 0 and 1. The location of the m th impulse of the VNS is calculated as

$$k(m) = \text{round}[mT_d + r_2[m](T_d - 1)], \quad (7)$$

where $r_2(m)$ is also a random number between 0 and 1 [34].

To convolve a VNS with another signal, we exploit the sparsity of the sequence, representing about 98% of the sequence for a density of $\rho = 1000$ at $f_s = 44100$ Hz, which allows a very efficient time-domain convolution computation [30]. By storing the VNS as a series of indices of the non-zero elements, all mathematical operations involving zeros can be skipped. The convolution with a basic VNS does not require multiplications either, only additions and subtractions [28, 29].

Furthermore, VNS sequences have been found to be suitable to decorrelate audio signals [30] by applying an exponentially decaying gain to each impulse to prevent the smearing of transients. For a given decay constant α , the gains are expressed as

$$s_e(m) = e^{-\alpha m} s(m) r_3(m), \quad (8)$$

where $r_3(m)$ is a random gain between 0.5 and 2.0 [30, 31]. The sparse convolution operation with a signal $x(n)$ can be written as

$$x * s_e = \sum_{m=0}^{M-1} x[n - k(m)] s_e(m), \quad (9)$$

where the asterisk (*) denotes the discrete convolution.

Since VNS filters do not have an exactly flat magnitude response, as seen in Fig. 3(a), they introduce a minor coloration to audio signals. For this reason, optimizing the random values is recommended to minimize the spectral deviation [31]. Instead of simply choosing random values, the impulse sign $r_1(m)$, the impulse location $r_2(m)$, and the impulse gain $r_3(m)$ are specified by a nonlinear optimization scheme. For the resulting optimized velvet-noise decorrelators, a peak magnitude-response deviation

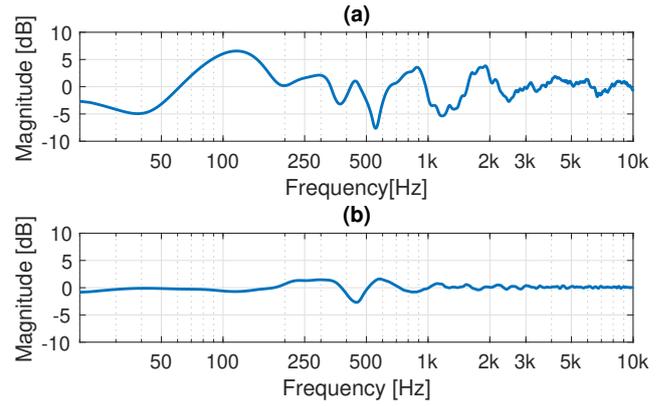


Figure 3: Smoothed magnitude spectra of the signals in Fig. 2: (a) a non-decaying and (b) an optimally decaying VNS.

of less than 1 dB can be achieved when third-octave smoothing is applied [31]. Fig. 2(b) shows an example optimized VNS, that decays approximately exponentially. Fig. 3(b) presents its magnitude spectrum, that has only a few dB of ripple, much less than in the spectrum of the non-optimized VNS in Fig. 3(a).

In this work, VNS filters are used to increase the echo density in artificial reverberation, a goal different from decorrelation. Some of the same requirements as above still apply, such as the desire for computational efficiency, minimizing the smearing of transients, and the need for a flat magnitude response.

3. PROPOSED STRUCTURE

The novel VFDDN structure is an extension of the conventional FDN in Fig. 1. As shown in Fig. 4, the input and output gains b_i and c_i , respectively, are replaced by sparse VNS filters $b_i(z)$ and $c_i(z)$. The VNS filters used here are relatively short, and their main purpose is to increase the echo density in the final output, which is otherwise sparse in the beginning of the impulse response due to the exponential nature of the recirculation. The transfer function of the proposed VFDDN structure is

$$H(z) = \mathbf{c}(z)^\top (\mathbf{D}_m(z)^{-1} - \mathbf{A})^{-1} \mathbf{b}(z), \quad (10)$$

where $\mathbf{c}(z)$ and $\mathbf{b}(z)$ are vectors containing the VNSs for the output and input delay lines, respectively. The form of the transfer function stays similar to the conventional FDN, as seen by comparing (3) and (10). Since the VNS filters are placed outside the feedback loops in Fig. 4, they cannot affect the decay rate of the system.

3.1. Optional Configurations

The VFDDN offers three distinct configurations. One configuration is to use a single set of VNS filters connected at either the input, using the $b_i(z)$, or the output, using the $c_i(z)$ of each delay line. Alternatively, the VNS filters can be connected to both the input and output, using both $b_i(z)$ and $c_i(z)$.

When considering the absolute echo density of the VFDDN output signal, each convolution operation with the VNS multiplies the number of echoes. If a single set of VNS filters is used at either

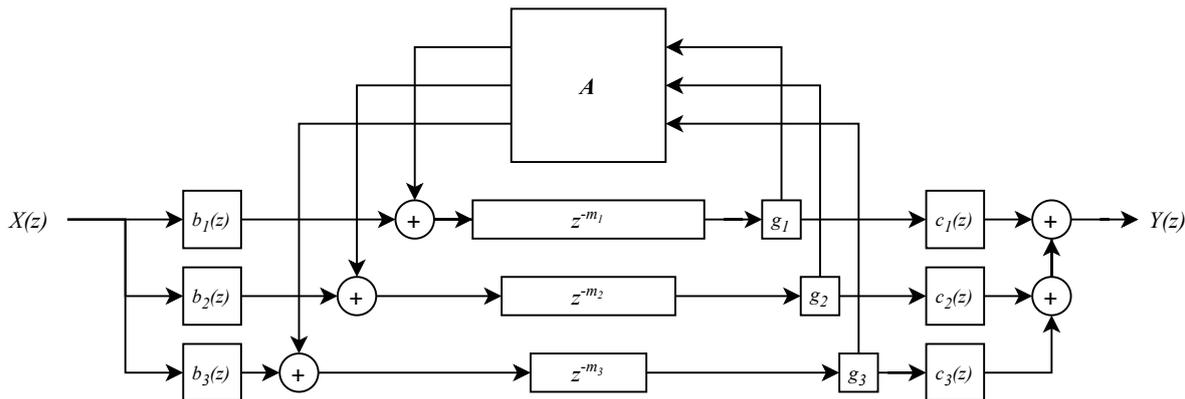


Figure 4: Block diagram of the proposed VFDN having a VNS filter in each input and output branch.

the input or output, and each VNS filter has $2M$ impulses, the multiplier is approximately

$$E_{\text{single}} = 2M. \quad (11)$$

Here, the superposition of impulses that inevitably occurs is not accounted for, but we assume for simplicity that each impulse appears separately in the impulse response.

If each VNS contains M impulses when using sets of VNSs at both the inputs and outputs, the multiplier of the absolute echo density is approximately

$$E_{\text{both}} = M^2. \quad (12)$$

Both configurations retain the same computational cost since they both use a total of $2M$ impulses. Since $E_{\text{both}} > E_{\text{single}}$ when $M > 1$, dividing the total number of impulses equally to both the input and output increases the echo density more when compared to using the same total number of impulses at either just the inputs or outputs.

Fig. 5 further demonstrates the benefit of using VNSs at both the input and output of each delay line. There, VN_1 and VN_2 are two VNSs with 15 impulses each, where the former can be thought to be connected at the input and the latter connected at the output VNS filters. The bottom pane in Fig. 5 shows the resulting convolved sequence, which is much denser than the two individual VNSs with almost $M^2 = 15^2 = 225$ pulses. This convolution result is equivalent to the filtering observed when a signal goes through both filters in the structure. Superimposed impulses result in samples with a value of 2 in Fig. 5 (bottom).

We could also consider having two sets of VNS filters connected in series either at the input or output. However, the resulting spectral coloration is flatter when having the filters at both the input and output. This difference in the coloration between these two configurations is derived in Appendix 7.2.

Fig. 6 shows the echo density plot for the three optional configurations of the VFDN. Each of the configurations has the same total number of impulses in the velvet sequences ($M = 30$), which also means they have the same computational cost. All other parameters, such as the feedback matrix and the delay-line lengths, remain unchanged in this comparison. The normalized echo densities of the single input or output VFDN configurations are equivalent, as expected. However, when using a configuration of the

VFDN that combines both input and output VNSs, the echo density grows faster. We observe that the measured echo density agrees with (11) and (12), demonstrating the benefits of using this configuration.

3.2. Velvet-Noise Sequence Types

There are two types of VNSs to be used in a VFDN, the basic non-decaying sequences containing $+1$'s and -1 's only and optimized decaying VNSs. The advantage of the non-decaying VNS is that it does not require multiplications in the convolution, and that all its impulses are equally tall. Therefore, no computation is necessary for perceptually negligibly small impulses that do not contribute to the overall echo density. The advantage of the decaying VNS is that it can be optimized to have a practically flat magnitude response. Unfortunately, towards its tail, the sample values get smaller, so some of them may not contribute to the increased echo density, as they may become inaudible at the output signal. It should be noted that, on modern hardware, some of these operations can be performed in parallel, which will improve the performance.

The optimized sequences used in this paper are the optimized VN sequences from [31]. They are 30-ms long with 15 or 30 impulses each and are denoted OVN_{15} and OVN_{30} , respectively. The non-decaying sequences are 10-ms long with 15 impulses each and are denoted VN_{15} . Initially, 30-ms-long non-decaying sequences were also considered, but they were found to cause transient smearing. Also, the improvement in the echo-density growth is superior with the 10-ms sequences.

4. RESULTS

The main motivation behind this research was to improve the echo density build-up of a conventional FDN. To quantify the resulting improvements, we compared the normalized echo density of two conventional FDNs, the proposed VFDN structures, and another recent extended FDN structure. The spectral coloration and implementation costs are also discussed. Audio examples are available at <http://research.spa.aalto.fi/publications/papers/dafx20-vfdn/> using the web-audio player from [35].

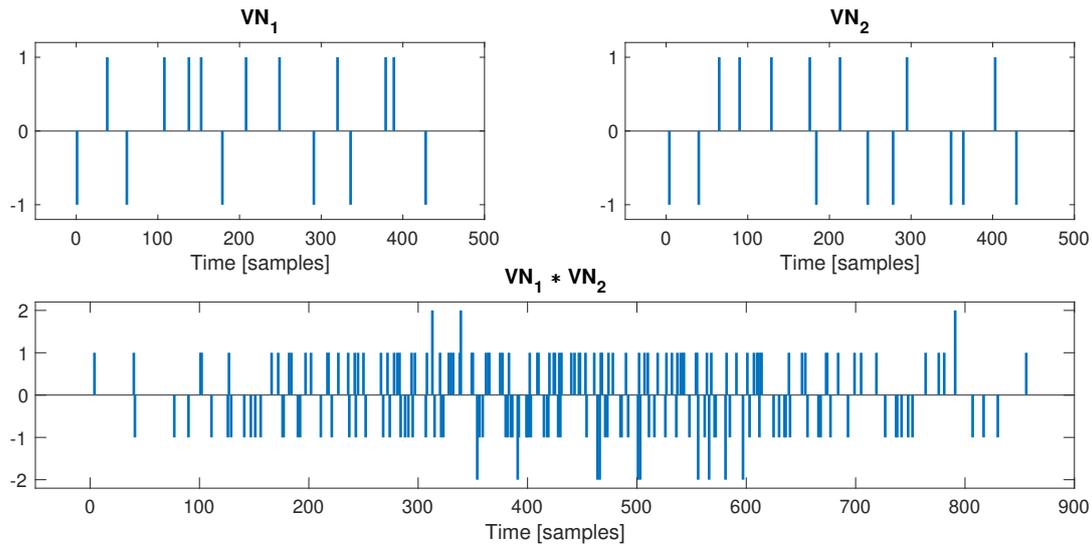


Figure 5: Two velvet-noise sequences (VN_1 , VN_2) and their convolution.

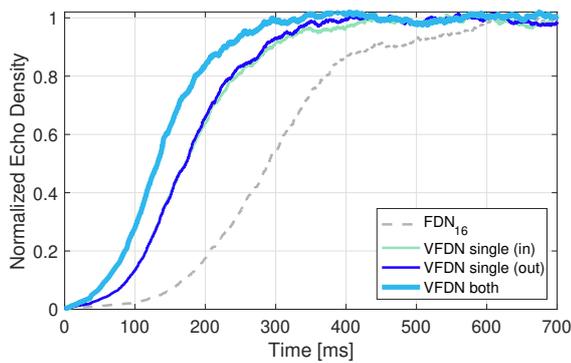


Figure 6: Growth of the normalized echo density over time with a conventional FDN and three optional VFDN configurations.

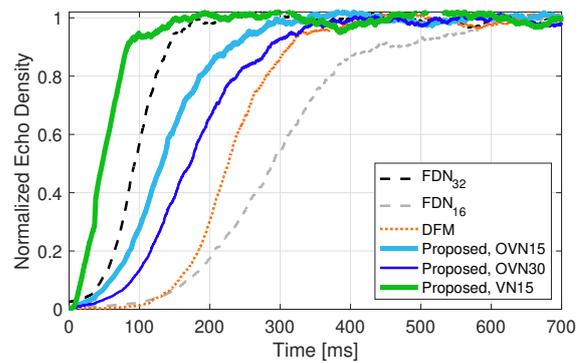


Figure 7: Normalized echo density of two conventional FDNs, a DFM-FDN [24], and the three proposed VFDN structures. The OVN30 configuration has VNSs only at the output, and the OVN15 and VN15 configurations have VNSs both at the input and the output.

4.1. Improvements in Echo Density

FDN structures with 32 and 16 delay lines, denoted as FDN_{32} and FDN_{16} , respectively, were used as the target and baseline methods in this study. Fig. 7 shows the normalized echo density of the FDN_{32} and FDN_{16} . FDN_{32} is considered to produce a sufficiently dense impulse response, whereas FDN_{16} has an impulse response that is slightly too sparse and would benefit from improvement. First, the target FDN_{32} was created with prime-number delay-line lengths (839, 881, 929, 971, 1013, 1049, 1091, 1123, 1181, 1223, 1277, 1301, 1361, 1423, 1451, 1487, 1531, 1571, 1609, 1657, 1699, 1747, 1789, 1861, 1889, 1949, 1997, 2029, 2083, 2129, 2161, and 2237).

The delay-line lengths of FDN_{16} were then computed by summing each of the two consecutive delay-line lengths of FDN_{32} and rounding them to the closest prime number length (1721, 1901, 2063, 2213, 2399, 2579, 2789, 2939, 3109, 3271, 3449, 3643, 3833, 4027, 4211, and 4397). This allowed keeping all the delay-line lengths prime, while having the total delay length in both configurations close to equal. Retaining the same total delay length

between different configurations ensures the modal density is not lowered, which can result in poor sound quality. The feedback matrices are random orthogonal matrices.

All the proposed structures in Fig. 7 have the same delay lines and feedback matrix as the FDN_{16} . The delay feedback matrix (DFM) corresponds to a recently proposed FDN structure having delay lines in its feedback matrix [24]. Fig. 7 shows the growth of the echo density of the $VFDN_{16}$ with the short VN15 sequences is even faster than that of the target FDN_{32} . When using the optimized OVN15 instead, the resulting echo-density growth still surpasses the DFM structure and gets close to the target FDN_{32} . As expected, the OVN30 configuration with VNS filters only on the output, introduces fewer echoes.

As mentioned earlier, since the decaying attenuation in an OVN sequence can preserve the transients better, the length used for the VN and OVN filters are 10 and 30 ms, respectively. Additionally,

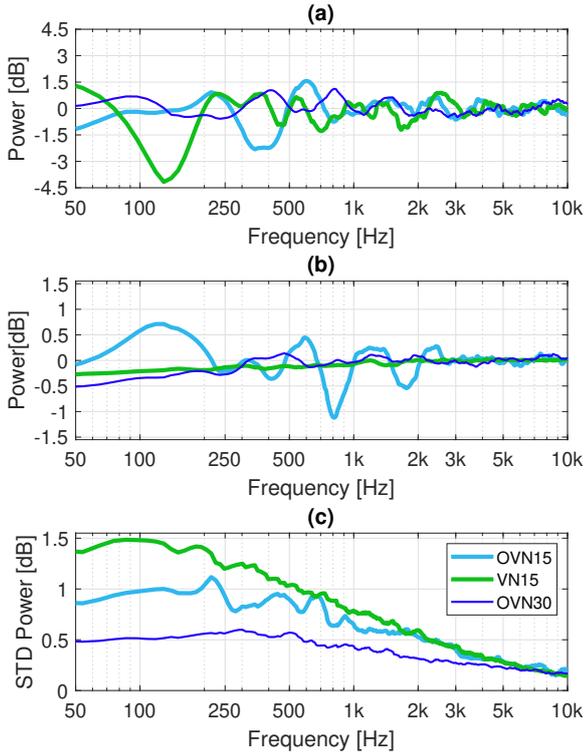


Figure 8: (a) Example power spectra, (b) mean, and (c) standard deviation of the smoothed power spectra for 500 instances.

we have experimented with a Schroeder allpass (SAP) filter structure with seven filters in series at each delay output [28]. The delay lengths of the SAP filter are 630, 555, 442, 209, 140, 64, and 1 samples, and the coefficient is 0.7 for all allpass filters. The benefit in the echo density was comparable to a VFDN structure with the same computational cost. However, for transient sounds the series SAP smears the signal, as is evident from the provided online sound examples. These results demonstrate that the proposed idea of inserting VNS filters at the input and output branches of the baseline method FDN_{16} substantially improves its echo density.

4.2. Spectral Coloration

Although the VNS filters improve the echo density, they also introduce some coloration [30, 31], which appears in the response of the reverberator. The amount of coloration depends on the particular sequences used, but some types of VNS filters introduce more coloration than others [30, 31].

The spectral coloration introduced by the VNS filters placed at the inputs and outputs of the FDN can be analyzed using the following method. The transfer function in (10) can be rewritten as

$$H(z) = \mathbf{1}^\top \text{diag}(\mathbf{c}(z))\mathbf{P}(z)\text{diag}(\mathbf{b}(z))\mathbf{1}, \quad (13)$$

where $\mathbf{P}(z) = (\mathbf{D}_m(z)^{-1} - \mathbf{A})^{-1}$ denotes the loop transfer

Table 1: Computational costs of different FDN and VFDN configurations.

Configuration	ADD	MUL	Total	Saving
FDN_{32}	1280	1440	2720	Reference
FDN_{16}	384	464	848	69%
DFM [24]	384	4464	848	69%
OVN15	864	912	1776	35%
OVN30 at outputs	864	912	1776	35%
VN15	864	432	1296	52%

function, or more succinctly

$$H(z) = \mathbf{1}^\top (\mathbf{P}(z) \circ \Gamma(z))\mathbf{1}, \quad (14)$$

where $\Gamma(z) = \mathbf{c}(z)\mathbf{b}(z)^\top$ is a frequency-dependent gain matrix and \circ denotes the Hadamard product, i.e., the element-wise multiplication of two matrices. The spectral coloration $E(\omega)$ is the frequency-dependent energy ratio of the FDN with and without the gain matrix $\Gamma(z)$. If the matrix entries in $\mathbf{P}(z)$ are uncorrelated and of the same energy, then only the input and output filters determine the spectral coloration, i.e.,

$$E(\omega) = \frac{\|\Gamma(e^{j\omega}) \circ \mathbf{P}(e^{j\omega})\|_F^2}{\|\mathbf{P}(e^{j\omega})\|_F^2} = \frac{\|\mathbf{c}(e^{j\omega})\|_2 \|\mathbf{b}(e^{j\omega})\|_2}{N^2}. \quad (15)$$

More details on this derivation are given in Appendix 7.1. Although the loop transfer function $\mathbf{P}(z)$ is not entirely uncorrelated, for the FDN configuration tested, the equation above yields an accurate estimate of the spectral coloration with the mean broadband error being 0.02 dB.

To compare the impact of the different variants of the proposed method on the spectrum, we computed the spectral coloration $E(\omega)$ of 500 configurations for each of the proposed VNS type. One configuration includes 32 sequences for VN15 and OVN15, having 16-input and 16-output filters, and 16 sequences for OVN30, which consist only of the output filters. Each of the VN15 sequences are random, whereas the optimized OVN15 and OVN30 sequences were picked randomly out of a set of 500 pregenerated sequences. Fig. 8(a) shows example spectral coloration plots of a single instance of each proposed VNS configuration. Fig. 8(b) and (c) show the mean, and standard deviation (STD) of the spectral coloration of the 500 instances, respectively.

Of the tested options, OVN30 sequences placed only at the delay-line outputs introduce the least amount of coloration to the system, followed by the OVN15 sequences placed at both input and output branches. The random VN15 sequences are the most coloring, but still their STD value remains less than 1.5 dB, as seen in Fig. 8(c). However, a formal perceptual study is necessary to determine whether this deviation is problematic in a practical setting, in which frequency-dependent attenuation is applied. Furthermore, since the spectral deviation is constant, it is possible to introduce a set of equalization filters at the final outputs to compensate for the coloration.

4.3. Computational Cost

Table 1 shows the number of operations per output sample for the configurations presented in Fig. 7. These numbers are computed for a practical setup using fourth-order attenuation filters consisting of a second-order low-shelf, a second-order high-shelf, and

gain for middle frequencies. The FDN₃₂ structure is used as the reference method to calculate the amount of savings, since it is the target we used for the echo density measure. A comparison between the VF_{FDN}₁₆ structures implemented using OV_N15 and VN15 highlights the added cost of using decaying VNSs, which shows in the number of multiplications in Table 1. The proposed configuration using OV_N15 and VN15 save 35% and 52% of operations, respectively, so they both still use considerably less operations than the reference method FDN₃₂.

Although counting additions and multiplications separately may not be relevant on modern hardware, Table 1 follows the typical method of presenting the computational cost and from which also the Multiply and Accumulate (MAC) operations can be derived. Here, we used a dense random orthogonal matrix in the FDN recirculation path. However, computational cost can be saved by using special matrices, such as the Hadamard matrix [36], which would reduce the computational benefit of using VN.

5. CONCLUSION

This paper proposes inserting velvet-noise filters at the input and output branches of an FDN to increase its echo density during the beginning of the impulse response. The sparseness of the impulse response is a known limitation of the FDN. This work shows that with the proposed VF_{FDN} an even faster growth in the echo density can be obtained than with the doubling of the number of delay lines in a conventional FDN. The short velvet-noise filters lead to computational savings, as they can be convolved very efficiently with a digital signal.

Various configurations for the VF_{FDN} are proposed in this work. A configuration with VNS filters both at the input and at the output of the FDN was shown to be a particularly effective solution, since the echo density increases more than when having VNS sequences with the same number of impulses only at the input or output of the FDN.

Non-decaying and decaying VNS filters were compared in the VF_{FDN}. The non-decaying sequences were found to help the echo density of the VF_{FDN} grow more rapidly than the decaying ones. However, the non-decaying VNSs cause more coloration than non-decaying sequences optimized to have a flat spectrum. Equalization of the resulting response of the VF_{FDN} to compensate for the coloration is left for future work.

6. REFERENCES

- [1] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223, July 1962.
- [2] J. M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Proc. Audio Eng. Soc. 90th Conv.*, Paris, France, 19–22 Feb. 1991.
- [3] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 20, no. 5, pp. 1421–1448, July 2012.
- [4] M. A. Gerzon, "Synthetic stereo reverberation: Part 1," *Studio Sound*, vol. 13, pp. 632–635, Dec. 1971.
- [5] J. Stautner and M. Puckette, "Designing multi-channel reverberators," *Computer Music J.*, vol. 6, no. 1, pp. 52–65, 1982.
- [6] S. J. Schlecht and E. A. P. Habets, "On lossless feedback delay networks," *IEEE Trans. Signal Processing*, vol. 65, no. 6, pp. 1554–1564, Mar. 2017.
- [7] N. Agus, H. Anderson, J.-M. Chen, S. Lui, and D. Herremans, "Minimally simple binaural room modeling using a single feedback delay network," *J. Audio Eng. Soc.*, vol. 66, no. 10, pp. 791–807, Oct. 2018.
- [8] S. J. Schlecht and E. A. P. Habets, "Modal decomposition of feedback delay networks," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5340–5351, Oct. 2019.
- [9] O. Das, E. K. Canfield-Dafilou, and J. S. Abel, "On the behavior of delay network reverberator modes," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2019, pp. 50–54.
- [10] S. J. Schlecht and E. A. P. Habets, "Scattering in feedback delay networks," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 28, pp. 1915–1924, June 2020.
- [11] B. Wiggins and M. Dring, "AmbiFreeVerb 2 – Development of a 3D ambisonic reverb with spatial warping and variable scattering," in *Proc. Audio Eng. Soc. Int. Conf. Sound Field Control*, Guildford, UK, Jul. 2016.
- [12] B. Alary, A. Politis, S. J. Schlecht, and V. Välimäki, "Directional feedback delay network," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 752–762, Oct. 2019.
- [13] B. Alary and A. Politis, "Frequency-dependent directional feedback delay network," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020.
- [14] M. Karjalainen and H. Järveläinen, "More about this reverberation science: Perceptually good late reverberation," in *Proc. Audio Eng. Soc. 111th Conv.*, New York, USA, Sept. 2001, paper no. 5415.
- [15] J.-M. Jot, "Proportional parametric equalizers—Application to digital reverberation and environmental audio processing," in *Proc. Audio Eng. Soc. 139th Conv.*, New York, USA, Oct. 29–Nov. 1, 2015.
- [16] S. J. Schlecht and E. A. P. Habets, "Accurate reverberation time control in feedback delay networks," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Edinburgh, UK, Sept. 2017, pp. 337–344.
- [17] K. Prawda, V. Välimäki, and S. J. Schlecht, "Improved reverberation time control for feedback delay networks," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Birmingham, UK, Sept. 2019.
- [18] K. Prawda, V. Välimäki, and S. Serafin, "Evaluation of accurate artificial reverberation algorithm," in *Proc. Sound and Music Computing Conf. (SMC)*, Turin, Italy, June 2020.
- [19] J. A. Moorer, "About this reverberation business," *Computer Music J.*, vol. 3, no. 2, pp. 13–28, June 1979.
- [20] D. Griesinger, "Improving room acoustics through time-variant synthetic reverberation," in *Proc. AES 90th Conv.*, Paris, France, Feb. 1991, paper no. 5679.
- [21] J. Frenette, "Reducing artificial reverberation algorithm requirements using time-variant feedback delay networks," M.S. thesis, University of Miami, FL, USA, 2000.

- [22] T. Lokki and J. Hiipakka, “A time-variant reverberation algorithm for reverberation enhancement systems,” in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Limerick, Ireland, Dec. 2001, pp. 28–32.
- [23] S. J. Schlecht and E. A. P. Habets, “Time-varying feedback matrices in feedback delay networks and their application in artificial reverberation,” *J. Acoust. Soc. Am.*, vol. 138, no. 3, pp. 1389–1398, Sept. 2015.
- [24] S. J. Schlecht and E. A. P. Habets, “Dense reverberation with delay feedback matrices,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2019, pp. 150–154.
- [25] J.-M. Jot, “Efficient models for reverberation and distance rendering in computer music and virtual audio reality,” in *Proc. Int. Computer Music Conf.*, Thessaloniki, Greece, Sept. 1997.
- [26] M. Karjalainen and H. Järveläinen, “Reverberation modeling using velvet noise,” in *Proc. Audio Eng. Soc. 30th Int. Conf. Intelligent Audio Environments*, Saariselkä, Finland, Mar. 2007.
- [27] K. S. Lee, J. S. Abel, V. Välimäki, T. Stilson, and D. P. Berners, “The switched convolution reverberator,” *J. Audio Eng. Soc.*, vol. 60, no. 4, pp. 227–236, Apr. 2012.
- [28] B. Holm-Rasmussen, H.-M. Lehtonen, and V. Välimäki, “A new reverberator based on variable sparsity convolution,” in *Proc. 16th Int. Conf. Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, Sept. 2013, pp. 344–350.
- [29] V. Välimäki, B. Holm-Rasmussen, B. Alary, and H.-M. Lehtonen, “Late reverberation synthesis using filtered velvet noise,” *Appl. Sci.*, vol. 7, no. 483, May 2017.
- [30] B. Alary, A. Politis, and V. Välimäki, “Velvet-noise decorrelator,” in *Proc. Int. Conf. Digital Audio Effects (DAFx-17)*, Edinburgh, UK, Sept. 2017, pp. 405–411.
- [31] S. J. Schlecht, B. Alary, V. Välimäki, and E. A. P. Habets, “Optimized velvet-noise decorrelator,” in *Proc. Int. Conf. Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, Sept. 2018, pp. 87–94.
- [32] S. J. Schlecht and E. A. P. Habets, “Feedback delay networks: Echo density and mixing time,” *IEEE/ACM Trans. Audio, Speech and Lang. Process.*, vol. 25, no. 2, pp. 374–383, Feb. 2017.
- [33] H. P. Tukuljac, V. Pulkki, H. Gamper, K. Godin, I. J. Tashev, and N. Raghuvanshi, “A sparsity measure for echo density growth in general environments,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 226–230.
- [34] V. Välimäki, H.-M. Lehtonen, and M. Takanen, “A perceptual study on velvet noise and its variants at different pulse densities,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 7, pp. 1481–1488, Jul. 2013.
- [35] N. Werner, S. Balke, F.-R. Stöter, M. Müller, and B. Edler, “trackswitch.js: A versatile web-based audio player for presenting scientific results,” in *Proc. 3rd Web Audio Conf.*, London, UK, Aug. 2017.
- [36] D. Rocchesso, “Maximally diffusive yet efficient feedback delay networks for artificial reverberation,” *IEEE Signal Process. Lett.*, vol. 4, no. 9, pp. 252–255, Sept. 1997.

7. APPENDIX

7.1. Spectral Deviation

Let $w_1(n)$ and $w_2(n)$ be two Gaussian noise sequences of length L . Then, the energy of $w_j(n)$ is given by

$$\|w_j(n)\|_2^2 = \sum_{n=0}^L |w_j(n)|^2. \quad (16)$$

If $w_1(n)$ and $w_2(n)$ are uncorrelated and have each a normalized energy of 1, then the energy of the scaled sum is

$$\|\gamma_1 w_1(n) + \gamma_2 w_2(n)\|_2^2 = |\gamma_1|^2 + |\gamma_2|^2, \quad (17)$$

where γ_1 and γ_2 are two scalar gains. This relation can be readily extended to a summation of an $N \times N$ matrix of uncorrelated and normalized noise sequences $\mathbf{W}(n)$ and a matrix of scalar gains $\mathbf{\Gamma}$, i.e.,

$$\left\| \sum_{i,j=0}^N \Gamma_{ij} W_{ij}(n) \right\|_2^2 = \sum_{i,j=0}^N |\Gamma_{ij}|^2, \quad (18)$$

or more succinctly

$$\|\mathbf{\Gamma} \circ \mathbf{W}(n)\|_F^2 = \|\mathbf{\Gamma}\|_F^2, \quad (19)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Now, for $\mathbf{\Gamma} = \mathbf{c}\mathbf{b}^\top$, which is a rank-1 matrix, the Frobenius norm is expressed as

$$\|\mathbf{\Gamma}\|_F^2 = \|\mathbf{c}\|_2 \|\mathbf{b}\|_2. \quad (20)$$

The spectral coloration is the energy ratio between summation with and without gain matrix $\mathbf{\Gamma}$, i.e.,

$$E = \frac{\|\mathbf{\Gamma} \circ \mathbf{W}(n)\|_F^2}{\|\mathbf{W}(n)\|_F^2} = \frac{\|\mathbf{c}\|_2 \|\mathbf{b}\|_2}{N^2}. \quad (21)$$

Analogously, the same considerations can be made for individual frequency bands, and a frequency-dependent spectral coloration from a frequency-dependent gain matrix can be derived.

7.2. Serialized Structure

If the gains \mathbf{c} and \mathbf{b} are replaced with VNS filters $\mathbf{v}_1(z)$ and $\mathbf{v}_2(z)$, respectively, we note in Sec. 4.2 that the gain matrix in (14) becomes

$$\mathbf{\Gamma}(z) = \mathbf{v}_1(z)\mathbf{v}_2(z)^\top, \quad (22)$$

and the resulting spectral coloration is given in (15). However, if we connect the filters $\mathbf{v}_1(z)$ and $\mathbf{v}_2(z)$ in series at the input side of the FDN, the resulting gain matrix can be written as

$$\mathbf{\Gamma}(z) = \mathbf{1}[\mathbf{v}_1(z) \circ \mathbf{v}_2(z)]^\top, \quad (23)$$

and, similarly, if the series connection of filters $\mathbf{v}_1(z)$ and $\mathbf{v}_2(z)$ are placed at the output side, the gain matrix becomes

$$\mathbf{\Gamma}(z) = [\mathbf{v}_1(z) \circ \mathbf{v}_2(z)]\mathbf{1}^\top. \quad (24)$$

The resulting spectral coloration for the gain matrices in (23) and (24) is then

$$E(\omega) = \frac{\|\mathbf{1}\|_2 \|\mathbf{v}_1(e^{j\omega}) \circ \mathbf{v}_2(e^{j\omega})\|_2}{N^2}. \quad (25)$$