Prawda, Karolina; Schlecht, Sebastian; Välimäki, Vesa

Multichannel Interleaved Velvet Noise

Published in:
Proceedings of the 25th International Conference on Digital Audio Effects (DAFx20in22)

Published: 08/09/2022

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

MULTICHANNEL INTERLEAVED VELVET NOISE

Karolina Prawda *, Sebastian J. Schlecht † and Vesa Välimäki

Acoustics Lab
Department of Signal Processing and Acoustics
Aalto University
FI-02150 Espoo, Finland
{karolina.prawda, sebastian.schlecht, vesa.valimaki}@aalto.fi

ABSTRACT

The cross-correlation of multichannel reverberation generated using interleaved velvet noise is studied. The interleaved velvet-noise reverberator was proposed recently for synthesizing the late reverb of an acoustic space. In addition to providing a computationally efficient structure and a perceptually smooth response, the interleaving method allows combining its independent branch outputs in different permutations, which are all equally smooth and flutter-free. For instance, a four-branch output can be combined in 4! or 24 ways. Additionally, each branch output set is mixed orthogonally, which increases the number of permutations from $M!$ to $M^2!$, since sign inversions are taken along. Using specific matrices for this operation, which change the sign of velvet-noise sequences, decreases the correlation of some of the combinations. This paper shows that many selections of permutations offer a set of well decorrelated output channels, which produce a diffuse and colorless sound field, which is validated with spatial variation. The results of this work can be applied in the design of computationally efficient multichannel reverberators.

1. INTRODUCTION

Velvet noise is a sparse quasi-random noise that consists of pulses assuming values -1 and 1, while zeros make up to 95% of all samples [1, 2]. Thus, convolution of an audio signal with velvet noise is extremely cheap in terms of computational cost [3, 4, 5]. This property, combined with the smooth perceptual quality of velvet noise [2] makes it a suitable tool in audio processing and synthesis [5, 6]. This paper focuses on the use of velvet noise in multichannel artificial reverberation and its properties regarding decorrelation.

One of the prominent applications of velvet noise is artificial reverberation. After the early reflections, reverberation begins to resemble noise with a decaying envelope, which can be modeled with a pseudo-random signal [7, 8, 9]. This led to numerous solutions employing velvet-noise artificial reverberation synthesis [1, 10, 3, 11], with the most recent one utilizing interleaved velvet noise (IVN) [12].

The majority of studies regarding velvet-noise reverberation focuses on an accurate reproduction of room impulse responses while maintaining low computational complexity. The topic of multichannel reproduction is somewhat less prominent [13], especially in comparison to an extensive literature on spatial reverberation synthesis [14], particularly with feedback delay networks [15, 16, 17].

In multichannel sound reproduction, cross-correlation between the channels impacts the auditory image. For instance, the auditory image of well decorrelated signals is perceived as wider than when the correlation is high [18, 19]. On one extreme, fully correlated monophonic signals are perceived as located inside of the listener's head. On the other, fully decorrelated ones might be perceived as originating from completely separated sources [20]. Thus, decorrelation is a useful tool in many applications, including artificial reverberation, where all the channels are ideally fully decorrelated [21, 22, 14]. Among various methods to lower the similarity between spatially reproduced signals [23, 24], velvet-noise-based techniques assumed a prominent spot [18, 19]. In his early publication on artificial reverberation, Schroeder proposes to use a so-called resistance matrix to alternate the signs of comb-filter outputs in an attempt to recreate properties of diffuse spatial reverberation [25].

This paper proposes to extend the IVN in achieving multiple mutually decorrelated signals. To minimize the additional computation, we refrain from using additional decorrelating filters, but rely exclusively on a single recombination of the IVN output using delay permutations and orthogonal mixing.

The paper is organized as follows. Section 2 describes the synthesis of the IVN. In Section 3, we discuss the correlation of IVN permutations and propose a way to further decrease it by orthogonal mixing. Section 4 presents the evaluation of the proposed algorithm, and Section 5 concludes the paper.

2. INTERLEAVED VELVET-NOISE REVERBERATOR

The main challenge of previous velvet-noise reverberation algorithms is caused by reusing a single velvet-noise sequence, which leads to audible repetition in the produced sound, reminiscent of flutter echo [12]. Attempts at overcoming this problem by using time-variant randomization of impulses introduced yet another artifact in the synthesized reverb: warbling [1, 10, 13, 12].

The interleaved velvet-noise (IVN) technique [12] hides the repetition by simultaneously using several extended velvet-noise (EVN) sequences [2] in parallel. The pulse location k_{EVN} in an EVN is defined by

$$k_{\mathrm{EVN}}(m) = \mathrm{round}[m\widehat{T}_{\mathrm{d}} + \Delta r_1(m)(\widehat{T}_{\mathrm{d}} - 1)], \quad (1)$$

where m is the pulse counter, $\widehat{T}_{\mathrm{d}} = M\,T_{\mathrm{d}}$, where T_{d} is the grid size (average distance between consecutive pulses) and M is

* This work was supported by the Nordic Sound and Music Computing Network—NordicSMC, NordForsk project number 86892.

† Also at: Media Lab, Department of Art and Media, Aalto University, FI-02150 Espoo, Finland

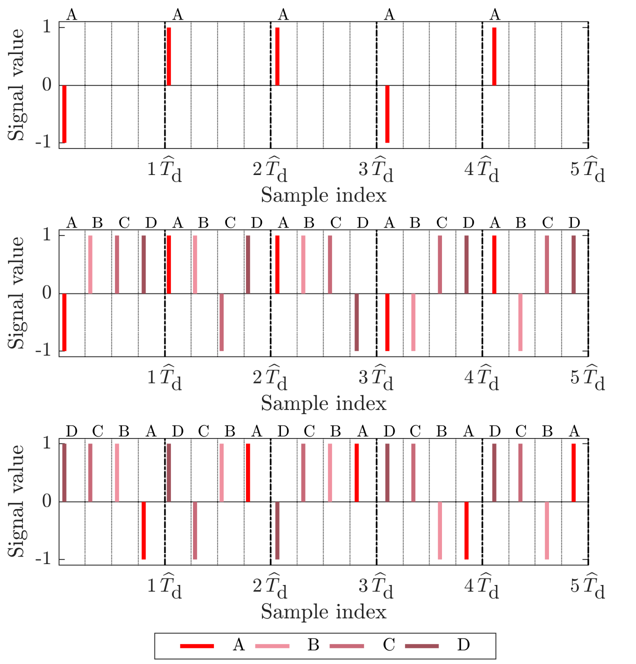

Figure 1: *(Top) Single EVN sequence (A), when $\Delta = 0.25$. (Middle) Four EVN sequences summed to an IVN signal (ABCD). Each EVN sequence is marked with a different shade of red. (Bottom) The same four IVN sequences, reordered (DCBA). Vertical dotted lines symbolize the grid size T_d, whereas $\widehat{T}_\mathrm{d}$ is indicated with dash-dotted vertical lines.*

the total number of EVN sequences in the signal, $r_1(m)$ is a random number generator, drawing from a uniform distribution in the range $[0, 1]$, and $\Delta \in [0, 1]$ limits the range of the pulse location, and can be tied to M by $\Delta = 1/M$. For example, when $M = 4$, and consequently, $\Delta = 0.25$, the pulse will appear only in the first quarter of the grid, whilst the rest of the samples within one $\widehat{T}_\mathrm{d}$ are zeros. Further delaying each EVN sequence by T_d creates an interleaved signal with a target pulse density [12]. The IVN signal composed of four EVN sequences is presented in Fig. 1.

IVN reverberation is made by convolving the attenuated and delayed input signal with an EVN sequence which has the same length as the delay line [12]. The output of each branch of the reverberator is offset by T_d and summed, creating a dense impulse response [12]. An example of such a structure is depicted in Fig. 2 (and also in Fig. 3), where $G_1(z), ..., G_M(z)$ represent feedback comb filters, each consisting of a delay line and an attenuation filter to obtain the frequency-dependent decay. The transfer function of $G_m(z)$ for $m = 1, ..., M$ is

$$G_m(z) = \frac{1}{1 - H_m(z)\, z^{-C_m \widehat{T}_\mathrm{d}}}\,, \tag{2}$$

where $H_m(z)$ is the attenuation filter and $C_m \widehat{T}_\mathrm{d}$ is the delay-line length in samples. The delay-line lengths of the feedback comb filters are the same as the lengths of the EVN signals, where C_m are prime numbers [12]. In Fig. 2, the blocks EVN1, ..., EVNM symbolize the EVN sequences, and the final offsets by T_d samples are indicated with delay lines.

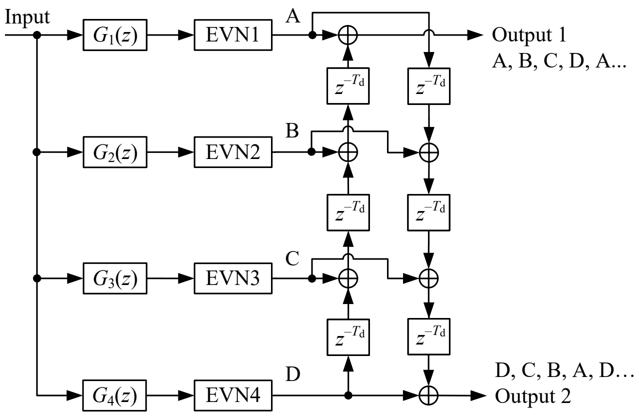

Figure 2: *Example stereo IVN structure ($M = 4$) producing two decorrelated output channels by delaying and combining the four branch signals (A, B, C, and D) in opposite orders [12]. These are only two of the 24 possible permutations.*

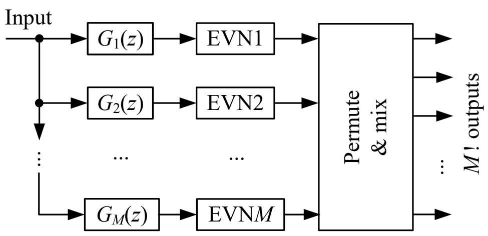

Figure 3: *Basic multichannel IVN structure producing $M!$ different output channels by permuting and mixing the M branch signals.*

In the present work, the properties of the IVN reverberator are presented on an example comprised of four EVN sequences ($M = 4$) having the lengths of 7760, 8080, 8240, and 8560 samples, which correspond to $\widehat{T}_\mathrm{d} = 80$ multiplied by prime numbers $C_1 = 97$, $C_2 = 101$, $C_3 = 103$, and $C_4 = 107$, respectively. These parameter values were found to be a good trade-off between the perceptual qualities of the velvet noise and the computational cost of the reverberation synthesis [12]. At the sample rate of $f_\mathrm{s} = 44.1$ kHz, these EVN sequence lengths correspond to 176 ms, 183 ms, 187 ms, and 194 ms.

3. CORRELATION OF IVN PERMUTATIONS

This section presents the analysis of the cross-correlation of different permutations of the IVN signals. In addition, orthogonal mixing is proposed as a way to extend the possible permutations and to reduce the cross-correlation between the IVN signals.

3.1. Correlation of Basic Permutations

Multichannel IVN reverberation can be achieved by reusing the same set of EVN sequences whilst changing their order in the output, thus preserving low-memory cost. This is shown in Fig. 2 for a stereo case, where the same four EVN sequences are offset

Table 1: *Permutations of four EVN sequences. Each sequence is represented with a capital letter A, B, C, or D, as in Fig. 1.*

Permutation #		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Offset in samples	0	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	C	C	C	D	D	D	D	D	D
	T_d	B	B	C	C	D	D	A	A	C	C	D	D	A	A	B	B	D	D	A	A	B	B	C	C
	$2T_\mathrm{d}$	C	D	B	D	B	C	C	D	A	D	A	C	B	D	A	D	A	B	B	C	A	C	A	B
	$3T_\mathrm{d}$	D	C	D	B	C	B	D	C	D	A	C	A	D	B	D	A	B	A	C	B	C	A	B	A

by different amounts of samples for the two stereo channels. The number of output channels produced with this simple operation is equal to the total number of permutations of all sequences, $M!$. The general multichannel case is illustrated in Fig. 3.

For the case of the IVN comprised of four sequences, all the permutations and their respective offsets are listed in Table 1. Some permutations share one or two sequences in the same place, i.e., the same sequences have equal offsets in both permutations, e.g., permutations #1 and #2 both start with the sequence A followed by the sequence B offset by T_d samples. On the other hand, some permutations do not share any sequences, e.g., permutations #1 and #24 have all sequences in different orders.

In the following, we describe how the cross-correlation is affected between different permutations. The i^{th} permutation is a unique ordering $\sigma_i(m) \in \{1, \ldots, M\}$ for $m = 1, \ldots, M$. The output signal of the i^{th} permutation is

$$y_i(k) = \sum_{m=1}^{M} x_m(k + \sigma_i(m) T_\mathrm{d}), \tag{3}$$

where k denotes the time index, x_m are the M EVN sequences, and $i = 1, 2, ..., M!$ numbers the permutations.

The correlation of IVN permutations is best estimated with measures that account for both the similarity of the two sets of data and the relative displacement of that similarity in time (lag). Therefore, in this study, the properties of IVN permutations are assessed by cross-correlation

$$R_{y_i,y_j}(n) = \sum_{k=1}^{N} y_i(k)\, y_j(k+n), \tag{4}$$

where n is the time lag in samples. The correlation between two outputs y_i and y_j depend on how different the permutations are, i.e., the number of shared sequences.

To this end, we introduce the Hamming distance between two permutations i and j

$$d_\mathrm{H}(i,j) = \sum_k \delta_{\sigma_i(k),\sigma_j(k)}, \tag{5}$$

where δ denotes the Kronecker delta. The possible values of $d_\mathrm{H}(i,j)$ according to Table 1 are: $d_\mathrm{H}(i,j) = 4$ if the sequence order in two permutations is different, $d_\mathrm{H}(i,j) = 0$ if the sequence order is identical, and $d_\mathrm{H}(i,j) = 3$ or $d_\mathrm{H}(i,j) = 2$ for one and two shared sequences, respectively.

In the case of the IVN permutations, however, the Hamming distance is time-dependent. An example of this is permutation #4 in Table 1, ACDB. When compared to permutation #1, ABCD, $d_\mathrm{H}(1,4) = 3$. However, shifting ACDB by 1 position results in a larger overlap, i.e., $d_\mathrm{H}^1(1,4) = 2$ for the shifted Hamming distance

$$d_\mathrm{H}^m(i,j) = \sum_k \delta_{\sigma_i(k),\sigma_j(k+m)}. \tag{6}$$

We know that the EVN sequences are fully decorrelated, i.e., $R_{x_i,x_j}(n) = \delta_{ij}$ for energy-normalized sequences, i.e., $E[x_i] = \sum_{k=1}^{N} |x_i|^2 = 1$. Thus, the number of fully correlated sequences between two permutations at lag $n = mT_\mathrm{d}$ determines the cross correlation, i.e.,

$$R_{y_i,y_j}(mT_\mathrm{d}) = M - d_\mathrm{H}^m(i,j). \tag{7}$$

As we are interested in the largest overlap, we define the maximum shifted Hamming distance, describing the maximal number of shared EVN sequences when one of the permutations is shifted

$$\widehat{d}_\mathrm{H}(i,j) = \min_m d_\mathrm{H}^m(i,j), \tag{8}$$

where $m = -M+1, ..., 0, ..., M-1$.

To improve the readability of the results, all of the values are normalized, so that the autocorrelation of analyzed signals is equal to one at zero lag $n = 0$. Hence, the cross-correlation is scaled to assume values from the range $[-1, 1]$:

$$\widehat{R}_{y_i,y_j}(n) = \frac{R_{y_i,y_j}(n)}{\sqrt{R_{y_i,y_i}(0)\, R_{y_j,y_j}(0)}}. \tag{9}$$

Thus Eq. (9) is rewritten following similar logic as in [26]

$$\widehat{R}_{y_i,y_j}(mT_\mathrm{d}) = \frac{M - d_\mathrm{H}^m(i,j)}{M}. \tag{10}$$

Two examples of the normalized cross-correlation of IVN permutations are presented in Fig. 4. In the top pane, the permutations have a $d_\mathrm{H}^m(i,j) = 2$ at $n = 0$ and $d_\mathrm{H}^m(i,j) = 3$ at $n = \pm T_\mathrm{d}$. In the bottom pane, $d_\mathrm{H}^m(i,j) = 3$ at $n = \{-3T_\mathrm{d}, -T_\mathrm{d}, T_\mathrm{d}, 3T_\mathrm{d}\}$, and $d_\mathrm{H}^m(i,j) = 4$ at $n = 0$.

In the present work, the maximal values of normalized cross-correlation are of interest, since they highlight the worst-case scenario of similarity between two signals. They can be stored in a matrix $\mathbf{R}$, defined as

$$\mathrm{R}_{i,j} = \max_n \left|\widehat{R}_{y_i,y_j}(n)\right|. \tag{11}$$

Eq. (10) established the relation between the values of cross- correlation and the Hamming distance between sequence orders. Thus, the highest $\widehat{R}_{y_i,y_j}$ is obtained with the lowest Hamming distance, i.e., the shifted Hamming distance $\widehat{d}_\mathrm{H}(i,j)$

$$\mathrm{R}_{i,j} = \left|\frac{M - \widehat{d}_\mathrm{H}(i,j)}{M}\right|. \tag{12}$$

The matrix $\mathbf{R}$ for $M = 4$ is displayed in Fig. 5. The diagonal presents the normalized autocorrelation of IVN signals, which is equal to one (black cells). The values of the remaining elements of the matrix are tied to the $\widehat{d}_\mathrm{H}(i,j)$ values. The correlation of IVN signals with $\widehat{d}_\mathrm{H}(i,j) = 2$ oscillates around 0.5 (dark gray cells in Fig. 5). For the pairs with $\widehat{d}_\mathrm{H}(i,j) = 3$, the correlation is close to 0.25 (light gray cells in Fig. 5).

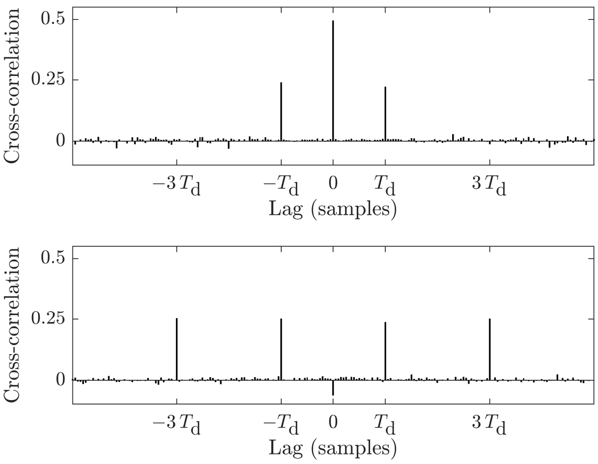

Figure 4: *Normalized cross-correlation of IVN permutations with (top) $\widehat{d}_H(i,j) = 2$ at $n = 0$ (e.g., ABCD and BACD) and (bottom) $\widehat{d}_H(i,j) = 3$ at $n = \{-3T_\mathrm{d}, -T_\mathrm{d}, T_\mathrm{d}, 3T_\mathrm{d}\}$ (e.g., ABCD and DCBA). The time lag n is relative to one of the signals, thus negative values are possible.*

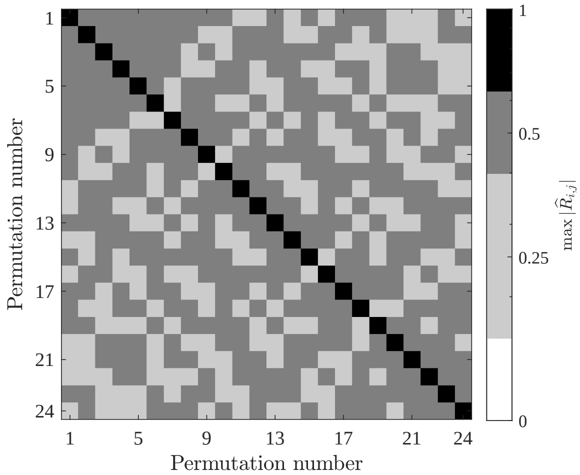

Figure 5: *Maximal cross-correlation for permutations of four IVN sequences given in Table 1. The cross-correlation values are calculated using Eq. (11).*

3.2. Subset of Decorrelated IVN Permutations

In many cases of spatial reproduction the number of channels is lower than the number of possible IVN permutations. In such situations, the combinations that show the smallest similarity among each other can be selected to achieve the best decorrelation.

The subset of K IVN signals with the lowest correlation within a set of $M!$ permutations may be found by solving the thinnest K-subgraph problem [19, 27]. In such an approach, the matrix $\mathbf{R}$ is used as an adjacency matrix of a graph. However, for small values of M, such as in the four-branch case ($M = 4$), a brute-force search entailing the examination of all possible combinations is feasible as well. The estimator of the lowest correlation may be either mean or median of the elements of $\mathbf{R}$. When the mean is used, the problem is formulated as

$$\min \frac{1}{K} \sum_{i,j \in \kappa} \mathrm{R}_{i,j} \, , \tag{13}$$

where $\kappa \in \{1, ..., M!\}$ and $|\kappa| = K$.

The results of searching for the permutations with the lowest correlation are shown in Fig. 6 for the mean and the median of the elements in $\mathbf{R}$ and the number of channels K from two to twenty-four. Both estimators show approximately the same values up to four channels. Between five and nine channels, the results obtained with median are lower, as less than half of the IVN permutations have a higher correlation. This changes for a ten-channel case and is reflected by a dramatic jump in the median correlation.

The mean values in Fig. 6 show a steady increase in normalized correlation, as the mean is more susceptible to outliers than median [28, 29, 26]. Since in this work we are interested in spotting the presence of highly correlated permutations, we choose mean as a correlation estimator in the remainder of this paper.

3.3. Orthogonal Mixing of IVN Signals

The ability of the IVN reverberator to produce a set of decorrelated signals can be extended by the means of orthogonal mixing. Knowing that

$$\mathbf{R} = \mathbb{E}[\mathbf{Y}\mathbf{Y}^\mathrm{T}], \tag{14}$$

where $\mathbb{E}[\cdot]$ is the expected value and $\mathbf{Y}$ contains signals $y_1, ...y_M$, we can perform multiplication with orthogonal matrix $\mathbf{A}$, which has the property $\mathbf{A}\mathbf{A}^\mathrm{T} = \mathbf{I}$ and has the same size as $\mathbf{Y}$. Thus

$$\mathbf{R} = \mathbb{E}[\mathbf{Y}\mathbf{A}\mathbf{A}^\mathrm{T}\mathbf{Y}^\mathrm{T}] = \mathbb{E}[\mathbf{Y}\mathbf{Y}^\mathrm{T}], \tag{15}$$

which proves that the cross-correlation of permutations does not increase after such an operation.

Choosing an appropriate orthogonal matrix allows for a reduction of correlation of some permutations from the set. From a set of $M! = 24$ IVN combinations let us consider a subset of $K = 4$, which are picked based on their mean normalized cross-correlation, as described in Sec. 3.2. Here, we propose to perform orthogonal mixing using a Hadamard matrix, which for size $K = 4$ is defined as

$$\mathbf{H}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \tag{16}$$

The Hadamard matrix changes the sign of two EVN sequences in each combination. An example of a stereo IVN structure in which two sequences are inverted according to the orthogonal mixing principle in shown in Fig. 7. The Hadamard matrix also offers advantages of fast and cheap computation (change of sign does not require multiplications, only a subtraction). However, such a matrix does not exist for every M, e.g., $M = 5$. In such cases, a different type of orthogonal matrix is needed.

The permutations for the orthogonal mixing were selected based on their mean cross-correlation, as described in Sec. 3.2. The order of sequences resulting from the orthogonal mixing is listed in Table 2. The matrix $\mathbf{R}$ for these permutations is displayed in Fig. 8. The light gray cells show the normalized cross-correlation value around 0.25 (*cf.* Fig. 5), whereas the white cells indicate the cross-correlation close to zero.

Table 2: *Orthogonal mixing of four IVN permutations with the Hadamard matrix (cf. Table 1).*

Permutation #		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Offset in samples	0	A	A	A	A	B	B	B	B	C	C	C	C	D	D	D	D
	T_d	B	−B	B	−B	D	−D	D	−D	A	−A	A	−A	C	−C	C	−C
	$2\,T_\mathrm{d}$	C	C	−C	−C	A	A	−A	−A	D	D	−D	−D	B	B	−B	−B
	$3\,T_\mathrm{d}$	D	−D	−D	D	C	−C	−C	C	B	−B	−B	B	A	−A	−A	A

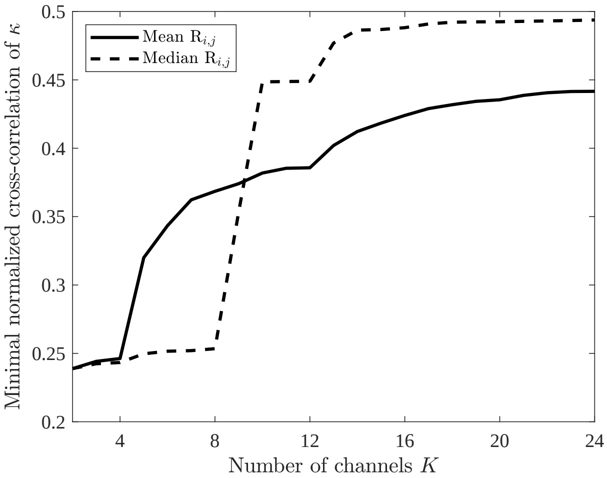

Figure 6: *Minimal values of mean and median cross-correlation for four-sequence IVN depending on the number of channels K for reproduction.*

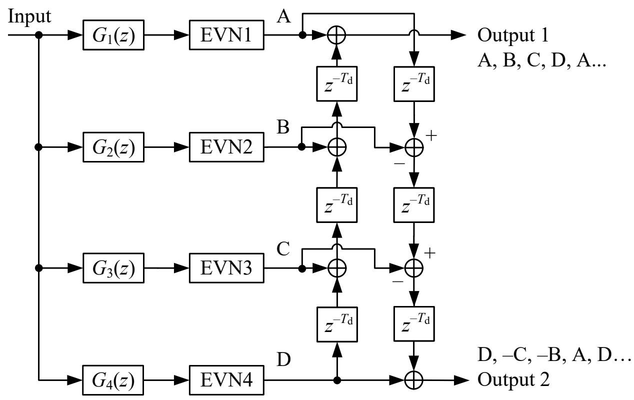

Figure 7: *Orthogonally mixed version of the stereo IVN structure of Fig. 2. Here, the signs of two sequences are inverted for the second output, which further improves the decorrelation between the output channels.*

The values of correlation for the permutations with $\widehat{d}_\mathrm{H}(i,j) = 2$, but the sign of two of them is negative, is visibly decreased compared to permutations for which $\widehat{d}_\mathrm{H}(i,j) = 3$ but the sign is not flipped. For example, the normalized cross-correlation value for permutations #1 and #2 from Table 2, ABCD and A−BC−D, is close to zero, whereas for permutations #1 and #13, ABCD and DCBA, it is 0.25. This is due to the fact that the EVN sequences with opposite signs are anticorrelated, i.e., their normalized cross-correlation is -1. Having two fully correlated and two fully anticorrelated sequences cancels out the peak expected at $n = 0$, bringing the maximal cross-correlation value to zero.

Two EVN sequences with the same pulse locations but opposite pulse signs will cancel each other if they arrive at the listening position at the same time. This might happen e.g., when the listener is at the midpoint between loudspeakers. To avoid the cancellation, the orthogonally mixed IVN signals are offset by a small number of samples. In this work, a random offset by up to 120 samples was found to be sufficient. The synthesized audio examples of multichannel IVN signals are available online[1].

4. EVALUATION

This section presents the evaluation of the proposed algorithm in terms of the spectral properties of the multichannel reverberation, as well as the computational cost of creating basic and orthogonally-mixed permutations. The assessment concerns the example reverberation as described in the previous sections of this paper.

[1] `http://research.spa.aalto.fi/publications/papers/dafx22-multichannel-ivn/`

4.1. Spatial Variance

An unwanted phenomenon in multichannel reproduction is called the comb-filtering effect. It occurs when two versions of the signal arrive at the receiver at different times, which is caused by different path lengths between the sound sources and the receiver. The resulting changes in phase between the delayed versions of the signal cause cancellation of frequencies when the phase difference is $180°$. The bigger the difference in path lengths, the longer the delay and consequently, the lower the cancellation frequency [30].

A method to assess the amount of comb-filtering effect of a multichannel setup is to compute its spatial variance (SV). The SV is a measure of the magnitude-response variation for a frequency range of interest and fixed audience area and is expressed in dB [30, 31, 32, 33]. The SV is defined as

$$\mathrm{SV} = \frac{1}{N_f} \sum_{f=f_{lo}}^{f_{hi}} \sqrt{\frac{1}{N_p - 1} \sum_{p=1}^{N_p} \left(L_\mathrm{p}(p,f) - \overline{L_\mathrm{p}(f)} \right)^2}, \quad (17)$$

where N_f is the number of frequency bins between f_{lo} and f_{hi}, N_p is the number of measurement points in the audience area, $L_\mathrm{p}(p,f)$ is the sound pressure level in dB and measurement point p and frequency bin f, and $\overline{L_\mathrm{p}(f)}$ is the mean sound pressure level in dB over all measurement points at frequency bin f. A SV of 0 dB indicates that there is no magnitude deviation between listening positions, while the higher the SV, the bigger the deviation.

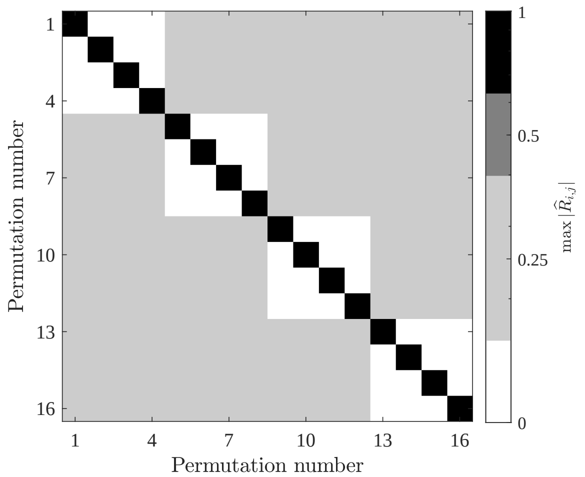

Figure 8: *Cross-correlation of four IVN sequences with the minimal mean correlation values (cf. Fig. 6) after the orthogonal mixing with the Hadamard matrix. The permutation numbering is based on Table 2. The light gray cells indicate the cross-correlation of about 0.25, the same as the best cells in Fig. 5. The white cells correspond to cases in which the correlation has been further decreased by the orthogonal mixing.*

In the present study, SV was calculated for two sets of multichannel IVN reverberators: one composed of the basic 24 permutations and the other consisting of 16 orthogonally mixed permutations. The $N_p = 50$ measurement points over the audience area were simulated by adding random delays to all but one signal in each set. That way the relative delay in arrival of sound at the listening position was replicated. The delays were chosen so that the distance between the receiver points was up to seven meters. For the calculations, the magnitude spectra at listening positions were smoothed in 1/12th-octave bands.

Figure 9 shows $L_\mathrm{p}(p, f)$ and $\overline{L_\mathrm{p}(f)}$ over the whole frequency spectrum for the three evaluated cases: monophonic one in the top pane, 24 basic permutations in the middle pane, and 16 orthogonally mixed permutations in the bottom pane. Although some comb-filtering occurs in all three examples, the peaks and notches are visibly flatter going from the top to the bottom pane. The reduction is particularly visible between 100 Hz and 1 kHz. Gray patches in all three panes of Fig. 9 show the areas within ± 6 dB from the $\overline{L_\mathrm{p}(f)}$. The permuted IVN reverberation stays within these limits in a broader frequency range compared to the monophonic case, especially above around 600 Hz.

The SV was calculated over two frequency ranges: 20 Hz to 500 Hz, following the previous research that focused on SV in low frequencies [30, 31, 32, 33], as well as for the whole frequency range from 20 Hz to 20 kHz. The results of the calculations in Table 3 are presented as a percentage of the SV reduction in relation to the monophonic case, where all delayed signals are the same. The numbers in Table 3 show that both sets of IVN signals offer 12% lower SV than mono in the low-frequency range. The advantage of the permuted IVN is more significant in Table 3 for the full audio spectrum (20 Hz to 20 kHz), where the reduction of the SV is over 30% for the 24 basic permutations and over 40% for the 16 orthogonally-mixed permutations.

The SV of the proposed algorithm was compared with two other methods to produce multichannel reverberation: using white-noise signals and the approach introduced by Schroeder [25]. In the former technique, different white-noise signal was fed to each channel, creating a set of maximally decorrelated outputs with minimal SV, treated here as a reference for the remaining methods.

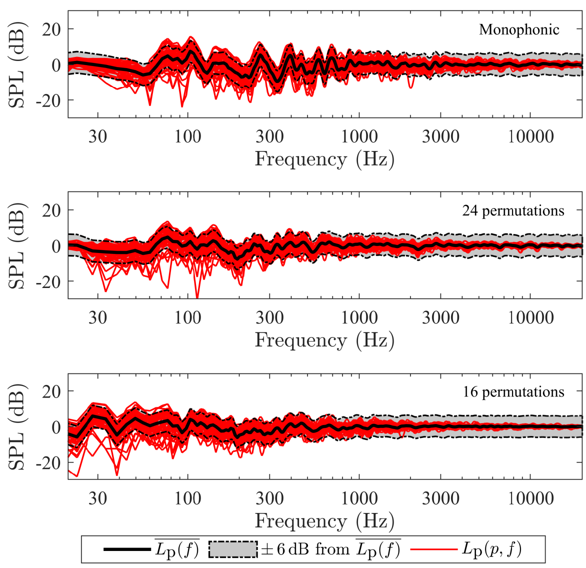

Figure 9: *Magnitude responses of (top) set of the same signals, (middle) 24 IVN permutations, and (bottom) 16 orthogonally mixed IVN permutations. All the magnitude responses have been 1/12th-octave smoothed.*

Table 3: *Reduction of SV for 24 and 16 IVN permutations in comparison to the monophonic case.*

Frequency Range	SV Reduction w.r.t. Mono	
	24 Permutations	16 Permutations
20–500 Hz	12%	12%
20–20k Hz	32%	43%

The Schroeder reverberator was implemented utilizing four parallel branches, each consisting of two all-pass filters (the same ones for all branches) and one feedback comb filter. The output of each branch was passed through a matrix containing 1's and -1's that affected the sign of output signals accordingly. In result, 16 channels were created (see Appendix for the block diagram and cross-correlation considerations). The delay-line lengths in the Schroeder reverberator were equal to the lengths of EVN sequences.

The SV of the white noise signals and Schroeder reverberator were calculated using the same setup as when calculated the SV for the IVN reverberators. The results of the comparison are presented in Table. 4. They show that all the analyzed signals display a high increase in the SV at low frequencies, with the IVN with 24 permutations being the most advantageous. In the case of broadband analysis, the rise in SV compared to white noise is less dramatic, with the orthogonally-mixed IVN leading the race. In both examples, the IVN reverberators outperform the Schroeder approach.

Table 4: *Comparison of SV between IVN with 24 and 16 permutations and the Schroeder reverberator, using the set of white-noise signals as a reference (100% SV). The best result in each row in shown in bold.*

Frequency Range	White Noise	Spatial Variance IVN 24	IVN 16	Schroeder
20–500 Hz	100%	**214%**	223%	255%
20–20k Hz	100%	157%	**136%**	200%

4.2. Computational Cost

Figure 2 shows that producing one decorrelated channel output, Output 2 in this case, requires only three short delay lines and three additions. This is considerably less processing than what is required for any single-channel reverberation algorithm. When orthogonal mixing is applied, some additions become subtractions but otherwise the computational requirements are the same.

In our example case, $T_\mathrm{d} = 20$ samples, which means that the three delay lines together consume only 60 samples of fast memory. This is the net increment in memory for one additional channel. For comparison, a reverberation algorithm often consumes nearly 1 s of sample memory for one channel. The IVN reverberator described in Sec. 2 uses 741 ms of sample memory [12]. Thus, the multichannel IVN method is very economical also in terms of memory consumption.

5. CONCLUSIONS

The paper discusses the application of IVN reverberation in multichannel audio reproduction. The reordering of the EVN sequences within the IVN structure achieves low correlation between a number of output channels, without the use of additional decorrelation filters. With M EVN sequences, the permutations yield $M!$ different output signals. This paper introduces orthogonal mixing of basic IVN permutations as well. It is shown that using a specific orthogonal matrix, e.g., a Hadamard matrix for a four-sequence case, allows for a further decrease of similarity between the IVN signals.

The evaluation of the proposed method shows that both sets of permutations—basic and orthogonally mixed—reduce the spatial variance of multichannel signals in comparison to the monophonic case. The comparison of magnitude responses reveals that pronounced peaks and notches—the result of comb-filtering—in the monophonic case are greatly flattened in permuted examples. Such a decrease is especially noticeable in mid- and high frequencies, reducing coloration in multichannel reproduction. The orthogonally mixed permutations yield the largest improvement. In a comparison with a similar approach to create multichannel reverberation, the IVN-based methods show a considerable improvement in terms of spatial variance.

The results of this paper offer an economical way to implement multichannel reverberation. Only one IVN reverb structure is needed, and the multiple decorrelated output signals are obtained by systematically delaying, inverting, and mixing its output signals. Thus, multichannel reverberation is no longer much more costly than monophonic reverberation.

6. REFERENCES

[1] M. Karjalainen and H. Järveläinen, "Reverberation modeling using velvet noise," in *Proc. Audio Eng. Soc. 30th Int. Conf. Intelligent Audio Environments*, Saariselkä, Finland, Oct. 2007.

[2] 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.

[3] B. Holm-Rasmussen, H.-M. Lehtonen, and V. Välimäki, "A new reverberator based on variable sparsity convolution," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Maynooth, Ireland, Sep. 2013, pp. 344–350.

[4] S. D'Angelo and L. Gabrielli, "Efficient signal extrapolation by granulation and convolution with velvet noise," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Aveiro, Portugal, Sep. 2018, pp. 107–112.

[5] V. Välimäki, J. Rämö, and F. Esqueda, "Creating endless sounds," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Aveiro, Portugal, Sep. 2018, pp. 32–39.

[6] J. Fagerström, S. J. Schlecht, and V. Välimäki, "One-to-many conversion for percussive samples," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Vienna, Austria, Sept. 2021, pp. 129–135.

[7] J. Moorer, "About this reverberation business," *Computer Music J.*, vol. 3, no. 2, pp. 13–28, 1979.

[8] P. Rubak and L. G. Johansen, "Artificial reverberation based on a pseudo-random impulse response, part I," in *Proc. Audio Eng. Soc. 104th Conv.*, Amsterdam, The Netherlands, May 1998, paper no. 4725.

[9] P. Rubak and L. G. Johansen, "Artificial reverberation based on a pseudo-random impulse response, part II," in *Proc. Audio Eng. Soc. 106th Conv.*, Munich, Germany, May 1999, paper no. 4900.

[10] K. S. Lee, J. S. Abel, V. Välimäki, T. Stilson, and D. B. Berners, "The switched convolution reverberator," *J. Audio Eng. Soc.*, vol. 60, no. 4, pp. 227–236, Apr. 2012.

[11] V. Välimäki, B. Holm-Rasmussen, B. Alary, and H-M. Lehtonen, "Late reverberation synthesis using filtered velvet noise," *Appl. Sci.*, vol. 7, no. 5, May 2017.

[12] V. Välimäki and K. Prawda, "Late-reverberation synthesis using interleaved velvet-noise sequences," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 29, pp. 1149–1160, Feb. 2021, https://ieeexplore.ieee.org/document/9360485.

[13] S. Oksanen, J. Parker, A. Politis, and V. Välimäki, "A directional diffuse reverberation model for excavated tunnels in rock," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Vancouver, Canada, May 2013, pp. 644–648.

[14] V. Pulkki and J. Merimaa, "Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.*, vol. 54, no. 1/2, pp. 3–20, Jan./Feb. 2006.

[15] J. Stautner and M. Puckette, "Designing multi-channel reverberators," *Computer Music J.*, vol. 6, no. 1, pp. 52–65, 1982.

[16] M. Kuster, “Multichannel room impulse response generation with coherence control,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 4, pp. 597–606, May 2009.

[17] 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.

[18] B. Alary, A. Politis, and V. Välimäki, “Velvet-noise decorrelator,” in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Edinburgh, UK, Sept. 2017, pp. 405–411.

[19] S. J. Schlecht, B. Alary, V. Välimäki, and E. A. P. Habets, “Optimized velvet-noise decorrelator,” in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Aveiro, Portugal, Sept. 2018, pp. 87–94.

[20] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Oct. 1996.

[21] G. S. Kendall, “The decorrelation of audio signals and its impact on spatial imagery,” *Computer Music J.*, vol. 19, no. 4, pp. 71–87, 1995.

[22] G. Potard and I. Burnett, “Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays,” in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Naples, Italy, Oct. 2004, pp. 280–284.

[23] E. K. Canfield-Dafilou and J. S. Abel, “Signal decorrelation using perceptually informed allpass filters,” in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Brno, Czech Republic, Sept. 2016, pp. 255–231.

[24] E. K. Canfield-Dafilou and J. S. Abel, “A group delay-based method for signal decorrelation,” in *Proc. 144th Audio Eng. Soc. Conv.*, May 2018.

[25] M. R. Schroeder, “Natural sounding artificial reverberation,” *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223, Jul. 1962.

[26] K. Prawda, S. J. Schlecht, and V. Välimäki, “Robust selection of clean swept-sine measurements in non-stationary noise,” *J. Acoust. Soc. Am.*, vol. 151, no. 3, pp. 2117–2126, Mar. 2022.

[27] U. Feige, D. Peleg, and G. Kortsarz, “The dense k-subgraph problem,” *Algorithmica*, vol. 29, no. 3, pp. 410—-421, Mar. 2001.

[28] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median,” *J. Exp. Soc. Psychol.*, vol. 49, no. 4, pp. 764–766, 2013.

[29] D. L. Donoho and P. J. Huber, “The notion of breakdown point,” *A festschrift for Erich L. Lehmann*, vol. 157184, 1983.

[30] J. B. Moore and A. J. Hill, “Dynamic diffuse signal processing for sound reinforcement and reproduction,” *J. Audio Eng. Soc.*, vol. 66, no. 11, pp. 953–965, Nov. 2018.

[31] A. Devantier and T. S. Welti, “In-room low frequency optimization,” in *Proc. 115th Audio Eng. Soc. Conv.*, New York, NY, USA, Oct. 2003.

[32] S. Birkedal Nielsen and A. Celestinos, “Optimizing placement and equalization of multiple low frequency loudspeakers in rooms,” in *Proc. 119th Audio Eng. Soc. Conv.*, New York, NY, USA, Oct. 2005.

[33] A. J. Hill and M. O. J. Hawksford, “On the perceptual advantage of stereo subwoofer systems in live sound reinforcement,” in *Proc. 135th Audio Eng. Soc. Conv.*, New York, NY, USA, Oct. 2013.

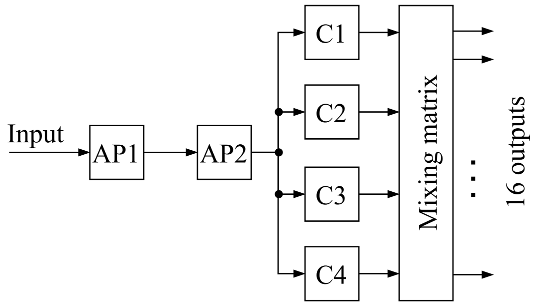

Figure 10: *Schroeder reverberator consists of two all-pass filters (AP1 and AP2) and four parallel branches with a feedback comb filter each (C1, C2, C3, C4). It is capable of producing 16 outputs by passing the comb-filtered signals through a matrix of 1's and −1's.*

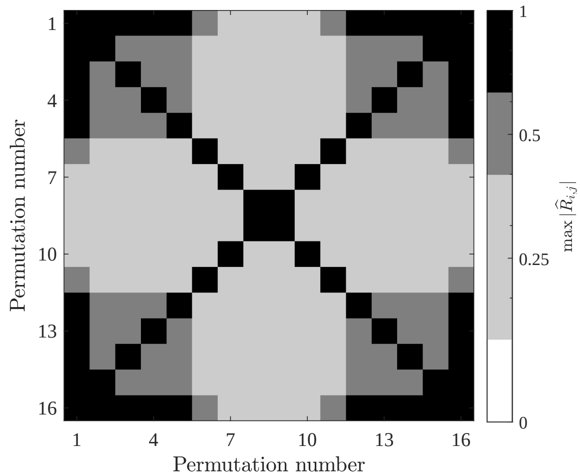

Figure 11: *Cross-correlation between channels of the Schroeder reverberator. Compared to cross-correlation of IVN permutations (*cf*. Figs. 5 and 8), the Schroeder system exhibits the biggest number of highly-correlated channels, represented here by the black cells.*

7. APPENDIX: CROSS-CORRELATION OF SCHROEDER REVERBERATOR CHANNELS

The structure of the Schroeder reverberator, as described in the original publication [25] and in Sec. 4.1, is presented in Fig. 10. The all-pass filters are symbolized by the blocks denoted with AP1 and AP2, whereas comb filters are marked with blocks C1, C2, C3 and C4. The cross-correlation between the channels of the structure was calculated according to Eqs. (9) and (11). The values stored in $\mathbf{R}$ for Schroeder's system are displayed in Fig. 11.

Compared to both, basic and orthogonally-mixed IVN permutations, the channels of Schroeder's algorithm show the biggest number of highly-correlated channels. Figs. 5 and 8 display only one diagonal of black cells, meaning that an IVN permutation is well-correlated only with itself. On the other hand, there are two such diagonals in Fig. 11, showing that individual channels exhibit high similarity with at least one other channel. Additionally, channels #1 and #16 display high values of cross-correlation with the majority of the remaining outputs.