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

*Published under the following license:*
CC BY

*Please cite the original version:*

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
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-channel 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

$$k_{EVN}(m) = \text{round}[m\hat{T}_d + \Delta r_1(m)(\hat{T}_d - 1)],$$

(1)

where $m$ is the pulse counter, $\hat{T}_d = M T_d$, where $T_d$ is the grid size (average distance between consecutive pulses) and $M$ is
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.

### 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.
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

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

where $k$ denotes the time index, $x_m$ are the $M$ EVN sequences, and $i = 1, 2, \ldots, M!$ 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 $j$

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

where $\delta$ denotes the Kronecker delta. The possible values of $d_H(i, j)$ according to Table 1 are: $d_H(i, j) = 4$ if the sequence order in two permutations is different, $d_H(i, j) = 0$ if the sequence order is identical, and $d_H(i, j) = 3$ or $d_H(i, j) = 2$ 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, $d_H(1, 4) = 3$. However, shifting ACDB by 1 position results in a larger overlap, i.e., $d_H(1, 4) = 2$ for the shifted Hamming distance

$$ d_H^M(i, j) = \sum_k \delta_{\sigma_i(k), \sigma_j(k+m)}, \tag{6} $$

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., $E[x_i] = \sum_{n=1}^{N} |x_i|^2 = 1$. Thus, the number of fully correlated sequences between two permutations at lag $n = m T_d$ determines the cross correlation, i.e.,

$$ R_{y_i, y_j}(mT_d) = M - d_H^n(i, j). \tag{7} $$

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

$$ \tilde{d}_H(i, j) = \min_m d_H^n(i, j), \tag{8} $$

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

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 $[-1, 1]$:

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

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

$$ \tilde{R}_{y_i, y_j}(mT_d) = \frac{M - d_H^n(i, j)}{M}. \tag{10} $$

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

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

$$ R_{i,j} = \max_n |\tilde{R}_{y_i, y_j}(n)|. \tag{11} $$

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

$$ R_{i,j} = \left| \frac{M - d_H^M(i, j)}{M} \right|. \tag{12} $$

The matrix $R$ 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 $d_H^M(i, j)$ values. The correlation of IVN signals with $d_H^M(i, j) = 2$ oscillates around 0.5 (dark gray cells in Fig. 5). For the pairs with $d_H^M(i, j) = 3$, the correlation is close to 0.25 (light gray cells in Fig. 5).

| 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 $T_d$ samples | 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_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_d | D C D B C B D C D A C A D B A B A C B C A B A |

Table 1: Permutations of four EVN sequences. Each sequence is represented with a capital letter A, B, C, or D, as in Fig. 1.
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 $R$ 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 $R$. When the mean is used, the problem is formulated as

$$\min \frac{1}{K} \sum_{i,j \in \kappa} R_{i,j},$$

where $\kappa \in \{1, \ldots, M!\}$ 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 $R$ 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

$$R = E[YY^T],$$

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

$$R = E[YYA^T Y^T] = E[YY^T],$$

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

$$H_4 = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1
\end{bmatrix},$$

The Hadamard matrix changes the sign of two IVN sequences in each combination. An example of a stereo IVN structure in which two sequences are inverted according to the orthogonal mixing principle is 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 $R$ 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.
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.

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

\[
SV = \frac{1}{N_f} \sum_{f=1}^{f_{hi}} \frac{1}{N_p - 1} \sum_{p=1}^{N_p} (L_p(p, f) - \bar{L}(f))^2,
\]

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, \(L_p(p, f)\) is the sound pressure level in dB and measurement point \(p\) and frequency bin \(f\), and \(\bar{L}(f)\) 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.
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 $\tilde{L}_p(p, f)$ and $\tilde{L}_p(f)$ 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 $\tilde{L}_p(f)$. 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 reverbiation: 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.

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>
<thead>
<tr>
<th>Frequency Range</th>
<th>SV Reduction w.r.t. Mono</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–500 Hz</td>
<td>12% 24 Permutations</td>
</tr>
<tr>
<td></td>
<td>12% 16 Permutations</td>
</tr>
<tr>
<td>20–20 kHz</td>
<td>32% 24 Permutations</td>
</tr>
<tr>
<td></td>
<td>43% 16 Permutations</td>
</tr>
</tbody>
</table>
### 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, $T_i = 20$ 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 reverberator 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


Proceedings of the 25th International Conference on Digital Audio Effects (DAFx2022), Vienna, Austria, September 6-10, 2022


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. Computed for 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 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.