Schlecht, Sebastian J.; Alary, Benoit; Välimäki, Vesa; Habets, Emanüel A P

Optimized velvet-noise decorrelator

Published in:
Proceedings of the International Conference on Digital Audio Effects

Published: 04/09/2018

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

Published under the following license:
Unspecified

Please cite the original version:
OPTIMIZED VELVET-NOISE DECORRELATOR

Sebastian J. Schlecht
International Audio Laboratories Erlangen *
Erlangen, Germany
Sebastian.Schlecht@audiolabs-erlangen.de

Benoit Alary†
Acoustics Lab, Dept. of Signal Processing and Acoustics
Aalto University, Espoo, Finland
Benoit.Alary@aalto.fi

Vesa Välimäki
Acoustics Lab, Dept. of Signal Processing and Acoustics
Aalto University, Espoo, Finland
Vesa.Valimaki@aalto.fi

Emanuel A. P. Habets
International Audio Laboratories Erlangen *
Erlangen, Germany
Emanuel.Habets@audiolabs-erlangen.de

ABSTRACT

Decorrelation of audio signals is a critical step for spatial sound reproduction on multichannel configurations. Correlated signals yield a focused phantom source between the reproduction loudspeakers and may produce undesirable comb-filtering artifacts when the signal reaches the listener with small phase differences. Decorrelation techniques reduce such artifacts and extend the spatial auditory image by randomizing the phase of a signal while minimizing the spectral coloration. This paper proposes a method to optimize the decorrelation properties of a sparse noise sequence, called velvet noise, to generate short sparse FIR decorrelation filters. The sparsity allows a highly efficient time-domain convolution. The listening test results demonstrate that the proposed optimization method can yield effective and colorless decorrelation filters. In comparison to a white noise sequence, the filters obtained using the proposed method preserve better the spectrum of a signal and produce good quality broadband decorrelation while using 76% fewer operations for the convolution. Satisfactory results can be achieved with an even lower impulse density which decreases the computational cost by 88%.

1. INTRODUCTION

In multichannel reproduction systems as well as binaural reproduction, the decorrelation of signals is key in controlling the spatial extent of a reproduced sound source. With decorrelation we aim to reduce the cross-correlation of the reproduction signals. For instance when reproducing a mono source on headphones, the spatial image is perceived in the head center. Decorrelation can extend the width of the auditory image such that it appears originating from a larger area. Fully decorrelated signals may even be perceived as separated auditory events [1]. Common applications of decorrelation include controlling the spatial extent, spatial audio coding, sound distance simulation, coloration reduction and headphone externalization [2–5]. This paper focuses on decorrelation methods suitable for controlling the perceived spatial extent of a sound source.

† The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.

† This work was supported by the Academy of Finland (ICHO project, grant no. 296390).

ABSTRACT

Decorrelation of audio signals is a critical step for spatial sound reproduction on multichannel configurations. Correlated signals yield a focused phantom source between the reproduction loudspeakers and may produce undesirable comb-filtering artifacts when the signal reaches the listener with small phase differences. Decorrelation techniques reduce such artifacts and extend the spatial auditory image by randomizing the phase of a signal while minimizing the spectral coloration. This paper proposes a method to optimize the decorrelation properties of a sparse noise sequence, called velvet noise, to generate short sparse FIR decorrelation filters. The sparsity allows a highly efficient time-domain convolution. The listening test results demonstrate that the proposed optimization method can yield effective and colorless decorrelation filters. In comparison to a white noise sequence, the filters obtained using the proposed method preserve better the spectrum of a signal and produce good quality broadband decorrelation while using 76% fewer operations for the convolution. Satisfactory results can be achieved with an even lower impulse density which decreases the computational cost by 88%.

1. INTRODUCTION

In multichannel reproduction systems as well as binaural reproduction, the decorrelation of signals is key in controlling the spatial extent of a reproduced sound source. With decorrelation we aim to reduce the cross-correlation of the reproduction signals. For instance when reproducing a mono source on headphones, the spatial image is perceived in the head center. Decorrelation can extend the width of the auditory image such that it appears originating from a larger area. Fully decorrelated signals may even be perceived as separated auditory events [1]. Common applications of decorrelation include controlling the spatial extent, spatial audio coding, sound distance simulation, coloration reduction and headphone externalization [2–5]. This paper focuses on decorrelation methods suitable for controlling the perceived spatial extent of a sound source.

† The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.

† This work was supported by the Academy of Finland (ICHO project, grant no. 296390).
2. VELVET NOISE

2.1. Velvet-Noise Sequences

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

$$T_d = f_s / N_d,$$

which is called the grid size [12]. The total number of impulses is

$$M = L_d T_d,$$

where $L_d$ is the total length in samples. The sign of each impulse is

$$\sigma(m) = 2 \left\lfloor r_1(m) \right\rfloor - 1,$$

where $\lfloor \cdot \rfloor$ denotes the rounding operation to the closest integer and $0 \leq m < M - 1$ is the integer impulse index, and $r_1(m)$ is a uniformly distributed random number between 0 and 1. The impulse location is

$$\tau(m) = \begin{cases} 0 & \text{for } m = 0, \\ \lfloor T_d (m - 1 + r_2(m)) \rfloor & \text{for } m > 0, \end{cases}$$

where $\lceil \cdot \rceil$ is the ceil operation to the next higher integer and $0 < r_2(m) \leq 1$ is a uniformly distributed random number.

Exponentially decaying impulse gains have been found to im-
prove the sharpness of transients and therefore the quality of the overall decorrelation [20]. The positive gain of each impulse is

$$\gamma(m) = e^{-\tau(m)\alpha},$$

where $\alpha > 0$ denotes the slope of the exponential decay

$$\alpha = -\ln 10^{-L_{dB}/20},$$

where $L_{dB}$ is the target total decay in dB. The exponentially de-
caying velvet noise is denoted $EVN_M$, where $M$ indicates the total number of impulses. In this work, we consider modifications to the $EVN_M$ by allowing deviations from the exponential pulse gains (5) to improve the sequence’s magnitude response. We refer to this non-exponential sequences as optimized velvet noise $OVN_M$ obtained using the method described in Sec. 3.

Since velvet noise is the sum of single delayed impulses, the impulse response $h(n)$ of the resulting sparse FIR filter with $M$ coefficients that are unequal to zero, is given by

$$h(n) = \sum_{m=0}^{M-1} \sigma(m)\gamma(m)\delta(n - \tau(m)),$$

where $\delta$ denotes the Kronecker delta function and $n$ denotes the time index in samples. An input signal $x$ can be decorrelated by convolution with the impulse response $h$. For this, we take advantage of the sparsity of the sequence. By storing the VNS as a series of non-zero elements, all mathematical operations involving zero can be skipped [17, 19]. For a sequence with a density of a 1000 impulses per second, which has been found sufficient for decorre-
lation [20], and a sample rate of 44.1 kHz, the zero elements repres-
t 97.7% of the sequence. Therefore, given a sufficiently sparse sequence, time-domain convolution can be more efficient than a fast convolution using the FFT for an equivalent white-noise se-
quence [20]. Furthermore, this sparse time-domain convolution offers the benefit of being latency-free.

For comparison, we use an exponentially decaying Gaussian white noise sequence $WN$, with the same envelope as given in (5). The spectral coloration, i.e., non-flatness of the magnitude re-
ponse, of the WN is reduced by replacing its magnitude response 
with the same envelope as given in (5). The spectral post-
processing of WN. Convolution with $OVN_M$ according to (7) uses 76% fewer operations than the fast convolution with WN, whereas $OVN_{15}$ decreases the computational cost by 88% [20].

2.2. Velvet Noise in Frequency Domain

In addition to the time-domain formulation given in [20], we for-
mulate the $z$-domain transfer function of the velvet noise. This formulation can be generalized to continuous impulse locations which is critical for the optimization procedure in Sec. 3. The cor-
The solid blue line indicates the exponential decay as defined in (5) with $L_{\text{DB}} = -60 \, \text{dB}$. The shaded blue area indicates the range of the optimized impulse gain with $\pm 6 \, \text{dB}$ and the enforced normalization of the first pulse to $\pm 1$.

The continuous formulation plays a critical role in the optimization of the first pulse to the optimized impulse gain with the optimization process; and iv) the performance results.

### 3. Magnitude Response Optimization

A central challenge in decorrelation is the coloration caused by the magnitude response of the decorrelator. In this work, we employ a third-octave smoothing of the magnitude response in dB between 20 Hz to 20 kHz [21]. The magnitude response is sampled at logarithmically spaced frequencies

$$ f_k = 20 \log_{10} \left( \frac{f_k}{20} \right), $$

where $f_k = [\ln(20), \ldots, \ln(f_s/2)]$ is a linearly spaced $1 \times K$ vector and $K$ is the number of frequency points. The corresponding frequencies in radians are $\omega_k = \frac{2\pi}{T} f_k$. The rectangular smoothing kernel $\kappa$ for a third-octave smoothing is then given by

$$ \kappa(k) = \left\{ \begin{array}{ll} \frac{1}{\pi \omega_k} & \text{for } |k| < \kappa_w \\ 0 & \text{otherwise}, \end{array} \right. $$

where the kernel width $\kappa_w$ is defined by

$$ \kappa_w = \frac{\ln(f_s/2)}{\ln(20)} = \frac{1}{6}. $$

The third-octave smoothed magnitude response $\mathcal{H}$ is then

$$ \mathcal{H}(k) = \left( \kappa * 20 \log |H(e^{i\omega_k / 2})| \right), $$

where $*$ denotes the convolution operation. The objective function $\mathcal{L}$ is given by the root mean squared error (RMSE) of the smoothed magnitude response

$$ \mathcal{L}(\tau, \gamma) = \left( \frac{1}{K} \sum_{k=0}^{K-1} (\mathcal{H}(k) - \overline{\mathcal{H}})^2 \right)^{1/2}, $$

where $\overline{\mathcal{H}} = \sum_{k=0}^{K-1} \mathcal{H}(k)/K$ is the mean smoothed magnitude response. The proposed optimization problem is then

$$ \min_{\tau, \gamma} \mathcal{L}(\tau, \gamma), $$

subject to $\tau(0) = 0$ and $\gamma(0) = 1$

$$ T_A m - 1 < \tau(m) \leq T_A m, $$

$$ e^{-\tau(m)\alpha}/\chi < \gamma(m) < \chi e^{-\tau(m)\alpha}, $$

3.1. Parameter Constraints

In the following, we impose heuristic constraints on the time location $\tau(m)$ and gain $\gamma(m)$ of the impulses of the velvet noise. An even distribution of impulses over time is desirable to ensure a smooth time-domain response [20]. Therefore, the impulse locations should not exceed the boundaries defined in (4).

An impulse with a long delay and a large gain is perceived as an echo, so it degrades the perceptual quality of decorrelated transients. The exponential decay of impulse gains over time as defined in (5) effectively minimizes the time-domain smearing of transients signals [20]. Nonetheless, small deviations from the exponential decay may be marginal for the perception. Informal experiments determined an appropriate range of $\pm 6 \, \text{dB}$ deviation from the exponential decay, which corresponds to a multiplicative gain factor $\chi$ up to 2. To enforce a normalization of the impulse gains, we set the first impulse gain to be $\pm 1$. Later for evaluation purposes, all sequences are normalized to the same energy. Figure 2 depicts the constraints on the impulse gain $\gamma$ over time. The positive and negative impulse gain ranges in Fig. 2 are not connected such that a continuous optimization process cannot change the impulse sign $\sigma$.

3.2. Objective Function

We establish the objective function as to represent the perceived quantity of coloration of the decorrelator. In this work, we employ a third-octave smoothing of the magnitude response in dB between 20 Hz to 20 kHz [21]. The magnitude response is sampled at logarithmically spaced frequencies
where the possible gain deviation $\chi = 2$ and the impulse sign $\sigma$ is a random, but fixed parameter in the objective function $\mathcal{L}$.

### 3.3. Optimization Process

The optimization problem (16) is a constrained, non-linear and non-convex problem such that the optimal solution, i.e., the global minimum, is generally difficult to find. However, local minima can be attained by various gradient descent algorithms. Here we employ a variant of the interior-point method [22]. The initial point is given by a randomly generated EVN according to (4) and (5).

To allow gradual changes of all parameters during optimization, we employ the continuous impulse location $\tilde{\tau}$ in the objective function

$$\min_{\tilde{\tau}, \gamma} \mathcal{L}(\tilde{\tau}, \gamma).$$  \hspace{1cm} (17)

The corresponding integer impulse location solution is then given by $\tau = \lfloor \tilde{\tau} \rfloor$. In the following, we evaluate the error introduced by the continuous impulse location solution.

The continuous impulse location $\tilde{\tau}$ introduces a phase error of the single impulse transfer function in (10). The maximum impulse location error is

$$|\tilde{\tau}(m) - \lfloor \tilde{\tau}(m) \rfloor| \leq 0.5.$$  \hspace{1cm} (18)

Consequently, the maximum phase error between the continuous and the closest integer transfer function is

$$\left| \angle \tilde{H}_m(e^{j\omega}) - \angle H_m(e^{j\omega}) \right| \leq \omega/2$$  \hspace{1cm} (19)

such that the maximum phase error increases linearly with frequency. The phase error of the single impulse transfer function $H_m$ results in a magnitude error of the full sequence transfer function $H$.

Figure 3a depicts the magnitude response error of an EVN$_{30}$ between a continuous impulse location $\tilde{\tau}$ and the closest integer impulse location $\lfloor \tilde{\tau} \rfloor$. The error between the non-smoothed magnitude responses in Fig. 3a increases with frequency up to 20 dB. However, for the third-octave smoothed response in Fig. 3b the error is within 1.3 dB.

Figure 3: Magnitude responses error of an EVN$_{30}$ between a continuous impulse location $\tilde{\tau}$ and the closest integer impulse location $\lfloor \tilde{\tau} \rfloor$. The error between the non-smoothed magnitude responses in Fig. 3a increases with frequency up to 20 dB. However, for the third-octave smoothed response in Fig. 3b the error is within 1.3 dB.

Figure 4: Performance evaluation of the proposed optimization process by comparing 500 sequences of the four decorrelator types: WN, EVN$_{30}$, OVN$_{30}$, and OVN$_{15}$.
In this subsection, we compare the magnitude response of four decorrelator types: WN, EVN, OVN, and OVN. The total length of the sequences is 30 ms and the total decay is $L_{1B} = -60$ dB. We generated 500 sequences for each decorrelation filter type. For the optimized sequence types, the initial sequences are EVN, and OVN, respectively, which were randomly generated. As convergence is not guaranteed, the optimization algorithm was limited to 60 iteration steps to comply with a time limit of 30 s. The mean absolute change in impulse location between the initial point and the local minima is 11 to 12 samples. The mean absolute gain deviation from the exponential decay is about 3 to 4 dB.

Figure 4a depicts the standard deviation of the smoothed magnitude response over 500 sequences. The EVN has the largest standard deviation over all frequencies indicating a relatively poor flatness of the magnitude response. The largest deviation is in the low frequencies with 5.3 dB, which decays with frequency to 1.5 dB. The standard deviation of the WN is similar in shape to the EVN with the largest deviation of 2.3 dB in the low frequencies and a minimum of 0.5 dB in the high frequencies. The standard deviations of the optimized sequences OVN and OVN are similar to WN for high frequencies, but is considerably lower for low frequencies. The minimum standard deviation at around 30 Hz is 1 dB and 1.6 dB, respectively, and by this up to 2.5 times lower than WN and up to 4 times lower than EVN. The low standard deviation of the OVN implies a successful minimization of the objective function (16).

Figure 4b depicts the smoothed magnitude response for the best sequences, i.e., with the lowest objective function value, out of all 500 sequences. The magnitude responses confirm the trends of the standard deviation, as shown in Fig. 4a. The best sequence demonstrates that optimization can yield sequences with less than a 1-dB maximum deviation from the mean magnitude. Despite the large standard deviation in the low frequencies, the best sequences have rather flat magnitude responses at low frequencies.

### 4. SET OF DECORRELATOR SEQUENCES

In many applications, a set of decorrelators is required such that each pair of decorrelation filters is as “different” as possible. In the following, we measure the difference using the coherence and present a method to choose a low-coherence set of multiple decorrelators. When a mono signal is required to be decorrelated to $N_D$ channels, we need $N_D$ decorrelation sequences where each pairwise coherence is minimal.

#### 4.1. Coherence

The effectiveness of decorrelation can be measured with the cross-correlation in different frequency bands, called coherence. Normally, a broadband decorrelator is more effective at higher frequencies than at lower, which is a result of the effective length of a decorrelation filter. Indeed, a longer filter will exhibit stronger decorrelation for longer wavelengths, but will also create potentially perceivable artifacts when the input signal contains transients. To study the decorrelation behavior on a frequency-dependent scale, we use a third-octave filterbank. The signals for the $j$th band are denoted as $a_j$ and $b_j$ and the normalized correlation coefficient as

$$\rho_{a,b}(j) = \frac{\sum_{n} a_j(n)b_j(n)}{\sqrt{\sum_{n} a_j^2(n)} \sum_{n} b_j^2(n)},$$

(20)

where $1 \leq j \leq J$, and $J$ is the number of third-octave bands. Between 20 Hz and 20 kHz, we have $J = 30$. A lower absolute value indicates a more effective decorrelation such that we are mainly interested in the absolute correlation $|\rho_{a,b}(j)|$. To summarize the broadband effectiveness of the decorrelation, we use the frequency mean absolute coherence

$$|\rho_{a,b}| = \frac{1}{Q} \sum_{j=1}^{J} |\rho_{a,b}(j)|.$$

(21)

Note that the sparse impulse locations of two velvet noise sequences rarely coincide such that the classic broadband decorrelation is ill-defined and (21) is preferred instead.

In the following, we evaluate the coherence between the 500 generated sequences of each decorrelation type explained in Sec. 3. Since the coherence is symmetric, there are $500 \times 499/2 = 124,750$ different pairs of sequences. Figure 5a depicts the mean
a selection process. More formally, the goal is to find a set $D$ and $0.22$, the coherence of a set of sequences can be improved by partitioning sequences with low pairwise coherence.

In the next subsection, we present methods to choose a set of decorrelators. The maximum absolute coherence at low frequencies is between 0.35 and 0.4 and the minimum absolute coherence of 0.1 and 0.2. These could be found with the proposed method. The gains $\gamma$ are given with a factor of 10.

Table 1: Best pair of optimized velvet noise $\text{OVN}_{30}$ found with the proposed method. The gains $\gamma$ are given with a factor of 10.

<table>
<thead>
<tr>
<th>$m$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau(a)$</td>
<td>1</td>
<td>46</td>
<td>91</td>
<td>134</td>
<td>175</td>
<td>182</td>
<td>239</td>
<td>271</td>
<td>351</td>
<td>359</td>
<td>407</td>
<td>484</td>
<td>531</td>
<td>536</td>
<td>581</td>
</tr>
<tr>
<td>$\gamma(a)$</td>
<td>4.71</td>
<td>7.37</td>
<td>-3.72</td>
<td>1.46</td>
<td>1.12</td>
<td>-1.84</td>
<td>0.64</td>
<td>-0.54</td>
<td>-0.64</td>
<td>1.08</td>
<td>-0.32</td>
<td>0.24</td>
<td>0.21</td>
<td>-0.49</td>
<td>0.14</td>
</tr>
<tr>
<td>$\tau(b)$</td>
<td>5</td>
<td>78</td>
<td>125</td>
<td>172</td>
<td>219</td>
<td>234</td>
<td>271</td>
<td>318</td>
<td>381</td>
<td>403</td>
<td>460</td>
<td>531</td>
<td>575</td>
<td>583</td>
<td>581</td>
</tr>
<tr>
<td>$\gamma(b)$</td>
<td>4.11</td>
<td>-3.91</td>
<td>5.58</td>
<td>4.30</td>
<td>-2.96</td>
<td>2.02</td>
<td>-0.61</td>
<td>-1.34</td>
<td>1.15</td>
<td>-0.93</td>
<td>0.81</td>
<td>-0.37</td>
<td>-0.26</td>
<td>0.16</td>
<td>0.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m$</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau(a)$</td>
<td>651</td>
<td>669</td>
<td>731</td>
<td>797</td>
<td>829</td>
<td>851</td>
<td>890</td>
<td>961</td>
<td>984</td>
<td>1027</td>
<td>1074</td>
<td>1130</td>
<td>1175</td>
<td>1232</td>
<td>1246</td>
</tr>
<tr>
<td>$\gamma(a)$</td>
<td>0.18</td>
<td>-0.14</td>
<td>-0.09</td>
<td>-0.08</td>
<td>-0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\tau(b)$</td>
<td>663</td>
<td>703</td>
<td>737</td>
<td>791</td>
<td>809</td>
<td>881</td>
<td>902</td>
<td>950</td>
<td>999</td>
<td>1041</td>
<td>1083</td>
<td>1135</td>
<td>1177</td>
<td>1216</td>
<td>1258</td>
</tr>
<tr>
<td>$\gamma(b)$</td>
<td>0.10</td>
<td>-0.19</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>-0.06</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Table 2: Best pair of optimized velvet noise $\text{OVN}_{15}$ found with the proposed method. The gains $\gamma$ are given with a factor of 10.

<table>
<thead>
<tr>
<th>$m$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau(a)$</td>
<td>1</td>
<td>51</td>
<td>101</td>
<td>200</td>
<td>291</td>
<td>372</td>
<td>476</td>
<td>581</td>
<td>627</td>
<td>736</td>
<td>827</td>
<td>913</td>
<td>998</td>
<td>1089</td>
<td>1180</td>
</tr>
<tr>
<td>$\gamma(a)$</td>
<td>4.80</td>
<td>-7.51</td>
<td>-4.18</td>
<td>-1.58</td>
<td>-0.48</td>
<td>0.29</td>
<td>0.21</td>
<td>0.43</td>
<td>-0.08</td>
<td>0.20</td>
<td>0.12</td>
<td>0.08</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\tau(b)$</td>
<td>1</td>
<td>10</td>
<td>140</td>
<td>215</td>
<td>279</td>
<td>365</td>
<td>485</td>
<td>579</td>
<td>668</td>
<td>756</td>
<td>836</td>
<td>892</td>
<td>1005</td>
<td>1071</td>
<td>1192</td>
</tr>
<tr>
<td>$\gamma(b)$</td>
<td>6.10</td>
<td>-2.94</td>
<td>6.63</td>
<td>-1.05</td>
<td>-2.88</td>
<td>-0.46</td>
<td>-0.28</td>
<td>-0.68</td>
<td>-0.36</td>
<td>0.06</td>
<td>0.04</td>
<td>-0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

The gains $\gamma$ are given with a factor of 10. The absolute coherence for each third-octave band over all sequence pairs. For all four decorrelator types, the absolute coherence decreases with frequency due to the effective length of the decorrelator. The maximum absolute coherence at low frequencies is between 0.35 and 0.4 and the minimum absolute coherence of 0.1 and 0.22. These could be found with the proposed method. The gains $\gamma$ are given with a factor of 10. The absolute coherence is optimal if the distributions of $\lambda(a,b)$ and $\lambda(b,a)$ overlap. The balance $\gamma$ is the normalization factor to balance the two objective functions with $\lambda = 0.5$. The balance is optimal if the distributions of $[\lambda(a,b)]$ and $\mu(L_a + L_b)$ overlap maximally. In this work, this is achieved by $\mu = 0.1$. The larger $\gamma$, the more emphasis is put on magnitude flatness rather than a low coherence value. Tables 1 and 2 give the best decorrelation pairs we have found through our proposed method with $\lambda = 0.8$. These sequences were evaluated via a formal listening test, as explained in the next section.

### 4.2. Choosing Set of Decorrelators

Although the mean absolute coherence is typically between 0.19 and 0.22, the coherence of a set of sequences can be improved by selecting them. More formally, the goal is to find a set $D$ of $N_D$ sequences such that

$$
\min_D \sum_{a,b \in D} |\rho_{a,b}|.
$$

Let us consider the coherence matrix, i.e., all pairwise frequency mean absolute coherences, to be the adjacency matrix of an undirected graph. The minimization problem (22) then corresponds to finding the thinnest $N_D$-subgraph. By taking the negative of the coherence matrix, this problem is equivalent to the better known densest $N_D$-subgraph problem [23]. Although finding the optimal solution is NP-hard, greedy algorithms can be applied to yield an approximative solution [24]. In this contribution, however, we are mainly concerned with pairs of sequences to allow decorrelated stereo reproduction. Thus, (22) is merely the minimum entry of the coherence matrix. Although, the frequency mean absolute coherence peaks around 0.2 in Fig. 5b, sequence pairs with coherence as low as 0.05 can be found for all decorrelator types.

### 4.3. Penalty Term

In the choice of the optimal set of decorrelators, the lowest coherence pairs are not necessarily those which have flat magnitude responses. To account for the coloration of the single sequences, we introduce a penalty term for (22):

$$
\min_D \sum_{a,b \in D} (1 - \lambda)|\rho_{a,b}| + \lambda \mu(\mathcal{L}_a + \mathcal{L}_b),
$$

where $\mathcal{L}_a$ and $\mathcal{L}_b$ are the objective functions (15) of sequences $a$ and $b$, $\lambda$ is the weighting factor, and $\mu$ is the normalization factor to balance the two objective functions with $\lambda = 0.5$. The balance is optimal if the distributions of $[\rho_{a,b}]$ and $\mu(L_a + L_b)$ overlap maximally. In this work, this is achieved by $\mu = 0.1$. The larger $\gamma$, the more emphasis is put on magnitude flatness rather than a low coherence value. Tables 1 and 2 give the best decorrelation pairs we have found through our proposed method with $\lambda = 0.8$. These sequences were evaluated via a formal listening test, as explained in the next section.

### 5. PERCEPTUAL EVALUATION

We conducted two formal listening tests to evaluate the perceived quality of the decorrelation filters obtained using the proposed method. The first test assessed the coloration introduced by the decorrelators via comparison of the processed signal to the unprocessed signal. The second test evaluated the effectiveness of the decorrelators to extend the auditory source width and overall quality. The tests were conducted in special listening booths built for sound isolation and high-quality reproduction over headphones. The test interface was based on a MUSHRA-type web interface with a subjective rating scale from 0 to 100 allowing seamless
switching between test conditions and looping of short sections. Each test page compared six conditions: OVN$_{30}$, OVN$_{15}$, EVN$_{30}$, WN, anchor, and reference. For each decorrelation type, we chose four decorrelator instances. Each test page was repeated once during the test. In total, 4 instances × 2 trials × 4 input signals = 32 test pages were presented for each test$^1$.

Each listening test was participated by 11 listeners (10 males and 1 female) who were all aged between 24 and 34. Due to the long test time, few participants performed both tests on the same day. Four different input signals were convolved with the decorrelation sequences: drums, guitar, singing, and speech. The order of the test conditions was individually randomized. From the difference between the identical trials, the test-retest reliability could be computed. The cross-correlation coefficient between the first and second trial was 0.96 suggesting that most participants were able to give consistent ratings.

5.1. Coloration Test

The first listening test evaluated how much the decorrelation filters distort the input signal. The input signal was convolved with a single decorrelation filter, and the difference to the unprocessed signal was rated by the participants. In MUSHRA terminology, the unprocessed mono signal was the reference, and the input signal processed with a lowpass filter having a 3.5-kHz cutoff frequency was the anchor. The resulting mono signals were reproduced on both headphone channels. The main coloration was expected to be caused by the change in timbre and smearing of transients.

The four decorrelation instances were selected out of the 500 sequences which were generated in Sec. 3. For OVN$_{30}$ and OVN$_{15}$, we selected the four best sequences according to spectral flatness as defined in (15). The EVN$_{30}$ sequences were selected as the initial sequences of the OVN$_{30}$, i.e., the original random sequence before the proposed optimization process to emphasize the improvement gained by the proposed. The WN sequences were generated randomly and spectrally flattened, as described in Sec. 2.

$^1$Audio examples are available at https://www.audiolabs-erlangen.de/resources/2018-DAFx-VND.

Figure 6a shows the resulting subjective rating of the coloration test. The median ratings for OVN$_{30}$, OVN$_{15}$, EVN$_{30}$, and WN are 90, 86, 26, and 75, respectively. All pairwise comparisons of the confidence interval suggests that the medians are significantly different at the 95% confidence level. The superior rating of both optimized velvet-noise sequences suggests a substantial reduction in spectral coloration compared to EVN$_{30}$, and this demonstrates the effectiveness of the optimization method and the corresponding objective function (15). Furthermore, both OVN$_{30}$ and OVN$_{15}$ were rated slightly superior to WN suggesting that they are valid alternatives.

5.2. Stereo Quality Test

The second listening test evaluated the effectiveness of the decorrelators in extending the auditory source width and the overall spatial quality. The input signal was convolved with a decorrelation filter for each channel (left and right) and the participants were asked to rate the perceived width, localization at the center, and overall quality. In this test, no ideal reference could be defined, so the unprocessed mono signal was provided only for guidance. The lowpass filtered mono signal was given as the anchor. The resulting stereo signal was reproduced on the left and right headphone channels. Once again, we selected the sequences from the generated set as in the coloration test. For OVN$_{30}$ and OVN$_{15}$, we selected the four best sequence pairs according to the rating function (23) and weighting factor $\lambda = 0.8$. Tables 1 and 2 present the top-rated sequence pairs. The EVN$_{30}$ sequence pairs were selected as the initial optimization sequences of the OVN$_{30}$ pairs. The WN sequence pairs were generated randomly according to Sec. 2.

Figure 6b shows the resulting subjective rating of the auditory source width test. The median ratings for OVN$_{30}$, OVN$_{15}$, EVN$_{30}$, and WN are 72, 71, 32, and 80, respectively. Pairwise comparison of the confidence interval suggests that the EVN$_{30}$ and WN medians are significantly different at the 95% confidence level. No significant difference between OVN$_{30}$ and OVN$_{15}$ was found. Here again, a superior rating was given to the optimized sequences over the EVN$_{30}$, which is expected due to the perceptible coloration of the EVN$_{30}$ found in the coloration test. A slightly inferior rating
was given to the optimized methods compared to WN. This may be a result of our pair selection process favoring a flat spectrum over low coherence. Nonetheless, these results suggest that OVN\textsubscript{30} and OVN\textsubscript{15} are valid alternatives to WN, since they can yield reduction in the computational cost without affecting significantly the overall sound quality.

6. CONCLUSION

We have proposed an optimization method to improve the perceived quality of velvet-noise decorrelators. The original method EVN employed short, sparse, and exponentially decaying sequences, which were generated randomly [20]. The proposed method OVN attempts to improve such sequences by allowing small deviations in the impulse gains and timings. The optimization minimizes the spectral flatness within given heuristic constraints. A continuous impulse location formulation facilitates simultaneous modifications of gains and times. Furthermore, we proposed a method to select a set of minimally correlated sequences according to a coherence metric. An additional weighting factor allows user-defined control over the trade-off between coherence and spectral flatness.

Two formal listening tests were conducted to evaluate possible coloration as well as the auditory source width and overall stereo quality. The subjective ratings show a substantial improvement of the proposed method against the original and perceptually satisfactory decorrelation. While convolving a signal with velvet noise can be performed using as much as 88% less operations compared with WN, the objective ratings as well as the subjective ones confirms that the proposed OVN method is a good alternative to the WN decorrelation, when it is possible to pre-compute sets of optimal sequences.

7. ACKNOWLEDGMENT

Part of this research was conducted in March 2018, when Dr. Sebastian Schlecht visited the Aalto Acoustics Lab for one week.

8. REFERENCES