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

Emanuël 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 center of the head. 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 separate 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.

Decorrelation may be achieved by randomizing the phase of a signal while maintaining its magnitude spectrum. In [2], Kendall proposed a decorrelation filter based on 20–30 ms sequences of white noise. Shorter decorrelation filters can preserve the quality of the transients and prevent a reverberation effect [2]. Indeed, since high frequencies have shorter wavelengths, randomizing their phases can produce a noticeable smearing effects on short transient signals if the delays are too long. Unfortunately, limiting the length of a filter will limit its ability to decorrelate low frequencies, since long wavelengths require long delays to alter their phase significantly. This duality illustrates the challenge of designing a good broadband decorrelator that can compromise between preserving the transients and low-frequency decorrelation. This is the reason why most modern decorrelation methods operate in the time-frequency domain and restrict the phase variation based on the wavelength of various frequency bands [6].

Laitinen et al. proposed to apply a random delay within perceptually motivated bounds at each frequency band [7]. Although this method can lead to audible artifacts in stereo reproduction, these artifacts are less perceivable in multichannel reproduction. An alternative and common method is to decompose the signal into transient and non-transient signals, and apply the decorrelation only to the non-transient signal. For time-domain methods, finite impulse response (FIR) filters are applied with the fast convolution technique which can be computationally prohibitive for long filters in multiple decorrelation stages of multichannel systems. Alternatively, infinite impulse response (IIR) filters such as single or cascaded allpass filters, which guarantee a flat magnitude response, are computationally efficient [2, 8, 9]. However, if the group delay of the filter becomes too large, higher-order allpass filters can cause an undesired chirping effect [10].

Karjalainen and Järveläinen proposed velvet-noise sequences (VNSes), i.e., sparse series of uniformly distributed ± 1s, as a perceptually smoother alternative to Gaussian white noise [11, 12]. At only a fraction of the computational cost of dense FIR filters, VNSes are suitable for artificial reverberation [13,14] and approximation of room impulse responses [11,15–19]. Short VNSes were proposed as an effective decorrelation method, although it suffered from spectral coloration [20]. In this work, we present a method to optimize the decorrelation properties of VNSes without altering the computational cost. We also conduct a formal listening test to evaluate the new method and to compare it with previous methods.

This paper is organized as follows. In Sec. 2, we review vel-

* 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).

vet noise and its time and frequency-domain representations. Section 3 proposes an optimization technique for VNSes to minimize spectral coloration. Section 4 proposes a selection process to improve the decorrelation in sets of sequences. Section 5 presents the listening tests we conducted to evaluate the proposed method.

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

$$T_\mathrm{d} = f_\mathrm{s}/N_\mathrm{d}, \tag{1}$$

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

$$M = L_\mathrm{s} T_\mathrm{d}, \tag{2}$$

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

$$\sigma(m) = 2\,\lfloor r_1(m) \rceil - 1, \tag{3}$$

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

$$\tau(m) = \begin{cases} 0 & \text{for } m = 0 \\ \lceil T_\mathrm{d}(m - 1 + r_2(m)) \rceil & \text{for } m > 0, \end{cases} \tag{4}$$

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

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

$$\gamma(m) = e^{-\tau(m)\alpha}, \tag{5}$$

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

$$\alpha = \frac{-\ln 10^{-L_\mathrm{dB}/20}}{L_\mathrm{s}}, \tag{6}$$

where L_dB is the target total decay in dB. The exponentially decaying 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

$$h(n) = \sum_{m=0}^{M-1} \sigma(m)\gamma(m)\delta(n - \tau(m)), \tag{7}$$

where δ 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 decorrelation [20], and a sample rate of 44.1 kHz, the zero elements represent 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 sequence [20]. Furthermore, this sparse time-domain convolution offers the benefit of being latency-free.

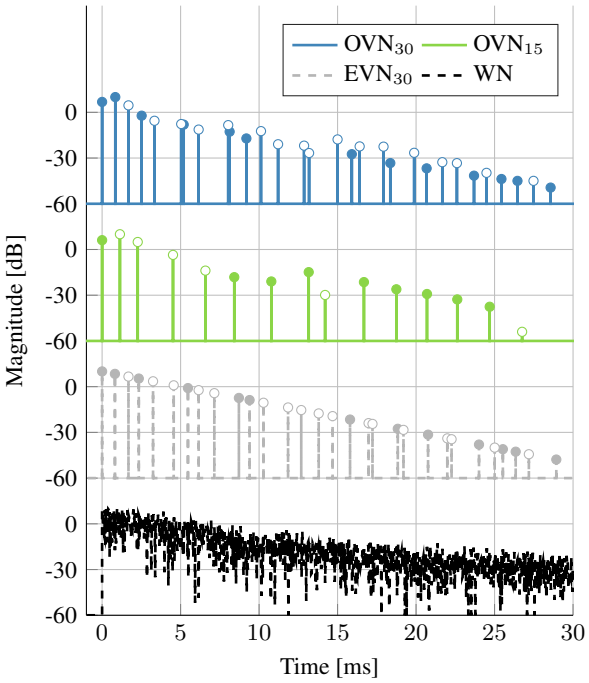


Figure 1: *Decorrelator sequences in the time domain: white noise WN, exponential velvet noise EVN_{30}, and two optimized velvet-noise sequences OVN_{15} and OVN_{30}. Positive impulses are indicated by • and negative gains by ∘ (except for WN).*

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 response, of the WN is reduced by replacing its magnitude response with a constant number, and re-synthesizing the time-domain sequence using the inverse Fourier transform.

Figure 1 depicts four decorrelation sequences: OVN_{30}, OVN_{15}, EVN_{30}, and WN. The total length of each sequence is 30 ms such that the VNS sequences have an impulse density of 1 and 0.5 impulse per ms, respectively. The total decay is $L_\mathrm{dB} = -60$ dB. However, the impulse response of WN decays only by about -30 dB in total, because of the spectral post-processing of WN. Convolution with OVN_{30} 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 formulate 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-

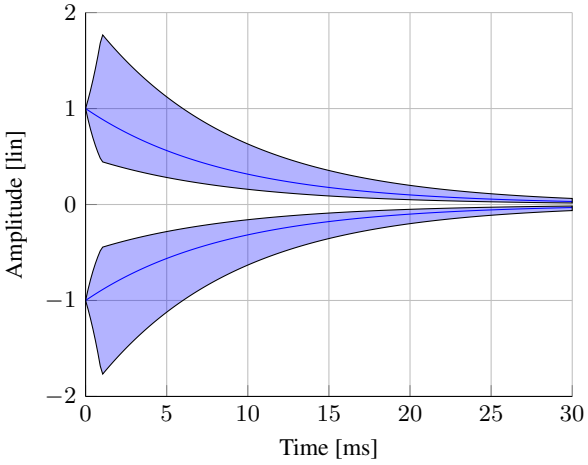


Figure 2: *Constraint on the optimized impulse gain γ over time. The solid blue line indicates the exponential decay as defined in* (5) *with $L_{\mathrm{dB}} = -60$ dB. The shaded blue area indicates the range of the optimized impulse gain with ±6 dB and the enforced normalization of the first pulse to* ± 1.

responding z-domain transfer function of (7) is

$$H(z) = \sum_{m=0}^{M-1} \sigma(m)\gamma(m) z^{-\tau(m)} = \sum_{m=0}^{M-1} H_m(z), \tag{8}$$

where $H_m(z)$ indicates the transfer function of the mth impulse. The magnitude response of the mth impulse is

$$|H_m(e^{\imath\omega})| = \gamma(m), \tag{9}$$

where ω is the frequency in radians and $\imath = \sqrt{-1}$. The corresponding unwrapped phase response is

$$\angle H_m(e^{\imath\omega}) = \begin{cases} -\omega\tau(m) & \text{for } \sigma(m) = 1 \\ \pi - \omega\tau(m) & \text{for } \sigma(m) = -1, \end{cases} \tag{10}$$

where $\angle$ denotes the radian angle of a complex number. The phase response formulation in (10) generalizes directly to continuous impulse locations $\widetilde{\tau}(m)$. The corresponding single impulse and summed transfer functions are denoted $\widetilde{H}_m$ and $\widetilde{H}$, respectively. The continuous formulation plays a critical role in the optimization process presented in the following section as it allows continuous modification of both impulse location and impulse gain.

3. MAGNITUDE RESPONSE OPTIMIZATION

A central challenge in decorrelation is the coloration caused by a non-flat magnitude response of the decorrelator. This section is concerned with modifying the impulse locations $\tau(m)$ and impulse gains $\gamma(m)$ of a VNS to improve the flatness of its magnitude response $|H(e^{\imath\omega})|$. In the following subsections, we describe: i) heuristic constraints on the velvet-noise parameters; ii) the objective function; iii) the optimization process; and iv) the performance results.

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 ±6 dB deviation from the exponential decay, which corresponds to a multiplicative gain factor χ up to 2. To enforce a normalization of the impulse gains, we set the first impulse gain to be ± 1. Later for evaluation purposes, all sequences are normalized to the same energy. Figure 2 depicts the constraints on the impulse gain γ 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 σ.

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

$$\boldsymbol{f}_{\log}(k) = e^{\boldsymbol{f}_{\mathrm{lin}}(k)}, \tag{11}$$

where $\boldsymbol{f}_{\mathrm{lin}} = [\ln(20), \ldots, \ln(f_{\mathrm{s}}/2)]$ is a linearly spaced $1 \times K$ vector and K is the number of frequency points. The corresponding frequencies in radian are $\boldsymbol{\omega}_{\log} = \frac{2\pi}{f_{\mathrm{s}}} \boldsymbol{f}_{\log}$. The rectangular smoothing kernel $\boldsymbol{\kappa}$ for a third-octave smoothing is then given by

$$\boldsymbol{\kappa}(k) = \begin{cases} \frac{1}{2\kappa_w + 1} & \text{for } |k| < \kappa_w \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where the kernel width κ_w is defined by

$$\frac{\kappa_w}{K} \frac{\ln(f_{\mathrm{s}}/2)}{\ln(20)} = \frac{1}{6}. \tag{13}$$

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

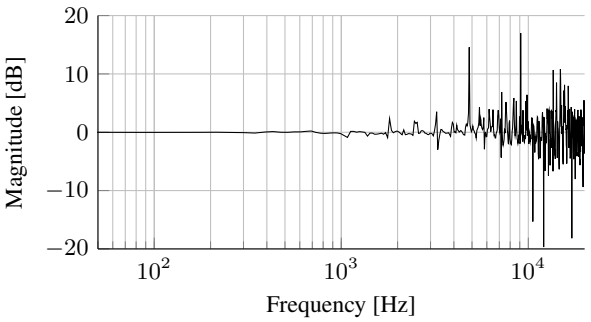
$$\mathcal{H}(k) = \left(\boldsymbol{\kappa} * 20 \log \left| H\left(e^{\imath \boldsymbol{\omega}_{\log}(k)}\right)\right|\right), \tag{14}$$

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

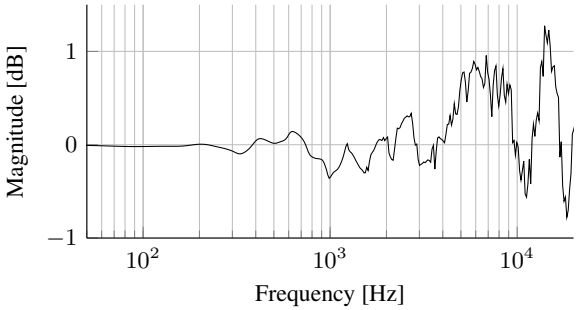
$$\mathcal{L}(\tau, \gamma) = \sqrt{\frac{1}{K} \sum_{k=0}^{K-1} \left(\mathcal{H}(k) - \overline{\mathcal{H}}\right)^2}, \tag{15}$$

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

$$\begin{aligned} &\min_{\tau,\gamma} \mathcal{L}(\tau, \gamma) \\ &\text{subject to } \tau(0) = 0 \text{ and } \gamma(0) = 1 \\ &\qquad T_{\mathrm{d}}(m-1) < \tau(m) \leq T_{\mathrm{d}}\, m \\ &\qquad e^{-\tau(m)\alpha}/\chi \leq \gamma(m) < \chi e^{-\tau(m)\alpha}, \end{aligned} \tag{16}$$

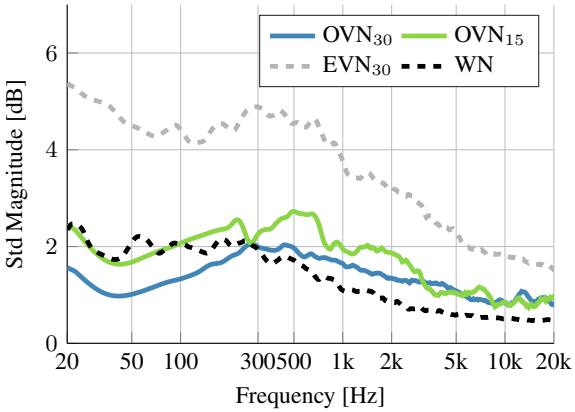


(a) Magnitude response error without smoothing.

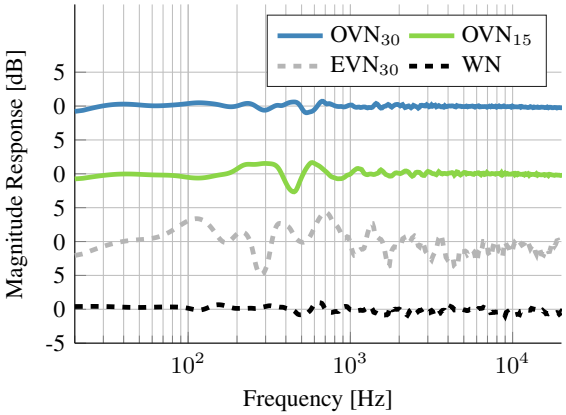


(b) Magnitude response error with third-octave smoothing.

Figure 3: *Magnitude responses error of an EVN$_{30}$ between a continuous impulse location $\widetilde{\tau}$ and the closest integer impulse location $\lfloor\widetilde{\tau}\rceil$. 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.*



(a) Standard deviation on the smoothed magnitude response for 500 sequences.



(b) Smoothed magnitude response of the best sequence, i.e., with the lowest objective function value, out of 500 sequences.

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$_{15}$.*

where the possible gain deviation $\chi = 2$ and the impulse sign σ 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 $\widetilde{\tau}$ in the objective function

$$\min_{\widetilde{\tau},\gamma} \mathcal{L}(\widetilde{\tau},\gamma). \tag{17}$$

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

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

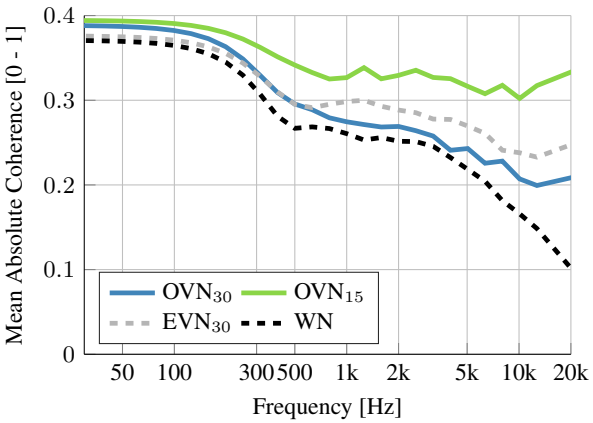
$$|\widetilde{\tau}(m) - \lfloor\widetilde{\tau}(m)\rceil| \leq 0.5. \tag{18}$$

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

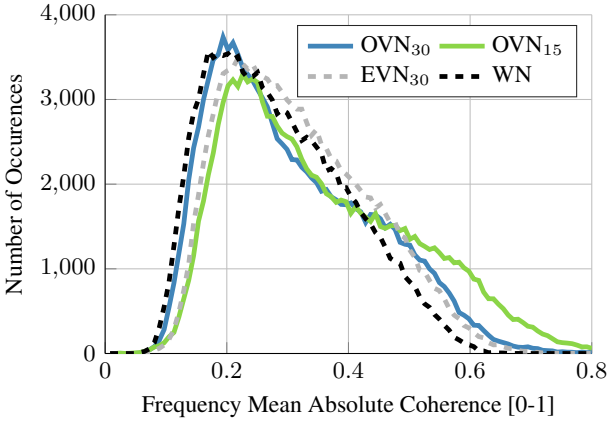
$$\left|\angle\widetilde{H}_m(e^{\imath\omega}) - \angle H_m(e^{\imath\omega})\right| \leq \omega/2 \tag{19}$$

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 $\widetilde{\tau}$ and the closest integer impulse location $\lfloor\widetilde{\tau}\rceil$. Whereas the magnitude error is below 1 dB for frequencies below 1 kHz, the error increases up to 20 dB for high frequencies. In Fig. 3b, the magnitude response error of the same two sequences are shown with third-octave smoothing. The maximum error over the complete frequency range stays below 1.3 dB. Similarly, Karjalainen and Järveläinen observed that increasing the time resolution beyond 44.1 kHz, does not improve velvet noise [11]. Hence, the proposed optimization using continuous impulse locations which are then rounded to the nearest integers introduces only minor deviations in the magnitude response.



(a) Mean absolute coherence.



(b) Distribution of frequency mean absolute coherence.

Figure 5: *Evaluation of the absolute coherence between over all sequence pairs of the 500 randomly generated sequences of four decorrelator types: WN, EVN$_{30}$, OVN$_{30}$, and OVN$_{15}$.*

3.4. Results

In this subsection, we compare the magnitude response of four decorrelation sequence types: WN, EVN$_{30}$, OVN$_{30}$ and OVN$_{15}$. The total length of the sequences is 30 ms and the total decay is $L_{\mathrm{dB}} = -60$ dB. We generated 500 sequences for each decorrelation filter type. For the optimized sequence types, the initial sequences are EVN$_{15}$ and EVN$_{30}$, 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$_{30}$ 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$_{30}$ 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$_{30}$ and OVN$_{15}$ 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$_{30}$. The low standard deviation of the OVN$_{30}$ 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 $N_{\mathcal{D}}$ channels, we need $N_{\mathcal{D}}$ 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 jth band are denoted as a_j and b_j and the normalized correlation coefficient as

$$\rho_{a,b}^{(j)} = \frac{\sum_n a_j(n) b_j(n)}{\sqrt{\sum_n a_j^2(n) \sum_n b_j^2(n)}}, \tag{20}$$

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 $\left|\rho_{a,b}^{(j)}\right|$. To summarize the broadband effectiveness of the decorrelation, we use the frequency mean absolute coherence

$$\overline{|\rho_{a,b}|} = \frac{1}{Q} \sum_{j=1}^{J} \left|\rho_{a,b}^{(j)}\right|. \tag{21}$$

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 absolute coherence for each third-octave band over all sequence

Table 1: *Best pair of optimized velvet noise OVN_{30} found with the proposed method. The gains γ are given with a factor of 10.*

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\tau_a(m)$	1	46	91	134	175	182	239	271	351	359	407	484	531	536	581
$\gamma_a(m)$	4.71	7.37	-3.72	1.46	1.12	-1.84	0.64	-0.54	-0.64	1.08	-0.32	0.24	0.21	-0.49	0.14
$\tau_b(m)$	1	5	78	125	172	219	234	271	318	381	403	460	531	575	583
$\gamma_b(m)$	4.11	-3.91	5.58	4.30	-2.96	2.02	-0.61	-1.34	1.15	-0.93	0.81	-0.37	-0.26	0.16	0.14
m	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
$\tau_a(m)$	651	669	731	797	829	851	890	961	984	1027	1074	1130	1175	1232	1246
$\gamma_a(m)$	0.18	-0.14	-0.09	-0.08	-0.08	0.07	0.05	0.04	-0.04	0.02	0.02	0.01	-0.01	0.01	-0.01
$\tau_b(m)$	663	703	737	791	809	881	902	950	999	1041	1083	1135	1177	1216	1258
$\gamma_b(m)$	0.10	-0.19	0.07	0.06	0.05	0.05	-0.06	-0.04	0.03	0.02	-0.02	0.01	-0.01	-0.01	-0.01

Table 2: *Best pair of optimized velvet noise OVN_{15} found with the proposed method. The gains γ are given with a factor of 10.*

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

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 0.33 at high frequencies. The coherence is generally slightly larger for EVN_{30} and OVN_{15} due to the systematic exponential gain, and higher sparsity, respectively. Since coherence is not modeled in the optimization process in Sec. 3, it is expected to have little influence on the overall coherence.

Figure 5b depicts the distribution of the frequency mean absolute coherence $\overline{|\rho_{a,b}|}$ over all pairs. The difference between the four decorrelation types is small, as expected, and a frequency mean absolute coherence of around 0.19 to 0.22 is most frequent. However, there are sequence pairs with rather large coherence values up to 0.8 suggesting poor decorrelation performance. In the next subsection, we present methods to choose a set of decorrelation sequences with low pairwise coherence.

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 a selection process. More formally, the goal is to find a set $\mathcal{D}$ of $N_{\mathcal{D}}$ sequences such that

$$\min_{\mathcal{D}} \sum_{a,b\in\mathcal{D}} \overline{|\rho_{a,b}|}. \tag{22}$$

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* $N_{\mathcal{D}}$-subgraph. By taking the negative of the coherence matrix, this problem is equivalent to the better known *densest* $N_{\mathcal{D}}$-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.

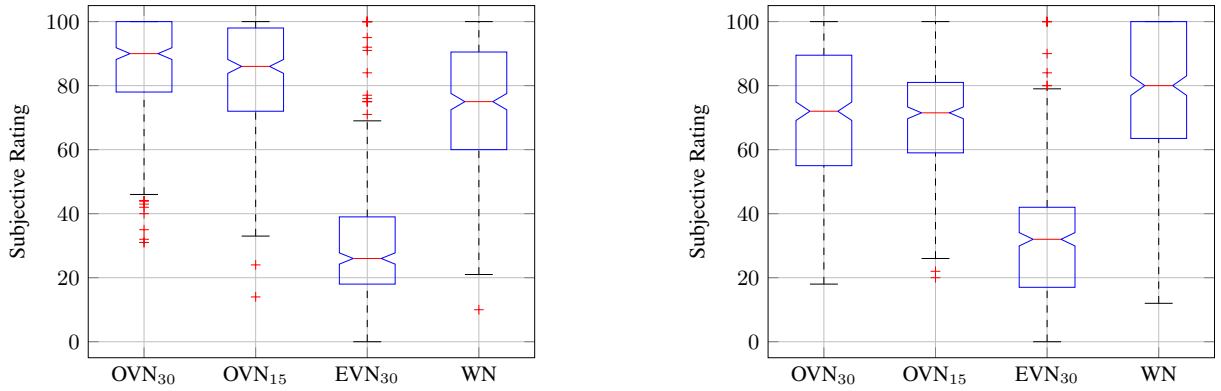
In the choice of the optimal set of decorrelators, the lowest coherence pairs are not necessarily those which have flat magnitude responses. To account for the coloration of the single sequences, we introduce a penalty term for (22):

$$\min_{\mathcal{D}} \sum_{a,b\in\mathcal{D}} (1-\lambda)\overline{|\rho_{a,b}|} + \lambda\,\mu(\mathcal{L}_a + \mathcal{L}_b), \tag{23}$$

where $\mathcal{L}_a$ and $\mathcal{L}_b$ are the objective functions (15) of sequences a and b, λ is the weighting factor, and μ is the normalization factor to balance the two objective functions with $\lambda = 0.5$. The balance is optimal if the distributions of $\overline{|\rho_{a,b}|}$ and $\mu(\mathcal{L}_a + \mathcal{L}_b)$ overlap maximally. In this work, this is achieved by $\mu = 0.1$. The larger λ, 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.



(a) Coloration test with mono reproduction.

(b) Stereo quality test with stereo reproduction.

Figure 6: *Results of two listening tests of four decorrelator types: OVN_{30}, OVN_{15}, EVN_{30}, and WN. In each box, the central red line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the + symbol. The box notches indicate the confidence intervals, i.e., two medians are significantly different at the 95% confidence level if their intervals do not overlap.*

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 $\times$ 2 trials $\times$ 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 optimization to emphasize the improvement gained by the proposed method. The WN sequences were generated randomly and spectrally flattened, as described in Sec. 2.

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

[1] Audio examples are available at `https://www.audiolabs-erlangen.de/resources/2018-DAFx-VND`.

low coherence. Nonetheless, these results suggest that OVN_{30} and OVN_{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 maximizes 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

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT press, 1997.

[2] 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.

[3] G. Potard and I. Burnett, "Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays," in *Proc. DAFX-04*, Naples, Italy, Oct. 2004, pp. 280–284.

[4] C. Faller, "Parametric multichannel audio coding: synthesis of coherence cues," *IEEE Trans. Audio, Speech, and Lang. Processing*, vol. 14, no. 1, pp. 299–310, Jan. 2006.

[5] 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.

[6] M. Bouéri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Proc. AES 117th Conv.*, San Francisco, CA, USA, Oct. 2004.

[7] M. Laitinen, F. Kuech, S. Disch, and V. Pulkki, "Reproducing applause-type signals with directional audio coding," *J. Audio Eng. Soc.*, vol. 59, no. 1/2, pp. 29–43, Jan./Feb. 2011.

[8] E. Kermit-Canfield and J. Abel, "Signal decorrelation using perceptually informed allpass filters," in *Proc. DAFx-16*, Brno, Czech Republic, Sept. 2016, pp. 225–231.

[9] E. K. Canfield-Dafilou and J. S. Abel, "A group delay-based method for signal decorrelation," in *Proc. AES 144th Conv.*, Milan, Italy, May 2018.

[10] V. Välimäki, J. S. Abel, and J. O. Smith, "Spectral delay filters," *J. Audio Eng. Soc.*, vol. 57, no. 7/8, pp. 521–531, Jul./Aug. 2009.

[11] M. Karjalainen and H. Järveläinen, "Reverberation modeling using velvet noise," in *Proc. AES 30th Int. Conf.: Intelligent Audio Environments*, Saariselkä, Finland, Mar. 2007.

[12] 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, and Lang. Processing*, vol. 21, no. 7, pp. 1481–1488, July 2013.

[13] P. Rubak and L. G. Johansen, "Artificial reverberation based on a pseudo-random impulse response," in *Proc. AES 104th Conv.*, Amsterdam, The Netherlands, May 1998.

[14] P. Rubak and L. G. Johansen, "Artificial reverberation based on a pseudo-random impulse response II," in *Proc. AES 106th Conv.*, Munich, Germany, May 1999.

[15] J. Vilkamo, B. Neugebauer, and J. Plogsties, "Sparse frequency-domain reverberator," *J. Audio Eng. Soc.*, vol. 59, no. 12, pp. 936–943, Dec. 2012.

[16] S. Oksanen, J. Parker, A. Politis, and V. Välimäki, "A directional diffuse reverberation model for excavated tunnels in rock," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 644–648.

[17] B. Holm-Rasmussen, H.-M. Lehtonen, and V. Välimäki, "A new reverberator based on variable sparsity convolution," in *Proc. DAFx-13*, Maynooth, Ireland, Sept. 2013, pp. 344–350.

[18] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "More than 50 years of artificial reverberation," in *Proc. AES 60th Int. Conf.*, Leuven, Belgium, Feb. 2016.

[19] V. Välimäki, B. Holm-Rasmussen, B. Alary, and H.-M. Lehtonen, "Late reverberation synthesis using filtered velvet noise," *Appl. Sci.*, vol. 7, no. 483, May 2017.

[20] B. Alary, A. Politis, and V. Välimäki, "Velvet-noise decorrelator," in *Proc. DAFx-17*, Edinburgh, UK, Sept. 2017, pp. 405–411.

[21] F. E. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, Focal Press, Burlington, MA, USA, 2008.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering. Springer Science & Business Media, New York, NY, USA, Jan. 1999.

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

[24] M. Charikar, "Greedy approximation algorithms for finding dense components in a graph," in *Proc. Int. Work. Approx. Algor. Comb. Optim.*, Berlin, Germany, 2000, pp. 84–95.