forum acusticum 2026
8-12 September · Graz, Austria

EFFICIENT DIFFUSE FIELD DECORRELATION: SPATIALIZED VELVET-NOISE

Leon Kaiser[1], Matthias Frank[1], Lukas Gölles[1], Franz Zotter[1]

[1] *Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria*

Abstract

Rendering diffuse sound fields of late reverberation with computational efficiency and accurate spatial impression is a challenging task. This paper introduces spatialized velvet-noise, a framework that efficiently generates a decorrelated sound field from a single monaural reverberation in real time. The method utilizes a single velvet-noise sequence that is distributed spatially. The approach is compared to other computationally efficient decorrelation strategies, including allpass filters, simple delays, and alternative velvet-noise configurations. This comparison involves a technical analysis of spectral deviations and interaural coherence, alongside a formal listening experiment. Results reveal that spatialized velvet-noise significantly outperforms most of the tested configurations in providing an enveloping sensation or maintaining a neutral timbre. Notably, none of the tested method was found to surpass spatialized velvet-noise in the listening experiment. While simple delays are the only computationally more efficient option, spatialized velvet-noise significantly outperforms them in the listening test.

Keywords: *velvet-noise, diffuse field, decorrelation, spatial audio*

1. Introduction

Once a high density of reflections is reached and the mixing time [1] is exceeded, a room impulse response (RIR) can be considered a diffuse statistical decay. The resulting sensation of listener envelopment (LEV) remains a primary focus of spatial audio research [2], [3], [4]. To reproduce these diffuse sound fields, the inherent timbre and LEV of physical spaces can be captured directly using microphone arrays combined with appropriate post-processing [5], [6], [7]. However, in the absence of spatial microphone arrays, a comparable spatial impression can be achieved by decorrelating a monaural RIR to derive multiple decorrelated reverberation tails [8].

For the generation of synthesized diffuse fields, computationally efficient methods utilizing velvet-noise (VN)—a sparse, pseudo-random sequence—have recently gained prominence [9], [10], [11], [12]. VN sequences offer a highly efficient way to decorrelate audio signals [13], [14], [15], with the advantage of becoming inherently colorless as their length approaches infinity [16]. For finite-length VN sequences, optimization techniques can be used to approximate a smooth frequency response [14]. Consequently, this study investigates the application of optimized VN sequences to generate decorrelated reverberation tails from monaural RIRs on an eight-channel loudspeaker setup. Furthermore, an additional novel approach is introduced where a single velvet noise sequence is distributed across the loudspeaker setup.

The remainder of this paper is organized as follows. Section 2.1 introduces the tested decorrelation strategies. Section 2.2 then details the specific criteria used to select individual VN sequences from a larger candidate pool. Section 2.3 presents an objective analysis of these decorrelation strategies within a simulated environment, focusing on spectral smoothness and interaural coherence (IC). Following this, Section 3 documents the subjective listening experiment designed to compare the methods, with the corresponding findings presented in Section 4. Finally, Section 5 concludes the paper with a summary of the main insights and potential directions for future work.

2. Method

2.1. Decorrelators

This section explains the different kind of decorrelators we applied to an eight channel loudspeaker setup depicted in Fig. 1. Generally, a dry signal $s[n]$ is first convolved with a monaural RIR $h_1[n]$. The resulting monaural reverberant signal $(s[n] * h_1[n])$ is then convolved with a channel-specific decorrelation filter $v_i[n]$

***Corresponding author**: leon.kaiser@kug.ac.at.*

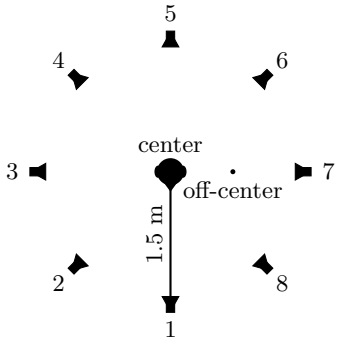

Figure 1: Loudspeaker setup for the evaluation of the decorrelation. Eight loudspeakers are arranged in a circle around a listening position at the center. An off-center position is additionally evaluated in the listening test. The numbers imply the channel numbers i.

for each channel i:

$$y_i[n] = (s[n] * h_1[n]) * v_i[n] \tag{1}$$

For sparse decorrelators $v_i[n]$, this architecture offers significant computational advantages over convolving the input signal with multiple decorrelated RIRs $h_i[n]$.

$$y_i[n] = s[n] * h_i[n] \tag{2}$$

For a 30-component VN sequence this can reduce the computational effort by 76 % [14].

2.1.1. Reference and anchor

For the reference, eight decorrelated white noise sequences were generated. Using an overlap-add procedure, the sequences were partitioned into blocks, and the magnitude spectrum of each block was matched to that of the original RIR. Thereby decorrelated reverberation tails $h_1[n]$ to $h_8[n]$ were created. The computational expensive reference stimuli were created by eight convolutions (see equation 2). For the anchor, which serves as a negative reference in the subsequent listening experiment, all eight loudspeakers played back the same first channel instance $y_1[n]$.

2.1.2. Original optimized velvet-noise (OVN)

Multiple versions of optimized VN decorrelators were inspected. Generally, a VN sequence $v[n]$ is defined as:

$$v[n] = \sum_{m=0}^{M-1} \sigma[m] A[m] \delta[n - \tau[m]] \tag{3}$$

where $\delta[n]$ denotes the Kronecker delta function and M represents the total number of Kronecker delta components. In this formulation, $\sigma[m] \in \{-1, 1\}$ is a random sign sequence (evenly distributed). The amplitude scaling $A[m]$ and the impulse locations $\tau[m]$ serve as the primary parameters for optimizing a VN sequence to achieve a flat frequency response. For a comprehensive derivation and details on these optimization procedure, the reader is referred to Schlecht et al. [14]. We performed the optimization utilizing the corresponding open source code [17].

Originally, VN decorrelators were designed with a 60 dB exponential decay within a 30 ms time window to prevent the smearing of transient signals [13]. Frequency-optimized VN sequences featuring an exponential decay are referred to as original Optimized Velvet-Noise decorrelators (OVN). The amplitude scaling of OVN $A_{\mathrm{OVN}}[m]$ is defined as:

$$A_{\mathrm{OVN}}[m] = g[m] 10^{\frac{-3\tau[m]}{f_s T}} \tag{4}$$

In this formulation, f_s represents the sampling frequency and $T = 0.03\,\mathrm{s}$ specifies the duration of the OVN. The factor $g[m]$ is optimized to ensure a smooth frequency response, with its values constrained such that:

$$20 \log_{10}(g[m]) \in [-6, 6]\mathrm{dB} \tag{5}$$

The 8 utilized OVN are depicted in Fig. 2 on the left.

2.1.3. Non-decaying optimized velvet-noise (NVN)

Since in diffuse fields temporal smearing is expected to be a less disturbing aspect, we also tested a VN version without exponential decay. Frequency-optimized VN sequences without exponential decay are refered to as Non-decaying Velvet-Noise decorrelators (NVN). Their amplitude scaling $A_{\mathrm{NVN}}[m]$ is:

$$A_{\mathrm{NVN}}[m] = g[m] \tag{6}$$

The 8 utilized NVN are also depicted in Fig. 2 in the middle.

2.1.4. Spatialized optimized velvet-noise (SVN)

Informal testing suggested that simple delays based on prime numbers can serve as an effective decorrelation strategy. Consequently, the same RIR was played back across all eight channels, with channels 1 through 8 being assigned delays of $\{1, 17, 3, 7, 13, 5, 2, 11\}$ ms, respectively.

Given the similarity to a VN sequence, we were motivated to try distributing a NVN across the 8 loudspeakers. Generally we define Spatialized Velvet-Noise (SVN) as:

$$v[n, \boldsymbol{\theta}] = \sum_{m=0}^{M-1} \sigma[m] A[m] \delta[n - \tau[m], \boldsymbol{\theta}] \tag{7}$$

Here, $\boldsymbol{\theta}$ denotes a direction of arrival in space. A greedy algorithm ensures approximately equal energy of $M = 32$ Kronecker delta components across all playback channels. Under the idealized conditions of an omnidirectional receiver and equidistant sources, SVN should theoretically exhibit the same frequency response as its non-spatial equivalent. However, it should be noted that Head-Related Transfer Functions (HRTFs) are expected to alter the perceived spectral smoothness of SVN. As previously mentioned, this work employs a NVN as the basis for the SVN; for the sake of readability, the term SVN refers to this specific optimized decorrelator in the following sections. The spatial distribution of the first NVN

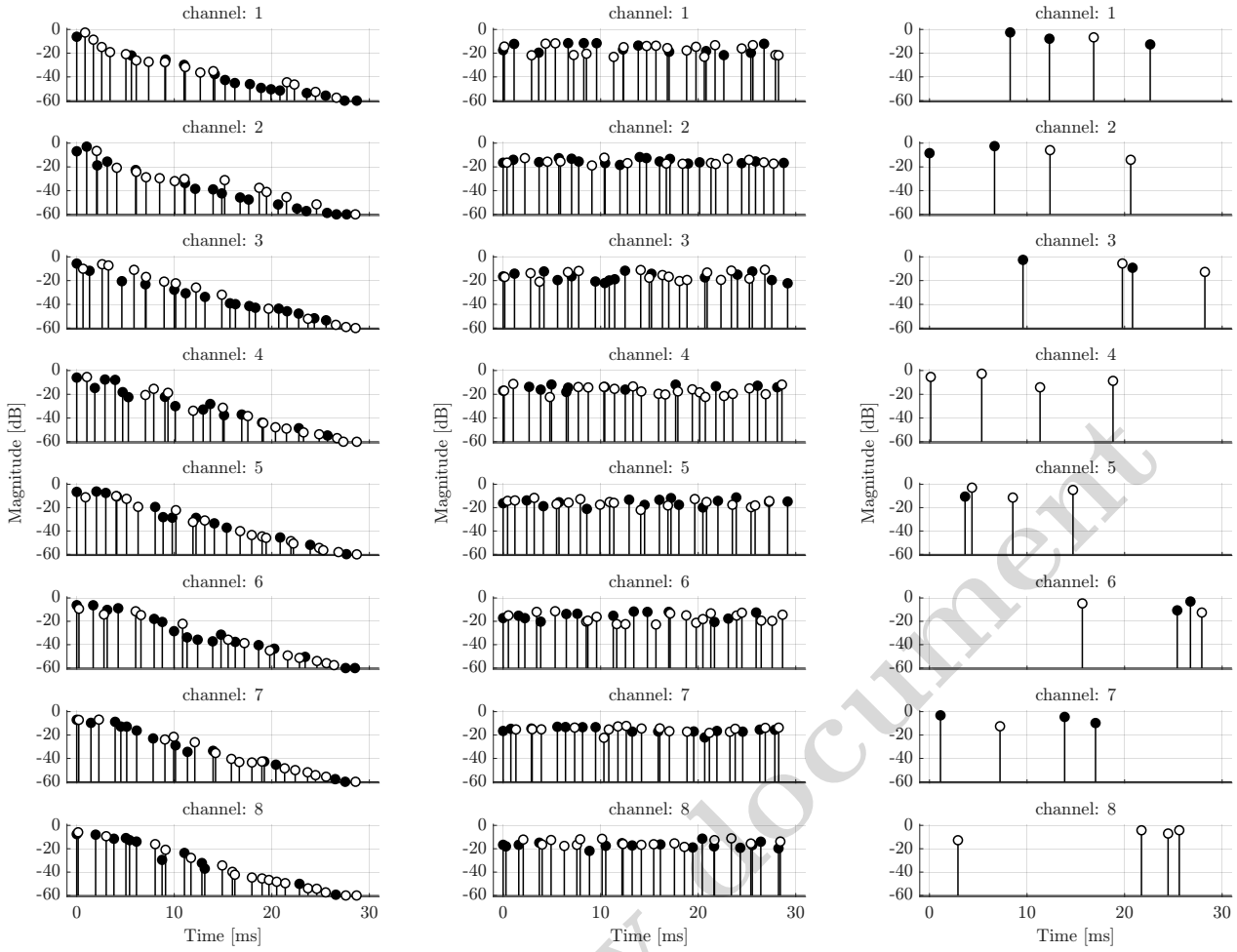

Figure 2: Comparison of OVN (left), NVN (center), and SVN (right). • indicates positive sign; ∘ indicates negative sign. All decorrelators are normalized so that their total energy is equal when added up over the 8 channels. The SVN is the NVN sequence of channel 1, distributed over the 8 channels and amplified with the factor $\sqrt{8}$.

channel is depicted in Fig. 2 on the right.

2.1.5. Allpass

The allpass filters were configured identically to the implementation described by Frank et al. [4]. The decays of the allpass filters ranged from approximately $T_{60} = 0.1\,\text{s}$ to $T_{60} = 0.2\,\text{s}$.

2.2. Selection of Decorrelators

For OVN and NVN, 500 sequences were generated using the framework provided by Schlecht [17]. From these data sets, eight sequences were selected in each case based on the smoothness of the frequency response $\mathcal{L}$ and the inter-sequence coherences $\overline{|\rho_{a,b}|}$. For details on these metrics the reader is refered to Schlecht et al. [14]. The final sequences were selected by imposing upper limits $\mathcal{L}_{max}$ and $\overline{|\rho_{a,b}|}_{max}$. The coherence limit $\overline{|\rho_{a,b}|}_{max}$ ensures that all pairwise coherence values remained below the threshold. To maintain a consistent trade-off between the criteria, the ratio was fixed at $B = \mathcal{L}_{max}/\overline{|\rho_{a,b}|}_{max} = 0.5$. An iterative algorithm varied both limits stepwise while preserving B until exactly eight sequences satisfied both conditions.

For OVN eight sequences were found as the limits reached $\mathcal{L}_{max} = 0.83\,\text{dB}$ and $\overline{|\rho_{a,b}|}_{max} = 0.413$. Empirically, a weighting of $B = 0.5$ favors coherence. This can be reasoned, since restricting $\overline{|\rho_{a,b}|}_{max}$ to 0.4 fails to yield eight sequences, even when neglecting the smoothness limit $\mathcal{L}_{max}$. NVN achieves tighter limits of $\mathcal{L}_{max} = 0.44\,\text{dB}$ and $\overline{|\rho_{a,b}|}_{max} = 0.22$. The improved performance of NVN can likely be attributed to the fact that Kronecker delta components with very small amplitudes ($|A[m]| < -30\,\text{dB}$) contribute only negligibly to both smoothing and decorrelation. Furthermore, an analysis of the frequency-dependent inter-sequence coherence reveals that while OVN optimization achieves a relatively low average coherence ($\overline{|\rho_{a,b}|}_{max} = 0.413$), it still exhibits high coherence peaks in specific frequency bands. In contrast, NVN successfully maintains consistently low coherence across the broadband spectrum.

The SVN sequence is derived from the NVN sequence with the smoothest frequency response.

2.3. Theoretical Evaluation

To maintain the original timbre, our goal is to ensure the channel-sum frequency response remains as smooth as possible, while keeping temporal roughness and smearing as minimal as possible. Concurrently, the frequency-dependent IC must align with the IC profiles of real physical environments [18].
The loudspeaker setup (see Fig. 1) was simulated at the center position with the IEM plugin suite[1] in fifth-order ambisonics. Specifically the Multi-Encoder is utilized to simulate the loudspeakers as point sources from the eight directions and the result is rendered by applying the binaural decoder [19]. Each simulated loudspeaker played 50 s of decorrelated white noise. Decorrelation was achieved either by generating different noise tracks (for the reference) or by applying the decorrelation strategies described in section 2.1 to a single noise track. Afterwards, Welch-Periodograms of the binaural decoded signals were used to estimate Power Spectral Densities (PSDs). The spectral similarity of the the p-th method to the reference is calculated by third octave smoothed Welch periodograms. S_{ll} represents the PSD of the reference on the left ear and $S_{ll,p}$ the PSD of the p-th method on the left ear.

$$\Delta S_p = 10 \log_{10} \left(\frac{S_{ll,p}}{S_{ll}} \right) \tag{8}$$

The results for ΔS_p are shown in Fig. 3.

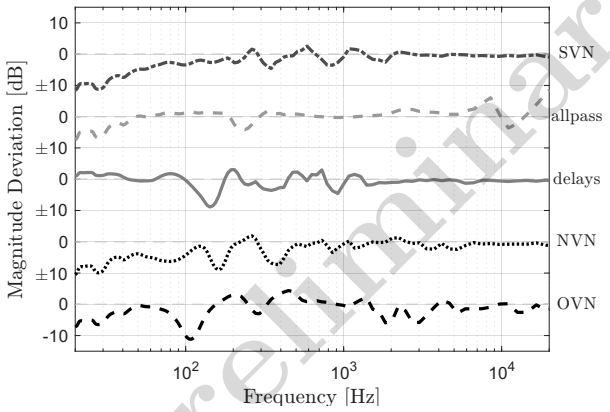

Figure 3: third-octave smoothed magnitude deviation ΔS_p from the reference

To calculate the IC, the PSDs at the right ear $S_{rr,p}$ and the cross-power spectral densities between the right and left ears $S_{lr,p}$ were calculated. Following the interaural coherence of the p-th method is estimated as:

$$\gamma_p = \sqrt{\frac{|S_{lr,p}|^2}{S_{ll,p} S_{rr,p}}} \tag{9}$$

The results are shown in Fig. 4.

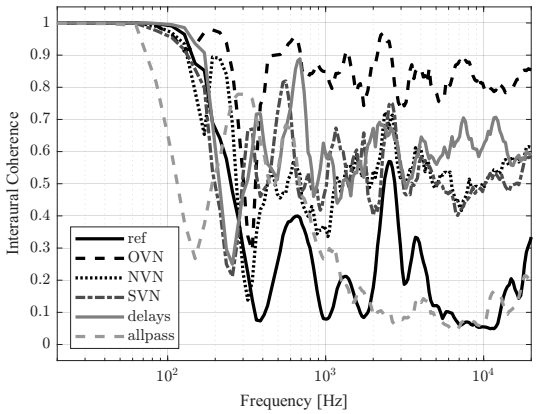

Figure 4: third-octave smoothed interaural coherence γ; 'ref' refers to the reference

3. Listening Experiment

A listening test was conducted in the anechoic chamber at the Institute of Electronic Music and Acoustics with 24 subjects (19 male, 4 female, 1 diverse; aged 20–62, median: 29). Stimuli were reproduced via Genelec 8020 loudspeakers arranged at a radius of 1.5 m and 1.2 m height. The experiment was performed at two listener positions: the geometric center and an off-center position 0.75 m laterally to the side (see Fig. 1). To rule out any order bias, the order in which the two positions were presented was alternated across the participants. At each position, subjects first completed an envelopment evaluation task, followed by a timbre evaluation task. In these tasks, participants had to rate the similarity between the reference and the presented stimuli in terms of envelopement and timbre. Each task consisted of four presentations, generated by convolving two dry excitation signals—speech (track 49) and music (track 70) from the EBU SQAM library [20]—with RIRs from two environments: an office ($T_{60} \approx 0.7$ s) and a church ($T_{60} \approx 2.4$ s). To construct the acoustic scenes, the direct sound and early reflection peaks were removed from the recorded RIRs. The remaining late reverberation tail was processed using the decorrelation strategies (see section 2.1) and reproduced via the 8-channel loudspeaker array. Artificial direct sound and early reflections were added using the dry signals: a direct sound originated from the center ($i = 1$) speaker, accompanied by left ($i = 3$) and right ($i = 8$) reflections delayed by 6 ms and 9 ms, respectively. The reflections were attenuated by 6 dB relative to the direct path. The salient-to-diffuse energy ratio (direct and reflections versus diffuse tail), measured using a white noise excitation for the reference condition, was -2 dB for the office scenario and -8 dB for the church scenario.
The similarity ratings followed the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) methodology [21]. The evaluated items comprised the hidden reference, the anchor, and five decorrelation

[1] https://plugins.iem.at/

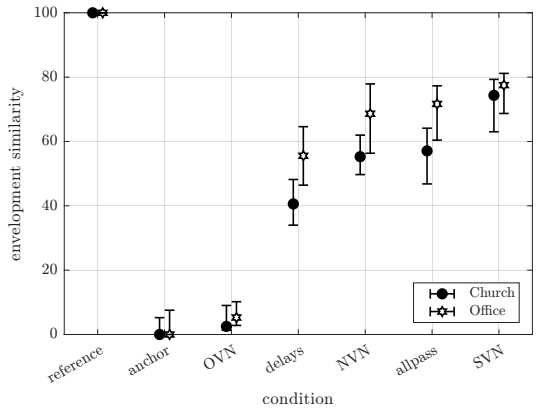

Figure 5: Medians and corresponding 95% confidence intervals for envelopment at center position

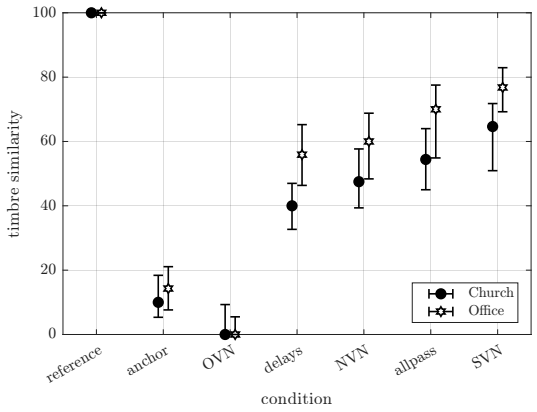

Figure 6: Medians and corresponding 95% confidence intervals for timbre at center position

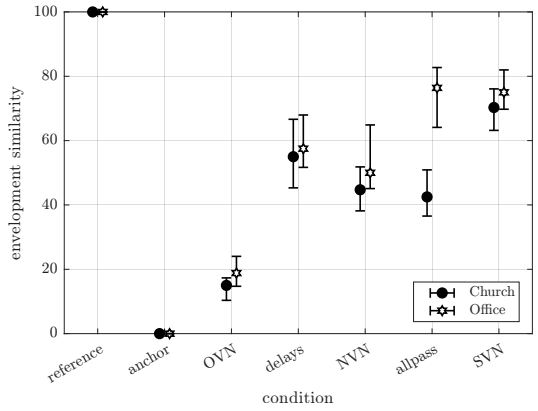

Figure 7: Medians and corresponding 95% confidence intervals for envelopment at off-center position

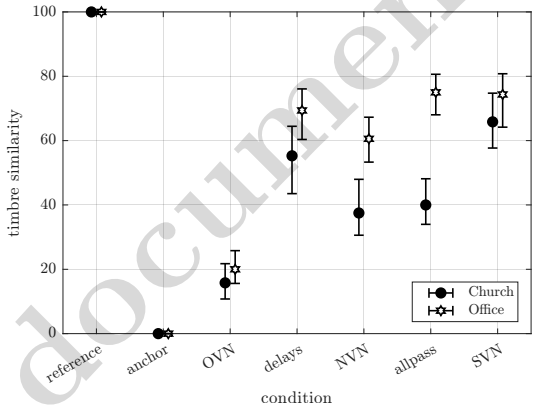

Figure 8: Medians and corresponding 95% confidence intervals for timbre at off-center position

methods under test (OVN, NVN, SVN, delays, and allpass filters). For every similarity rating, participants used sliders with a resolution of 20 steps to rate the eight stimuli against the reference. The presentation order of the four MUSHRA pages in a task (2 types of rooms (office, church) × 2 types of excitation (speech, music)), as well as the alignment of the seven individual stimuli within each MUSHRA page, was fully randomized. Binaural renders from the listening test (simulated with IEM-plugins in fith-order ambisonics) and the VN decorrelators are available via open access on Zenodo: `https://zenodo.org/records/21625598`

4. Results

The subjective ratings for envelopment and timbre across both listener positions are illustrated in Fig. 5–8. Prior to analysis, all raw scores were min-max normalized to a standard 0–100 scale. Statistical significance was assessed using Wilcoxon signed-rank tests with a Bonferroni-Holm correction. Across all conditions, the hidden reference was consistently rated most similar to itself. Conversely, the anchor performed worst in all but one case, where the OVN method was rated lower. Among the efficient decorrelation strategies, SVN was not significantly outperformed on any of the tasks.

4.1. Results – Center Position

For envelopment (Fig. 5), SVN significantly outperformed all other methods ($p < 0.03$), except NVN and the allpass filters. In the church scenario SVN showed marginal significance ($p < 0.06$) and in the office scenario no significance ($p = 1.0$) against NVN and the allpass filters.
For timbre (Fig. 6), SVN significantly outperformed all methods ($p < 0.01$) except NVN ($p = 0.15$) and the allpass filters ($p = 0.61$) in the church scenario, and all methods ($p < 0.03$) except the allpass filters ($p = 0.35$) in the office scenario.

4.2. Results – Off-Center Position

For envelopment (Fig. 7), SVN proved highly robust to the lateral shift, significantly outperforming all other methods in the church scenario ($p < 0.01$) and all methods ($p < 0.02$) except the allpass filters ($p = 0.43$) in the office scenario.
For timbre (Fig. 8), SVN again yielded significantly higher ratings than all other methods in the church scenario ($p < 0.05$). In the office scenario, however, no significant differences ($p > 0.14$) were observed in comparison to the SVN, with the notable exception

of OVN, which performed significantly worse than the other tested methods ($p < 10^{-4}$).

5. Conclusion

This work demonstrates that, for an eight channel setup, an OVN with a decay factor of 60 dB is insufficient to generate a diffuse sound field. While eliminating the decay introduces perceptual roughness, it significantly enhances decorrelation, which appears to be more critical for overall preference. Further perceptual improvements were achieved by spatially distributing a single optimized, non-decaying sequence—a method introduced here as spatialized velvet-noise (SVN). In our eight-channel scenario, distributing a single VN sequence across the channels results in an eightfold reduction in computational effort compared to using one VN sequence per channel. Consequently, SVN serves as most viable alternative to using multiple decorrelated RIRs when the computational advantages outweigh the reductions in perceptual quality. Furthermore, our findings suggest that VN sequences with a 60 dB decay can simply be truncated without compromising performance, as low-amplitude components appear to contribute minimally. Future work may focus on optimizing SVN parameters, specifically regarding decay duration and spatial configuration aspects.

References

[1] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of model-and signal-based predictors of the mixing time in binaural room impulse responses," *Journal of the Audio Engineering Society*, vol. 60, no. 11, pp. 887–898, 2012.

[2] S. Riedel, M. Frank, and F. Zotter, "The effect of temporal and directional density on listener envelopment," *Journal of the Audio Engineering Society*, vol. 71, no. 7/8, pp. 455–467, 2023.

[3] S. Riedel, M. Frank, and F. Zotter, "Towards models for listener envelopment based on binaural multi-source localization," *DAS—DAGA2025 Copenhagen*, 2025.

[4] M. Frank and L. Gölles, "Decorrelating the diffuse part in ambisonic spatial decomposition method," *Fortschritte der Akustik, DAGA*, 2026.

[5] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer, 2019.

[6] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc*, vol. 61, no. 1/2, pp. 17–28, 2013.

[7] L. Gölles and M. Frank, "Ambisonic spatial decomposition method with salient/diffuse separation," in *Proceedings of the 158th AES Convention, Warsaw, Poland*, 2025.

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

[9] J. Fagerström, B. Alary, S. Schlecht, and V. Välimäki, "Velvet-noise feedback delay network," in *International Conference on Digital Audio Effects*, DAFx, 2020, pp. 219–226.

[10] V. Välimäki and K. Prawda, "Late-reverberation synthesis using interleaved velvet-noise sequences," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1149–1160, 2021.

[11] J. Fagerström, *Velvet noise in audio processing*. Aalto University, 2025.

[12] K. Prawda, S. Schlecht, and V. Välimäki, "Multichannel interleaved velvet noise," in *International Conference on Digital Audio Effects*, DAFx, 2022, pp. 208–215.

[13] B. Alary, A. Politis, V. Välimäki, et al., "Velvet-noise decorrelator," in *Proc. Int. Conf. Digital Audio Effects (DAFx-17), Edinburgh, UK*, 2017, pp. 405–411.

[14] S. Schlecht, B. Alary, V. Välimäki, and E. A. Habets, "Optimized velvet-noise decorrelator," in *International Conference on Digital Audio Effects*, University of Aveiro, 2018, pp. 87–94.

[15] J. A. Belloch, J. M. Badia, G. Leon, and V. Välimäki, "Efficient velvet-noise convolution in multicore processors," *AES: Journal of the Audio Engineering Society*, vol. 72, no. 6, pp. 383–393, 2024.

[16] N. Meyer-Kahlen, S. J. Schlecht, and V. Välimäki, "Colours of velvet noise," *Electronics Letters*, vol. 58, no. 12, pp. 495–497, 2022.

[17] S. J. Schlecht, *OptimizedVelvetDecorrelators: Implementation of Optimized Velvet Noise Decorrelators*, `https://github.com/SebastianJiroSchlecht/OptimizedVelvetDecorrelators`, 2020.

[18] J. Fagerström, N. Meyer-Kahlen, S. J. Schlecht, and V. Välimäki, "Binaural dark-velvet-noise reverberator," in *Proceedings of the International Conference on Digital Audio Effects*, 2024, pp. 246–253.

[19] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[20] EBU. "Sqam cd: Sound quality assessment material. "[Online]. Available: `https://tech.ebu.ch/publications/sqamcd`.

[21] *Method for the subjective assessment of intermediate quality level of coding systems*, ITU-R BS.1534-1, 2001.