# DELAY OPTIMIZATION TOWARDS SMOOTH SPARSE NOISE

*Cristóbal Andrade and Sebastian J. Schlecht*

Multimedia Communications and Signal Processing,
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
cristobal.andrade@fau.de

## ABSTRACT

Smooth sparse noise sequences are applied to efficiently model reverberation. This paper addresses the problem of optimizing sparse noise sequences for perceptual smoothness using gradient-based methods. We demonstrate that sinc-shaped artifacts introduced by fractional delay create non-convexities in an envelope-based roughness loss function, hindering delay optimization. By temporarily removing pulse polarity and omitting envelope rectification, we obtain a convex loss suitable for gradient descent. Pulse signs are reintroduced after optimization during synthesis. Optimization results show roughness reduction across various pulse densities, with the optimized sequences approaching the perceptual smoothness of velvet noise.

## 1. INTRODUCTION

Sparse noise sequences have been widely explored in the literature, with various pulse distributions proposed to shape their temporal roughness and spectral properties. Examples of sparse noise sequences include Velvet Noise [1], Dark Velvet Noise [2], Totally Random Noise [3], and Additive Random Noise [4]. Velvet noise is known for its smooth perceptual quality, yet it remains unclear whether its pulse distribution achieves optimal smoothness. Obtaining smoother distributions involves optimizing pulse placements, a task that is computationally challenging due to its combinatorial nature.

Optimizing the pulses' placements is, in essence, equivalent to designing the positions of non-zero coefficients in sparse FIR filter design. This problem is typically formulated by representing the non-zero coefficient positions via an $l_0$-norm and optimizing it, yielding an inherently nonconvex challenge. Existing methods involve approximating the $l_0$-norm by an $l_p$-norm [5], zero coefficient positions search algorithms [6, 7], and greedy methods [8, 9]. Optimizing sparse noise sequences by allowing continuous values and rounding to the nearest integer has been proposed [10]. However, little research has explored gradient-based or continuous optimization approaches.

This paper investigates the use of gradient descent for pulse placement optimization in sparse noise sequences. Fractional delays introduce sinc-shaped impulse responses, where the sinc function is defined as $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$. We show that these sinc-like artifacts from fractional delays are the main source of non-convexities, which hinder convergence. Using a roughness model based on [1], we expose these issues, modify the loss formulation to enable delay optimization, and finally demonstrate the resulting perceptual improvements.

The organization of this paper is as follows. Section 2 introduces the fundamentals of fractional delays, describes the noise sequences used for optimization, and introduces a roughness-based loss metric. Section 3 analyzes the convexity of the proposed loss function and presents the delay optimization results.



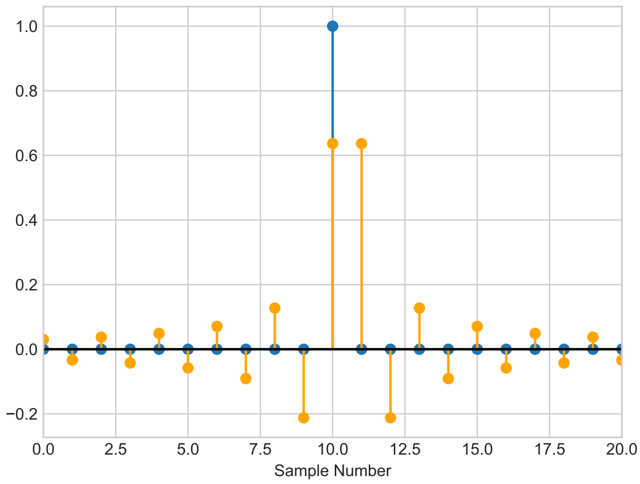


Figure 1: Blue: Impulse shifted by 10 samples. Yellow: Impulse shifted by 10.5 samples. Applying a delay that includes a fractional portion results in a sinc-shaped time response.

## 2. BACKGROUND

### 2.1. Fractional Delays

A fractional delay refers to a delay, which is not a multiple of the sampling interval. The time response of a fractional delay is a shifted sample sinc function [11]. That is

$$h(n) = \mathrm{sinc}(n - D) \tag{1}$$

where $n$ is the integer sample index and D is the delay decomposed into an integer and fractional part. Figure 1 shows the impulse response of an integer delay of 10 samples and the impulse response of a delay containing an integer portion of 10 samples and a fractional portion of 0.5 samples.

In Section 3, we show that the sinc-shaped renspose introduced by fractional delays, degrade the convexity of the roughness loss, which must be addressed to enable gradient-based optimization.

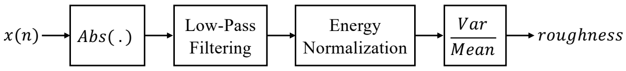


Figure 2: Proposed roughness model based on [1].

### 2.2. Sparse Noise Sequences

In this section, two sparse noise sequences are introduced: Velvet Noise and Totally Random Noise. The main difference between them lies in their pulse distributions, which give rise to different levels of temporal roughness.

#### *2.2.1. Velvet Noise*

Velvet Noise (VN) is synthesized by placing unit-impulses at pseudo-random sample positions and assigning each impulse a random sign. Formally

$$s_v(n) = \sum_{m=0}^{M} a(m)u\left(n - \text{round}\left(\frac{T_\text{d}}{T_s}(m + \text{rnd}(m))\right)\right) \quad (2)$$

where $rnd(m)$ is a uniform random number in $[0, 1)$, $a(m)$ randomly takes values $\pm 1$, $round(\cdot)$ rounds to the nearest integer, $T_d$ is the average impulse interval (inverse of impulse density), and $T_s$ is the inverse of the sampling rate.

Velvet Noise was designed to minimize pulse density while maximizing perceptual smoothness [1], yet whether its pulse distribution is truly optimal remains an open question.

#### *2.2.2. Totally Random Noise*

In Totally Random Noise (TRN), pulses can be placed anywhere in the sample grid with equal probability [12]. TRN was originally proposed to be generated by rounding a scaled, offset uniform random sequence to achieve the desired pulse density following

$$s_{\text{tr}}(n) = round\left(\frac{T_\text{d}}{T_\text{d} - 1}\left[r(n) - \frac{1}{2}\right]\right) \quad (3)$$

where $T_d$ denotes the inverse of the impulse density, $r(n)$ is a sequence of values uniformly distributed between 0 and 1, and $round()$ rounds to the nearest integer.

Alternatively, TRN can be synthesized by randomly selecting impulse positions to meet the target density and assigning each pulse a random ±1 sign. TRN is reported to sound rougher than Velvet Noise [12].

### 2.3. Roughness Loss Function

The psychoacoustic sensation of roughness arises from quick changes in modulating frequencies within the range of 15 to 300 Hz [13]. This psychoacoustical phenomenon is often associated with the sound produced when articulating a rolling "r".

To model the roughness [1] proposes to obtain the envelope fluctuation by rectifying the noise signal and low-pass filtering it to match auditory temporal resolution. Then the roughness is computed as the variance-to-mean ratio of the envelope, expressed in decibels. To ensure fair comparisons across different pulse densities, we add an energy-normalization step that removes differences caused purely by signal energy. Figure 2 shows a flow diagram of the proposed model.

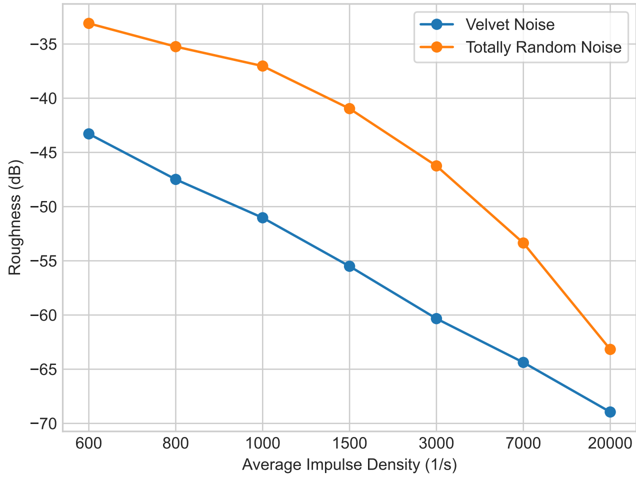


Figure 3: Roughness values obtained with the proposed model for Velvet and Totally Random Noise across various pulse densities. The results demonstrate that the pulse-density distribution influences the roughness of the noise sequence.

Figure 3 shows the roughness values of Velvet and Totally Random Noise across various pulse densities, as computed by our model. These results illustrate the importance of optimizing pulse placement to minimize roughness at any given impulse density.

## 3. ROUGHNESS OPTIMIZATION

In this section, we begin by examining the convexity of our proposed loss function and identifying how fractional-delay artifacts compromise its convexity. We then introduce a modification that yields a convex loss formulation. Finally, we apply the revised loss to optimize the roughness of sparse noise sequences, using TRN as the starting point.

### 3.1. Loss Function Analysis

To determine whether the roughness measure defines a convex loss, we synthesize a TRN sequence, select a single pulse, shift it incrementally between its neighbors, and evaluate the loss at each position.

Figure 4 shows the roguhness obtained by shifting a single pulse in a TRN sequence of 600 pulses per second, for both integer and fractional delay shifts. The integer-only shifts produce a smooth, discrete convex loss curve, whereas the inclusion of fractional delays presents dips leading to a non-convex curve. This behavior must be addressed to be able to utilize the proposed roughness model as a loss metric.

The dips observed in Figure 4 results from the absolute value operation used in our model, which rectifies the signal prior to computing its statistics. A rectified pulse should yield a consistent mean value regardless of its position. However, due to the sinc-shaped pulses introduced by fractional delays, this assumption becomes invalid, causing undesired fluctuations in the loss function. Since these sinc artifacts are purely digital and roughness perception is insensitive to pulse polarity, we can eliminate the sign information and corresponding absolute-value step during loss evaluation and then reassign the original signs when synthesizing the

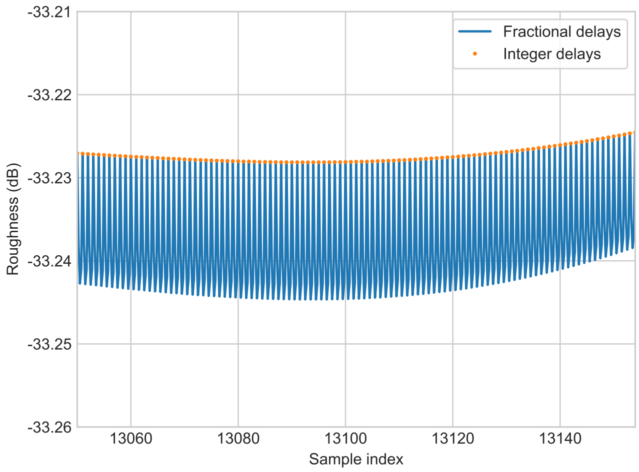


Figure 4: Roughness loss function evaluated by shifting a single pulse of a TRN. The blue curve includes fractional delay shifts, while the orange curve considers only integer delays. The dips observed when using fractional delays illustrate the instability of the delay optimization problem.

final sequence. Figure 5 compares the original loss function, incorporating pulse signs and absolute-value rectification, with a modified loss function that omits rectification by considering signless pulses. Figure 5 shows that removing pulse signs and excluding

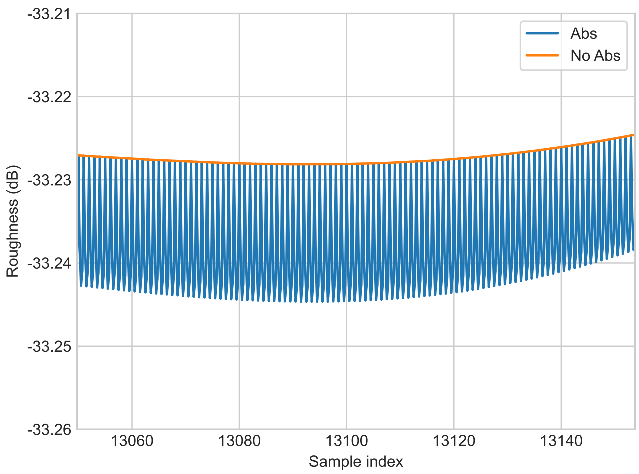


Figure 5: Roughness loss function evaluated by shifting a single pulse, including fractional sample shifts. The blue curve includes both pulse signs and absolute-value rectification. The orange curve omits pulse signs and rectification, resulting in a convex loss.

the absolute-value operation during analysis results in a convex loss function. This modified approach is therefore well-suited for delay optimization.

### 3.2. Delay Optimization

We begin with a TRN sequence and optimize its pulse positions to reduce roughness. To ensure the loss function remains convex, we temporarily ignore the signs of the pulses and omit the rectification

Table 1: *Roughness values (in dB) computed using the proposed model. Results are presented for different pulse densities (1/s), detailing values for the initial totally random noise, optimized sequence. Velvet Noise (VN) is provided as a reference.*

| Density | Init | Optimized | VN |
|---|---|---|---|
| 600 | -33.54 | -40.14 | -43.08 |
| 800 | -35.73 | -43.33 | -47.44 |
| 1000 | -37.28 | -46.69 | -50.74 |
| 1500 | -40.80 | -51.53 | -55.22 |
| 3000 | -46.53 | -58.6 | -60.34 |
| 7000 | -52.46 | -62.83 | -64.37 |

step in the roughness model. After optimization, the original pulse signs are reintroduced during synthesis. Throughout this procedure, the pulse densities remain constant. The optimization is carried out using the open-source Frequency-Domain Differentiable Audio Processing library, FLAMO [14].

Table 1 compares the roughness of the initial TRN, the resulting optimized sequence, and VN at various densities. Delay optimization consistently lowers the roughness of the initial TRN across all densities, bringing it close to the levels achieved by VN. The audio results from this optimization are available on the supplementary GitHub page[1], including a variant in which optimized pulse positions are rounded to the nearest integer.

Informal listening confirms that the optimized sequences sound noticeably smoother than TRN and approach the smoothness of VN, in agreement with the numerical results in Table 1. Moreover, rounding the optimized pulse positions has negligible impact on the perceived roughness, indicating that integer delay implementations can effectively replicate the optimized results.

## 4. CONCLUSION

We identified fractional delays as the main obstacle in delay optimization and proposed a method to minimize the roughness of sparse noise sequences. Fractional delays create sinc-shaped artifacts that distort the envelope used in roughness calculations. These distortions lead to non-convexities in the loss function, which hinder gradient-based optimization. By removing the sign of the pulses and skipping rectification during optimization, we avoid these artifacts and ensure a convex loss. This is valid since the envelope is polarity-invariant, and polarity can be reintroduced after optimization during synthesis.

The proposed method successfully reduces the roughness of Totally Random Noise for various pulse densities. Informal listening confirms a perceptual improvement, with the optimized sequences approaching the smoothness of velvet noise.

## 5. REFERENCES

[1] M. Karjalainen and H. Järveläinen, “Reverberation modeling using velvet noise,” *Proceedings of the AES International Conference*, Jan. 2007.

[2] J. Fagerström, N. Meyer-Kahlen, S. J. Schlecht, and V. Välimäki, “Dark velvet noise,” in *Proceedings of the 25th International Conference on Digital Audio Effects*

[1] https://cristobalandrade.github.io/Delay-Optimization-DAFx-2025/

*(DAFx20in22)*, Austria, 2022, Proceedings of the International Conference on Digital Audio Effects, pp. 192–199, DAFx.

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

[4] I. Bilinskis and A. K. Mikelson, *Randomized Signal Processing*, Prentice-Hall, Inc., USA, 1992.

[5] T. Baran, D. Wei, and A. V. Oppenheim, "Linear programming algorithms for sparse filter design," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1605–1617, 2010.

[6] J. T. Kim, W. J. Oh, and Y. H. Lee, "Design of non-uniformly spaced linear phase fir filters using mixed integer linear programming," in *1995 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1995, vol. 2, pp. 877–880.

[7] Song Y.S. and Lee Y. H., "Design of sparse fir filters based on branch-and-bound algorithm," in *Proceedings of 40th Midwest Symposium on Circuits and Systems.*, 1997, vol. 2, pp. 1445–1448.

[8] Y. Wenbin and Y. Ya Jun, "Greedy algorithm for the design of linear-phase fir filters with sparse coefficients," *Circuits, Systems, and Signal Processing*, vol. 35, 07 2015.

[9] C. Rusu and D. Bogdan, "Iterative reweighted l1 design of sparse fir filters," *Signal Processing*, vol. 92, no. 4, pp. 905–911, 2012.

[10] Sebastian J Schlecht, Benoit Alary, Vesa Välimäki, and Emanuël A P Habets, "Optimized velvet-noise decorrelator," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Aveiro, Portugal, 2018, pp. 1 – 8.

[11] V. Valimaki and T.I. Laakso, "Principles of fractional delay filters," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, 2000, vol. 6, pp. 3870–3873.

[12] V. Välimäki, H. Lehtonen, and M. Takanen, "A perceptual study on velvet noise and its variants at different pulse densities," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, pp. 1481–1488, Jul. 2013.

[13] E. Zwicker and H. Fastl, *Roughness*, pp. 257–264, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[14] G. Dal Santo, G. De Bortoli, K. Prawda, S. J. Schlecht, and Vesa Välimäki, "Flamo: An open-source library for frequency-domain differentiable audio processing," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.