# Threshold for pulse density of velvet noise used in artificial reverberation

**Juan Almaraz**

Ingenieria de Sonido, Universidad Nacional de Tres de Febrero.
*juan.almaraz097@gmail.com*

Abstract - *This study explores the minimum pulse density required for velvet noise to be subjectively perceived as equal to Gaussian white noise when used in artificial reverberation. For this purpose, a subjective test was conducted with 28 participants in which three musical excerpts were evaluated. Considering the minimum effective duration ($\tau_{e(min)}$) of the autocorrelation function of each excerpt, it was observed that the higher the $\tau_{e(min)}$, the lower the required pulse density is. Particularly, the music fragment having a $\tau_{e(min)}$ equal to 4 ms was associated with a pulse density threshold of 330 p/s. The one with a minimum effective duration of 42 ms resulted in a pulse density threshold equal to 31 p/s. The excerpt with 108 ms of $\tau_{e(min)}$ could not be associated with any valid pulse density threshold, indicating a subjective perception limitation of stimuli with a high enough $\tau_{e(min)}$.*

## 1. INTRODUCTION

When recording musical instruments, there is generally no access to spaces with desired acoustic characteristics, such as a theater, a concert hall, etc. This is mainly due to the cost of moving all the necessary equipment for the recording, not only economically but also in time and effort. For this reason, the addition of artificial reverb is a very recurrent line of research that records contributions since the publication of 'Natural Sounding Artificial Reverberation' (Schroeder, 1961). There are many artificial reverberation algorithms, all of which aim to emulate the propagation of sound interacting with the room surfaces, carrying with it to the listener an imprint of the space, including objects, its architecture and geometry.

Valimaki et al. (2012) compiled fifty years of research on the topic, dividing the most common algorithms into three categories:

- Delay networks, in which the input signal is delayed, filtered and fed back along a number of paths according to parametrized reverberation characteristics.
- Convolutional, wherein the input signal is simply convolved with a recorded or estimated impulse response of an acoustic space.
- Computational acoustic, wherein the input signal drives a simulation of acoustic energy propagation in the modeled geometry.

This investigation focuses on the convolutional reverberation, especially when using synthesized RIRs based on noise sequences.

Moorer (1979) first suggested that the late part of the RIR can be well characterized as exponentially decaying white noise. Later on, Rubak and Johansen (1998) proposed the first sparse-noise-based reverberation algorithm, using a type of pseudorandom noise called totally random noise (TRN). This sequence has an equal probability of any sample being zero or non-zero. For this reason, large groups of impulses can appear, as well as large gaps of consecutive zeros.

A major improvement in pseudorandom noise sequences has been the proposal of velvet noise made by Järveläinen and Karjalainen (2007). They named it velvet noise because it sounded smoother than other known noise sequences. The steps to generate this sequence are the following: first generate an evenly spaced impulse train with a determined pulse density, then randomly change each impulse position and finally randomly assign a positive or negative value to each impulse. Following the steps described above, the resulting signal consists mostly of zeros, with some samples being -1 or 1.

This smooth-sounding ternary random noise is considered featureless and has a flat power spectrum like white noise. In addition, broadband velvet noise has been shown to retain its perceived smoothness with lower pulse densities in comparison to other types of sparse noise sequences (Välimäki et al., 2013).

Another application for velvet noise can be found in audio signal decorrelation, a critical process for spatial sound reproduction on multichannel configurations. (Alary et al., 2017; Schlecht et al., 2018). Additionally, velvet-noise has been used in vocoder-based speech generation by serving as excitation signals (Kawahara et al., 2018), and has also been used in acoustic measurements (Kawahara et al., 2019).

An important variable when using velvet noise is its pulse density. According to Järveläinen and Karjalainen (2007), a minimum of 1500 pulses per second is required to sound as smooth as Gaussian white noise. Roberts et al. (2023) studied the frequency-dependant roughness of velvet noise, finding that this minimum pulse density can be decreased when using lowpass or octave filters.

This work also aims to further investigate the number of pulses per second required for the velvet noise to be perceived subjectively equal to white noise (when used in artificial reverberation). The problem is that this threshold is deeply related to the characteristics of the stimuli. For this reason, this study proposes to evaluate the stimuli through their minimum effective duration ($\tau_{e(\mathrm{min})}$) (Sato & Wu, 2010). Because of the scale of this investigation, it is not possible to fully characterize the relation between these two parameters. Future work may further investigate the relationship between the $\tau_{e(\mathrm{min})}$ and the pulse density threshold using the results obtained in this investigation as a starting point.

# 2. METHOD

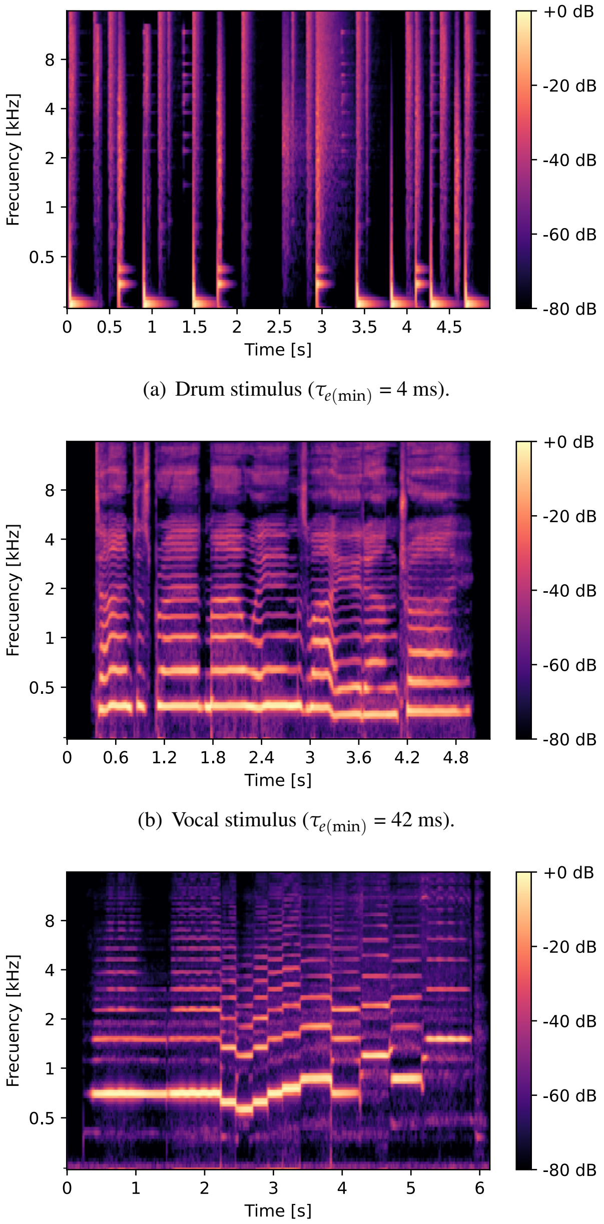
## 2.1 Test stimuli

The audio signals selected are based on excerpts of a synthesized drum kit (Figure 1(a)), a female singing voice (Figure 1(b)) and a flute (Figure 1(c)). From now on, each stimulus described above will be referenced as drum, vocal and flute, respectively.

These excerpts have a similar duration around 5 seconds, a sample rate of 48 kHz and a bit depth of 16 bits. The minimum effective duration value $\tau_{e(\mathrm{min})}$ is equal to 4 ms for the drum stimulus, 42 ms for the vocal stimulus and 108 ms for the flute stimulus. These values were obtained by performing a running-ACF analysis of the audio signals using a MATLAB program (Sato, 2014). The integration interval, running step and maximum delay time of the ACF were set to 1 s, 0.1 s and 0.2 s, respectively.

The test stimuli were obtained by performing a convolution between the excerpt and a synthetic RIR with 1s of reverberation time. Direct sound was added as an impulse in the first sample and a 9 ms time delay gap was added before the beginning of the reverberation tail. The magnitude of the direct sound impulse was calculated to obtain a direct to reverberant ratio of 10 dB. These values were taken from a subjective study conducted by Rubak and Johansen (1998). Figure 2 shows a block diagram of the process.

Since the purpose of this research is to find a threshold, different velvet noises were generated with an evenly increasing pulse density. For the vocal stimulus, the pulse density ranges from 10 to 70 p/s with a 10 p/s step. In the case of the drum stimulus, it ranges from 50 to 350 p/s with a 50 p/s step. For the flute stimulus, it ranges from 3 to 13 p/s with a 1 p/s increment. This is because the hypothesized threshold is different in each stimulus.



(a) Drum stimulus ($\tau_{e(\mathrm{min})}$ = 4 ms).

(b) Vocal stimulus ($\tau_{e(\mathrm{min})}$ = 42 ms).

(c) Flute stimulus ($\tau_{e(\mathrm{min})}$ = 108 ms).

Figure 1: Mel spectrogram of the audio signals.

For the reference stimulus, the same process is performed using white noise instead of velvet noise.

## 2.2 Subjecive test (forced ABX)

Before starting the test, listeners were asked their age, their listening experience and the type of headphones used for the test (in ear, over ear or speakers). In addition, it was suggested to use headphones during the test. For this reason, if the participant answered the test using speakers, their response was discarded.

A training stage was added to introduce the listener to the test procedure. It consisted of an ABX comparison with a clear difference between the two stimuli. Audio A was the drum excerpt convolved with GWN and audio B was the same drum excerpt convolved with velvet noise with a pulse density equal to 20 p/s. In this phase, the

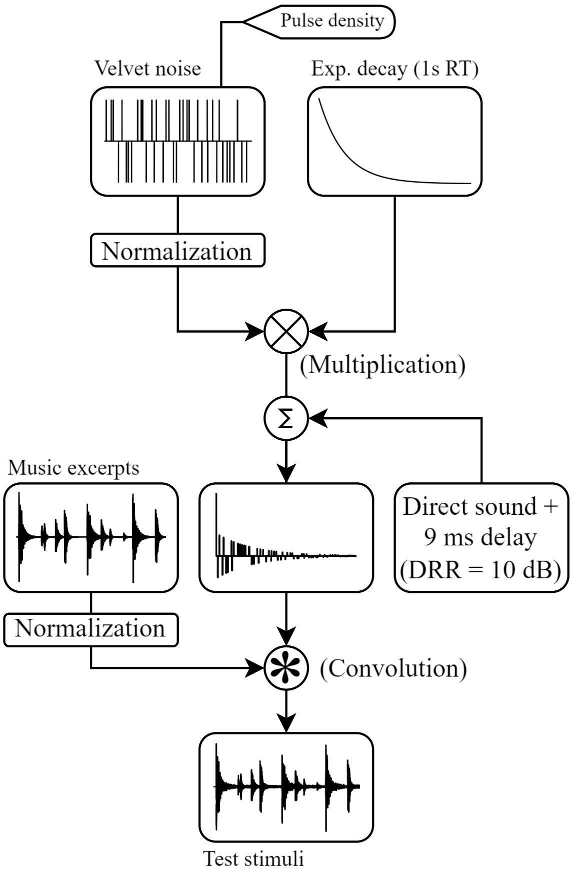


Figure 2: Block diagram of the creation of the stimuli.

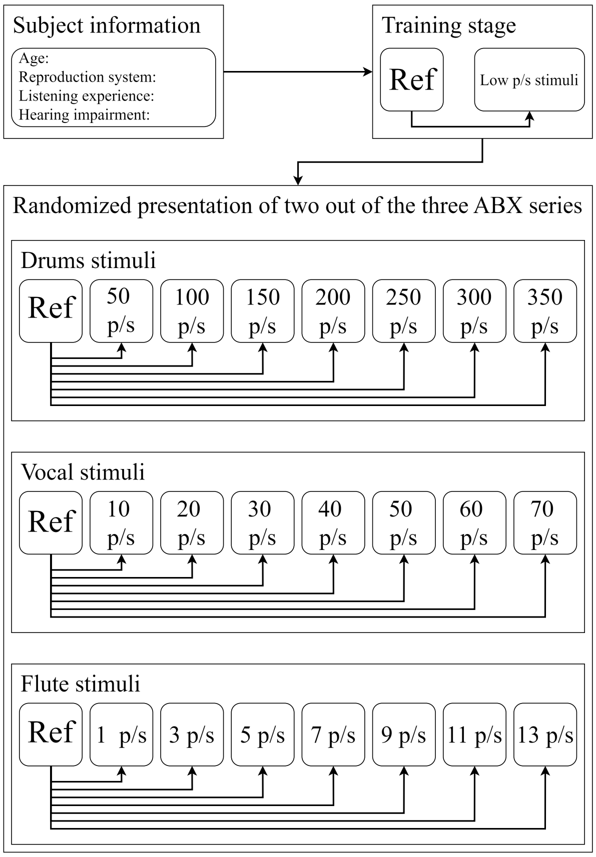


Figure 3: Block diagram of the conducted test.

listener is indicated to set a comfortable listening volume and keep it fixed for the rest of the test.

For each music excerpt and for each pulse density, an ABX test was performed. The test consisted of presenting three stimuli to the subject, with the ability to instantly switch between them. Stimuli A and B were the reference stimuli (Gaussian white noise) and a test stimulus with a determined pulse density value. Stimulus X was either one of the previously mentioned. The subject had to indicate if X is A or if X is B, and it was suggested to listen to each stimulus no more than three times before making a decision. If there were no perceived differences, the subject had to choose randomly one of the options.

For each stimulus (drum, vocal or flute), seven trials were performed, each one with a different pulse density in the described range.

From the three music excerpts, subjects were randomly shown only two of them. This is because each stimulus requires seven comparisons with the reference stimulus, and if presenting all three of them, the length of the test may have caused hearing fatigue and inaccurate responses. Also, the order of the stimuli was randomized. Figure 3 shows an example block diagram of the performed subjective test.

Each subject is required to perform 15 trials in total (1 training stage, 7 trials with the music excerpt #1 and 7 with the music excerpt #2) and the estimated duration of the test is around 6 minutes. The subjective test was conducted online using the website "ABX - Listening Tests as a service" (n.d.).

## 3. RESULTS

A total of 28 participants completed the test, resulting in 18 responses for the drum stimulus, 19 for the vocal stimulus and 19 for the flute stimulus.

The age of the subjects range from 17 to 59 years old. The average age of the listeners of participants was 30 years, with a standard deviation of 10 years.

Regarding the playback system used for the test, 68% of the participants used over-ear headphones and 32% used in-ear headphones. Because there were no participants who used speakers for the test, no answers were discarded by this criteria.

From the 28 listeners, 39% had some listening experience, 32% were considered musicians or music producers, 18% had a music related hobby and 10% selected the option 'none of the above'.

The number of correct answers is modeled as a random variable that follows a binomial distribution. If the answer is correct, the value 1 is assigned to that response. If the answer is incorrect, it corresponds to the value 0.

On the other hand, the number of correct answers required for a 95% confidence level on the existence of an audible difference is expressed in Equation (1).

$$\text{95\% confidence level} = N/2 + \sqrt{N} \qquad (1)$$

with N being the total number of trials.

Following Equation (1), the percentage of correct answers for a 95% confidence level is 73.6% for N = 18, and 72.9% for N = 19.

These values are represented as a dashed red line in Figure 4. The percentage of correct answers is represented as a black dot, and the linear interpolation of these values is indicated with a solid blue line. The pulse density threshold for each stimulus is obtained with the interception of the 95% confidence level line and the linear interpolation of the data.

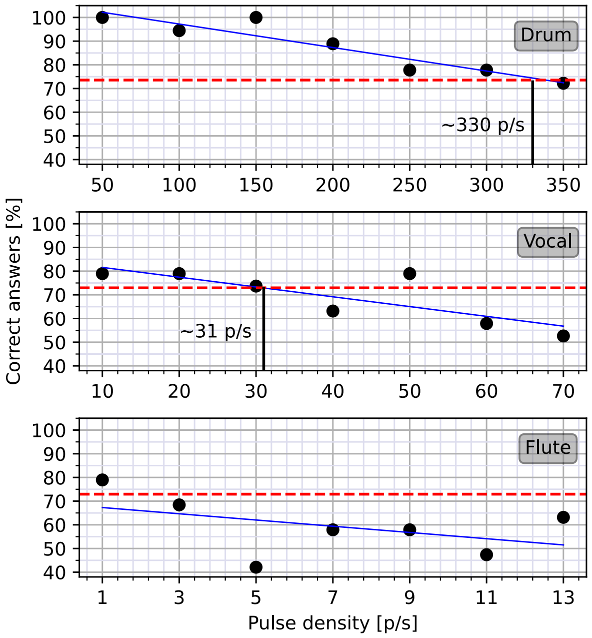


Figure 4: Pulse density thresholds for each stimulus obtained from the conducted test.

As seen in Figure 4, the pulse density threshold for the drum stimulus was found to be around 330 p/s. In the case of the vocal stimulus, the pulse density threshold was estimated to be around 31 p/s. On the other hand, it was not possible to deduct a valid threshold from the obtained data of the flute stimulus.

$\chi^2$-test can also be used to determine the necessary percentage of correct answers for a 95% confidential level of significant difference. Tables 1, 2 and 3 show the result of the $\chi^2$-test and the p-value for every comparison of each stimulus, rounded to three decimals.

Table 1: $\chi^2$ and p-value of the percentage of correct answers corresponding to the drum stimulus.

| Pulse density | $\chi^2$ | p-value |
|---|---|---|
| 50 [p/s] | 0.000 | 0.000 |
| 100 [p/s] | 0.000 | 0.000 |
| 150 [p/s] | 0.000 | 0.000 |
| 200 [p/s] | 0.001 | 0.001 |
| 250 [p/s] | 0.018 | 0.015 |
| 300 [p/s] | 0.018 | 0.015 |
| 350 [p/s] | **0.059** | 0.048 |

Table 2: $\chi^2$ and p-value of the percentage of correct answers corresponding to the vocal stimulus.

| Pulse density | $\chi^2$ | p-value |
|---|---|---|
| 10 [p/s] | 0.012 | 0.010 |
| 20 [p/s] | 0.012 | 0.010 |
| 30 [p/s] | 0.039 | 0.032 |
| 40 [p/s] | **0.251** | **0.180** |
| 50 [p/s] | 0.012 | 0.010 |
| 60 [p/s] | **0.491** | **0.324** |
| 70 [p/s] | **0.819** | **0.500** |

Table 3: $\chi^2$ and p-value of the percentage of correct answers corresponding to the flute stimulus.

| Pulse density | $\chi^2$ | p-value |
|---|---|---|
| 1 [p/s] | 0.012 | 0.010 |
| 3 [p/s] | **0.108** | **0.084** |
| 5 [p/s] | **0.491** | **0.820** |
| 7 [p/s] | **0.491** | **0.324** |
| 9 [p/s] | **0.491** | **0.324** |
| 11 [p/s] | **0.819** | **0.676** |
| 13 [p/s] | **0.251** | **0.180** |

These tables reflect roughly the same information shown in Figure 4, where each value above 0.05 (highlighted in bold) indicates audible difference. The same interpretation can be made with the black dots above the red dashed line in Figure 4.

## 4. DISCUSSION

The pulse density thresholds found in this study are significantly lower than those found in previous studies that directly assessed perception of the velvet noise itself (i.e., without being applied as an artificial reverberation method). Moreover, considering studies that did evaluate velvet noise applied to artificial reverberation, the thresholds found in this work were lower because the characteristics of the music fragment were taken into account.

An important factor in the success of the conducted test is the design of an evenly spaced scale of increasing pulse densities of velvet noise. It is not possible to determine a threshold if it is not confined in between the limits of the designed scale.

Looking at Figure 4, it can be observed that the designed scale almost fails to contain the estimated pulse density threshold for the drum stimulus. For example, a scale between 150 and 450 p/s would have been more appropriate. However, the threshold obtained for the drum stimulus is still considered valid, and no systematic errors were detected.

Moving to the vocal stimulus, the determined threshold is more centered in the pulse density scale. The obtained results show that there is no audible difference between GWN and velvet noise when the pulse density is greater than 31 p/s. However, this result is controversial, since the comparison between the 50 p/s velvet noise and the GWN shows a confidence level above 95%. This inconsistency in the subjects' responses can be attributed to a systematic error related to the test methodology. It was later discovered that, when instantly switching between specific audio tracks, a click sound could be occasionally detected. But if the audio tracks were the same, the click sound didn't exist. This could lead the listeners to choose the correct answer based on a technical artifact, instead of following their subjective perception. For this reason, the obtained threshold is considered valid, taking into account that if the test was conducted again without this specific systematic error, the new threshold could be slightly lower.

Finally, there is no threshold obtained related to the flute stimulus. This is because the linear interpolation of the percentage of correct answers doesn't intercept the 95% confidence level in a valid pulse density value. It can be observed that this intersection happens under 0 p/s. This observation is more related to the characteristics of the music excerpt rather than the test itself. Another discussion is whether it makes sense to consider a handful of impulses in the span of one second as a room impulse response.

For these reasons, it can be hypothesized that there are no subjective differences for any fragment of music with a $\tau_{e(min)}$ greater than 108 ms. If a future investigation pretends to find the relationship between $\tau_{e(min)}$ and the pulse density threshold, it is recommended to study fragments with a minimum effective duration of less than 108 ms.

## 5. CONCLUSIONS

This investigation provides evidence that the required pulse density for the velvet noise to be perceived as subjectively equal to Gaussian white noise significantly decreases depending on the characteristics of the music excerpt to which the synthetic reverberation is applied.

Decreasing the pulse density of velvet noise could potentially lead to better performing artificial reverberation algorithms. Understanding the cases where this can be done can avoid the perception of subjective differences.

Systematic errors may have interfered with the results, but it is considered that they did not affect the results to any great extent.

This study can serve as a starting point for future work to characterize in more detail the relationship between $\tau_{e(min)}$ and the pulse density threshold.

## REFERENCES

*Abx - listening tests as a service*. (n.d.). Retrieved June 20, 2024, from https://abxtests.com/

Alary, B., Politis, A., & Välimäki, V. (2017). Velvet-noise decorrelator. *International Conference on Digital Audio Effects*, 405–411.

Järveläinen, H., & Karjalainen, M. (2007). Reverberation modeling using velvet noise. *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*.

Kawahara, H., Sakakibara, K. I., Morise, M., Banno, H., Toda, T., & Irino, T. (2018). Frequency domain variants of velvet noise and their application to speech processing and synthesis. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2018*, 2027–2031.

Kawahara, H., Sakakibara, K.-I., Mizumachi, M., Banno, H., Morise, M., & Irino, T. (2019). Frequency domain variant of velvet noise and its application to acoustic measurements. *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1523–1532.

Moorer, J. A. (1979). About this reverberation business. *Computer music journal*, 13–28.

Roberts, J., Fagerström, J., Schlecht, S., & Välimäki, V. (2023). How smooth do you think i am: An analysis on the frequency-dependent temporal roughness of velvet noise. *International Conference on Digital Audio Effects*, 312–318.

Rubak, P., & Johansen, L. G. (1998). Artificial reverberation based on a pseudo-random impulse response: Part i. *Audio Engineering Society Convention 104*.

Sato. (2014). Matlab program for calculating the parameters of the autocorrelation and interaural cross-correlation functions based on ando's auditory-brain model. *Audio Engineering Society Convention 137*.

Sato & Wu. (2010). Definition of the effective duration ($\tau$e) of the running autocorrelation function of music signals. *Acta Acustica*.

Schlecht, S. J., Alary, B., Välimäki, V., Habets, E. A., et al. (2018). Optimized velvet-noise decorrelator. *Proc. Int. Conf. Digital Audio Effects (DAFx-18), Aveiro, Portugal*, 87–94.

Schroeder, M. R. (1961). Natural sounding artificial reverberation. *Audio Engineering Society Convention 13*.

Valimaki, V., Parker, J. D., Savioja, L., Smith, J. O., & Abel, J. S. (2012). Fifty years of artificial reverberation. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(5), 1421–1448.

Välimäki, V., Lehtonen, H.-M., & Takanen, M. (2013). A perceptual study on velvet noise and its variants at different pulse densities. *IEEE Transactions* *on Audio, Speech, and Language Processing*, *21*(7), 1481–1488.