Fagerström, Jon; Schlecht, Sebastian; Välimäki, Vesa

One-to-Many Conversion for Percussive Samples

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

Published: 08/09/2021

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

ONE-TO-MANY CONVERSION FOR PERCUSSIVE SAMPLES

Jon Fagerström[1], *Sebastian J. Schlecht*[1,2] *and Vesa Välimäki*[1]

[1]Acoustics Lab, Dept. of Signal Processing and Acoustics
[2]Media Lab, Dept. of Media
Aalto University
Espoo, Finland
jon.fagerstrom@aalto.fi

ABSTRACT

A filtering algorithm for generating subtle random variations in sampled sounds is proposed. Using only one recording for impact sound effects or drum machine sounds results in unrealistic repetitiveness during consecutive playback. This paper studies spectral variations in repeated knocking sounds and in three drum sounds: a hihat, a snare, and a tomtom. The proposed method uses a short pseudo-random velvet-noise filter and a low-shelf filter to produce timbral variations targeted at appropriate spectral regions, yielding potentially an endless number of new realistic versions of a single percussive sampled sound. The realism of the resulting processed sounds is studied in a listening test. The results show that the sound quality obtained with the proposed algorithm is at least as good as that of a previous method while using 77% fewer computational operations. The algorithm is widely applicable to computer-generated music and game audio.

1. INTRODUCTION

Sampling is a widely used sound synthesis technique in music synthesizers and drum machines as well as in sound effects for video games [1, 2, 3, 4]. However, repeatedly using a single sample results in a mechanical output, often called the "machine-gun effect" [5]. An endless variety of signal-processing techniques can be used to dramatically modify samples, such as filtering and temporal envelope shaping [1, 6]. However, natural and subtle variations between consecutive sounds are needed to achieve realism, and these are difficult to generate.

In computer games, the desired variations are often achieved by storing multiple recordings of each sound, such as gunshots or footsteps, and using round-robin for playback [7]. Morphing techniques have been proposed to synthesize new sounds with subtle differences, when at least two example recordings are available [8, 9]. The cross synthesis of different samples using linear prediction is another possibility [10]. However, these techniques are limited to variations between two or more samples, requiring more memory than a single sample and not necessarily yielding the full range of realistic alterations.

Linear prediction has been applied to produce different versions of a sampled sound by using the same prediction filter and replacing the excitation signal every time with a different random sequence [11]. However, high filter orders, like $N = 1000$, are needed for accurate spectral details. Modal synthesis techniques can also be used for creating random variations in impact sounds [12] or variations based on physical properties of the resonant object and the exciting force [13, 14]. However, the latter two do not run in real time. Recently, various machine-learning methods have also been applied to synthesize impact sounds [15, 16]. A deep-learning architecture was proposed for efficient real-time modal synthesis of impact sounds [15]. Another approach uses conditional WaveGAN for synthesizing knocking sounds [16].

Lloyd et al. suggested the use of a multi-notch filter with random adjustment to produce plausible variations [12]. To obtain good results, they recommended using ten biquad filters. The center frequencies were distributed uniformly on the logarithmic frequency axis, and the filter gains and Q-values randomized. However, appropriate ranges for these random filter parameters were not discussed in detail.

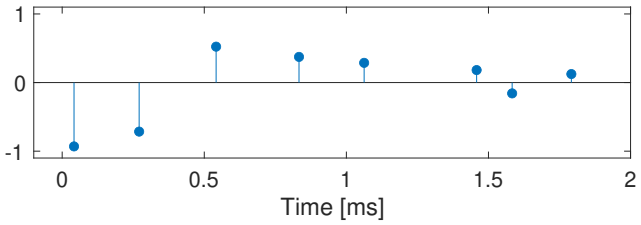
This paper introduces a new solution to the problem in which only one sample is available, and with the goal to produce many realistic variations efficiently at playback time. We refer to this as the small-data problem [17], as opposed to the more common big-data problem, which leads to very different challenges. We propose a signal-processing method, that applies a short and sparse velvet-noise filter (VNF) to the source sample and further filters it using a shelf filter. The VNFs and the shelf filter are designed so that the resulting spectral changes convey natural variations, which are learned from a set of recordings in this work.

Velvet noise was originally proposed as an efficient and perceptually smooth alternative to white noise for modeling late reverberation [18]. Velvet noise can also be used as the input signal for subtractive synthesis of stationary sounds [17, 19], but the same approach is not directly applicable to percussive samples. Alary et al. proposed the use of short VNFs as efficient decorrelation filters [20]. These filters were shown to produce some spectral coloration, which was subsequently reduced by an optimization scheme [21]. In this work, we exploit the spectral coloration effect of VNFs.

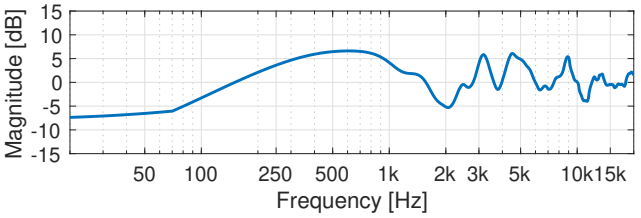
This paper is organized as follows. Section 2 recapitulates the basics of VNFs. Section 3 introduces the new algorithm for one-to-many mapping. Section 4 discusses the calibration of the algorithm parameters for four test signals. Section 5 validates the proposed method by analyzing the spectrum of the chosen signals, by assessing them in a listening test, and by comparing the computational cost with a previous method. Section 6 concludes the paper.

2. SPARSE FILTERING BASED ON VELVET NOISE

Velvet noise is a sparse pseudo-random signal, comparable to white noise, having a small percentage of non-zero samples [18]. By



(a) Time-domain response



(b) Magnitude response

Figure 1: *(a) Impulse response of an exponentially decaying VNF ($L_s = 2$ ms) and (b) its Bark-smoothed magnitude response.*

taking advantage of the sparsity of the VNFs, computing its time-domain convolution with another signal becomes very efficient [18, 22]. Conceptually, the first step in generating velvet noise is to create a sequence of evenly spaced impulses with a desired density [22]. The sign and location of each impulse are then randomized, but impulses still remain within a given interval, having a range dictated by the desired impulse density [18]. Fig. 1 shows an example of a VNF consisting of eight non-zero samples and its magnitude response.

For a given density ρ and sampling rate f_s, the average spacing between two neighboring impulses in a VNF is

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

which is called the grid size [23]. The VNF consists of M impulses, where the sign of each impulse is

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

where $m = 0, 1, 2, ..., M-1$ is the impulse index, $\lfloor \cdot \rceil$ is the rounding operation to the nearest integer, and $r_1(m)$ is a uniformly distributed random number between 0 and 1, i.e, $r_1(m) \sim \mathcal{U}(0,1)$ [23]. The length of the filter is then $L_\mathrm{s} = MT_\mathrm{d}$. The mth impulse of the sequence is located at

$$k(m) = \lfloor mT_\mathrm{d} + r_2(m)(T_\mathrm{d} - 1) \rceil, \tag{3}$$

where $r_2(m) \sim \mathcal{U}(0,1)$ [23].

To prevent the smearing of transients, consecutive pulses are attenuated exponentially, i.e.,

$$s_\mathrm{e}(m) = e^{-\alpha m} s(m)\, r_3(m), \tag{4}$$

where $\alpha > 0$ is the decay rate and $r_3(m) \sim \mathcal{U}(0.5, 2)$ allows some extra variation [20, 21]. The VNF $v(n)$ is then

$$v(n) = \begin{cases} s_\mathrm{e}(m) & \text{for } k(m) = n, \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

and the total energy decay is

$$L_\mathrm{dB} = 20 \log_{10} e^{-\alpha M}. \tag{6}$$

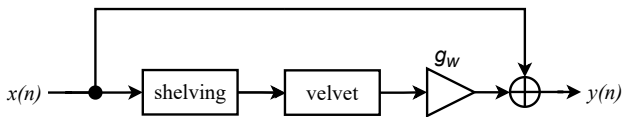
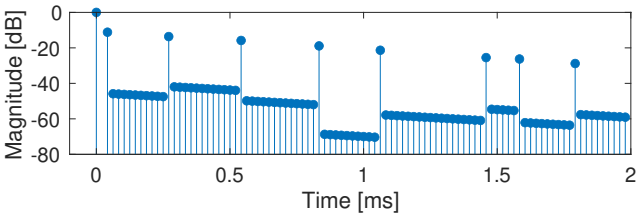
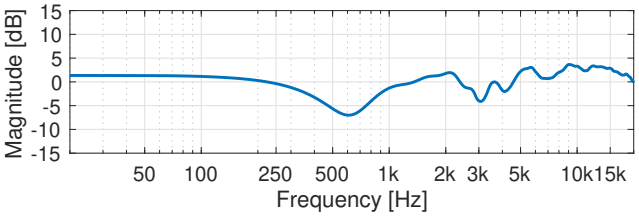


Figure 2: *Block diagram of the proposed variation filter structure. The shelving block consists of a single first-order low shelf.*



(a) Time-domain response



(b) Magnitude response

Figure 3: *(a) Example impulse response of the proposed variation filter and (b) its Bark-smoothed magnitude response. The VNF here is the one that was also presented in Fig. 1.*

The sparsity of the VNF $v(n)$ allows an efficient time-domain convolution with a signal $x(n)$ [20], i.e.,

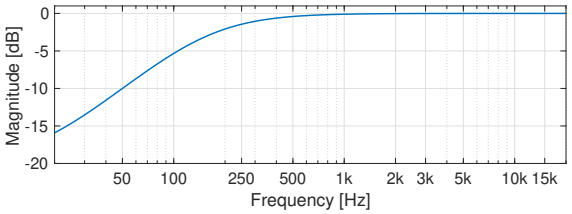
$$(x * v)(n) = \sum_{m=0}^{M-1} x(n - k(m)) s_\mathrm{e}(m), \tag{7}$$

where $*$ denotes the discrete convolution.

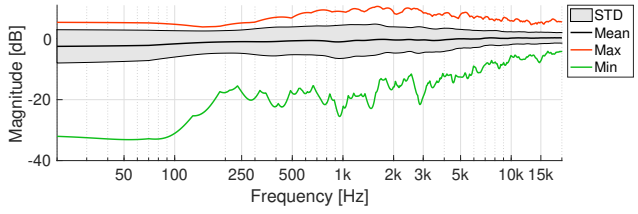
3. PROPOSED VARIATION FILTER

The block diagram of the proposed variation filter structure, shown in Fig. 2, consists of a direct and a filtered path. The filtered path contains a first-order low shelf [24] followed by the VNF. Additionally, the filtered path has a gain g_w to control the magnitude of variations introduced at the output. An example impulse response (IR) of the proposed variation filter is shown in Fig. 3a. The effect of the direct path is seen in the IR at time zero and that of the low-shelf filtered response of the VNF follows shortly after. The corresponding Bark-smoothed magnitude response is shown in Fig. 3b.

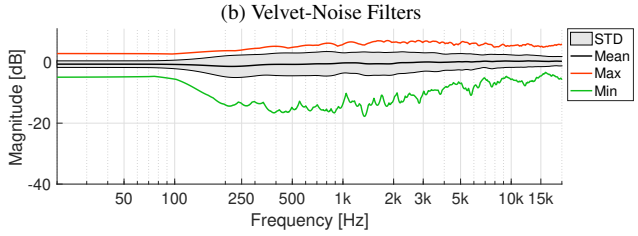
The core of the variation filter structure is the VNF that generates random coloration in the input sample. The VNF is a sparse FIR filter parametrized by its length L_s in ms, the number of impulses M, and the total target decay L_dB in dB. For each unique output sample a new random VNF filter is needed. The purpose of the low-shelf filter is to target the generated variations at higher frequencies. The effect of the low-shelf filter is further explained by Fig. 4.



(a) Low-Shelf Filter



(b) Velvet-Noise Filters



(c) Proposed Filter

Figure 4: *Magnitude response of (a) the low-shelf filter, (b) VNF filters, and (c) the proposed variation filter. Statistical values in (b) and (c) are computed from 500 Bark-smoothed responses.*

Figure 4a shows an example magnitude response of the low-shelf filter with $F_c = 50$ Hz and $G_{lo} = -20$ dB. The magnitude response of a VNF is shown in Fig. 4b. The response is analyzed from 500 random 3-ms long VNFs with $M = 8$ impulses. Finally, Fig. 4c shows the spread of the magnitude responses of the complete filter structure. Figure 4b shows that the VNF produces a wide spread of variations, especially at the low frequencies. Introducing the direct path and adding the low-shelf filter on the filtered path reduces the low-frequency variations. Thus, large variations are only produced on the passband of the low-shelf filter, as shown in Fig. 4c.

4. TEST SAMPLES AND FILTER PARAMETERS

The proposed variation filter was tested on four different sound types: hihat, snare-drum, tomtom, and door knock sounds. A set of 30 recordings of each sound type were used as the ground truth of the real variations. The recordings were captured by hitting each object multiple times, concentrating on using consistent force and location of each individual stroke on the drum. All the drum sounds (hihat, snare, tomtom) were performed by a semi-professional drummer with over 20 years of practise on the instrument, whereas the door-knocking sounds were performed by an amateur percussionist.

Table 1 shows the filter parameters used in this study. These parameters were calibrated with the help of a statistical spectral analysis, and by listening to the generated samples within each sound type and comparing them to the ground truth.

The proposed filtering algorithm was compared against a version of a previous random filter method [12]. Instead of the randomized center frequencies we opted to use a graphical equalizer (GEQ) consisting of ten biquad sections [24] and randomized only the filter gains. By using fixed center frequencies, we ensured each frequency area would always have only a single filter. The filter gains were limited to the maximum of the spectral analysis of the ground truth at each center frequency. Only negative gains were used, as suggested by Lloyd et al. [12].

Table 1: *Proposed filter parameters for the four test sounds.*

Sound	F_c	G_{lo}	L_s	M	L_{dB}	g_w
Hihat	50 Hz	−20 dB	30 ms	8	20 dB	0.5
Snare	100 Hz	−5 dB	4 ms	8	20 dB	0.2
Tomtom	75 Hz	−6 dB	2 ms	8	20 dB	0.25
Knock	50 Hz	−25 dB	2 ms	8	20 dB	0.4

The capability of the proposed one-to-many mapping was further tested using various unique samples. These sound examples, generated with the proposed method, are available online[1] using the web audio player from [25].

5. RESULTS

In this section, both objective and subjective test results are presented to assess the sound quality of the proposed and previous methods. Additionally, the computational cost is discussed.

5.1. Spectral Analysis

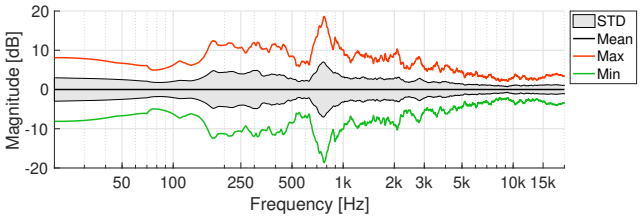
The spectral variations generated by the proposed filtering structure and the previous random GEQ filter were compared against the variations present in real recordings of each test sound type described in Sec. 4. The spectral analysis was conducted by investigating the pairwise magnitude spectrum differences of 30 samples resulting in 870 permutations. Bark-smoothing was used on the spectra to approximate the frequency resolution of human hearing.

Figure 5a shows the actual variations retrieved from 30 recorded hihat sounds. Figures 5b and 5c show the same analysis of 30 filtered samples generated with the proposed and previous methods, respectively. The source sample for each filtered signal was the first sample in the recorded set.

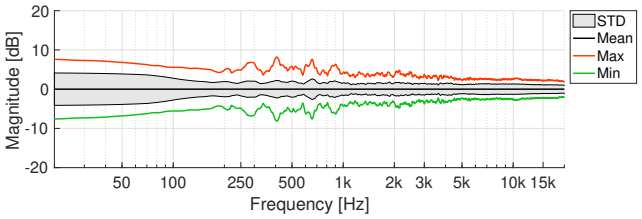
As seen in Fig. 5c, the GEQ filter approximates the overall spread of variations in different frequency areas. However, it fails to capture some of the details, e.g., the strong resonance around 800 Hz. On the other hand, the proposed filter in Fig. 5b models also some of the finer details, as seen from the jagged edge of the variations. However, in the case of the hihat sound, we opted to use longer 30-ms VNFs, that produce less variations than the shorter VNFs, and fail to reach the level of variation seen in Fig. 5a. Additionally, the longer 30-ms VNF results in some temporal smearing. We suspect that the more salient differences between the hihat recordings are in the time domain and were better modeled with the longer 30-ms VNFs.

Figure 6 shows the same spectral analysis conducted on the snare-drum sounds. Variations between the recorded set of 30 snare strokes are shown in Fig. 6a. The 4-ms long VNFs used for the snare drums are relatively short compared to the ones used

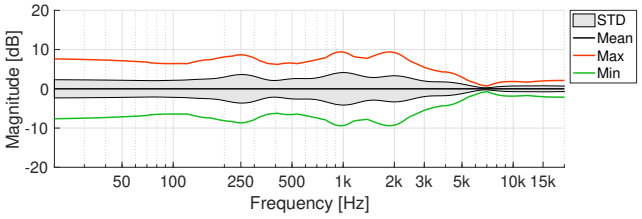
[1] `http://research.spa.aalto.fi/publications/papers/dafx21-one2many/`



(a) Actual

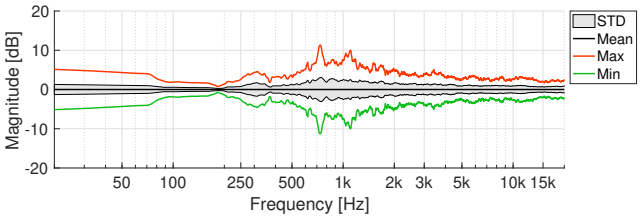


(b) Proposed

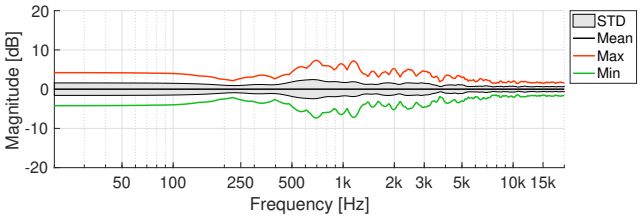


(c) GEQ

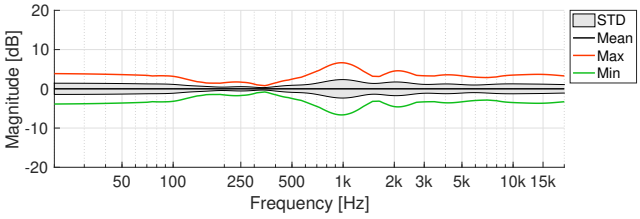
Figure 5: *Statistical analysis of pairwise magnitude-spectrum differences in 30 hihat sounds. A comparison between (a) the unique recordings, (b) the samples filtered with the proposed method, and (c) the samples filtered with the random-GEQ method. Each spectrum is Bark-smoothed.*



(a) Actual



(b) Proposed



(c) GEQ

Figure 6: *Statistical spectral analysis of the (a) recorded and (b), (c) processed snare-drum sounds, cf. Fig. 5.*

Table 2: *Computational operations count of the proposed method compared with the GEQ based on ten second-order sections.*

Method	ADD	MUL	Total	Saving
GEQ	40	50	90	Reference
Proposed	10	11	21	77%

with the hihat sound. This results in smoother variations and does not produce the ragged detail, as seen in Fig. 6b. Overall, the magnitude of the variations is matched well with the actual recordings. Again, the GEQ filter in Fig. 6c behaves similarly as with the hihat sound, capturing the overall shape of the variations, while missing some of the narrow resonant details.

The variations in the tomtom sounds are shown in Fig. 7a. Whereas the actual variations in the hihat and snare sounds were concentrated at the middle frequencies, with the tomtom sound there are large variations also at higher frequencies. The GEQ filter models the overall shape of variations accurately, as seen in Fig. 7c, but again lacks the finer details of the actual variations. On the other hand, the proposed filter in Fig. 7b has some more detail but does not produce as much variations above 6 kHz as in the ground truth in Fig. 7a.

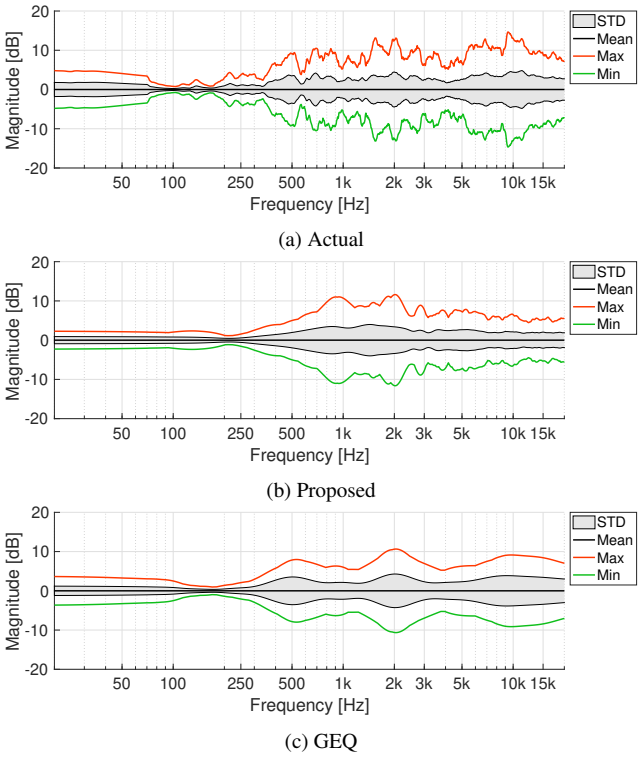
Finally, Fig. 8a shows the variations between 30 door-knock recordings. These sounds have the largest variations of all the sound types with maximum variations reaching up to 20 dB. The larger variance may be explained by the more complex excitation produced by multiple knuckles hitting the door as compared to the drum sounds with the tip of the stick as the exciting object. Another explanation may be that the amateur performing the door knocking was less meticulous than the semi-professional drummer who played the drum sounds. Figure 8b shows the variations generated by the proposed filter structure. It produces smaller variations at the higher frequencies as compared to the ground truth, but otherwise matches the actual variations well. The GEQ filter in Fig. 8c performs similarly as with all the other sound types, by capturing the smoothed shape of the spectrum while failing to produce finer details.

5.2. Computational Cost

The numbers of operations per output sample for the compared methods are shown in Table 2. Here, the GEQ corresponds to the previous random-notch-filter method with ten biquad filters [12]. Each biquad uses four adders and five multipliers. Our proposed method uses eight adders and multipliers for the VNF computation and two additional adders and three multipliers for the low-shelf filter. A saving of 77% in the total number of operations is achieved with the proposed method compared to the previous technique.

5.3. Perceptual Test

A Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test was conducted to assess the realism of the generated samples. In total, 21 participants took the test. All of the

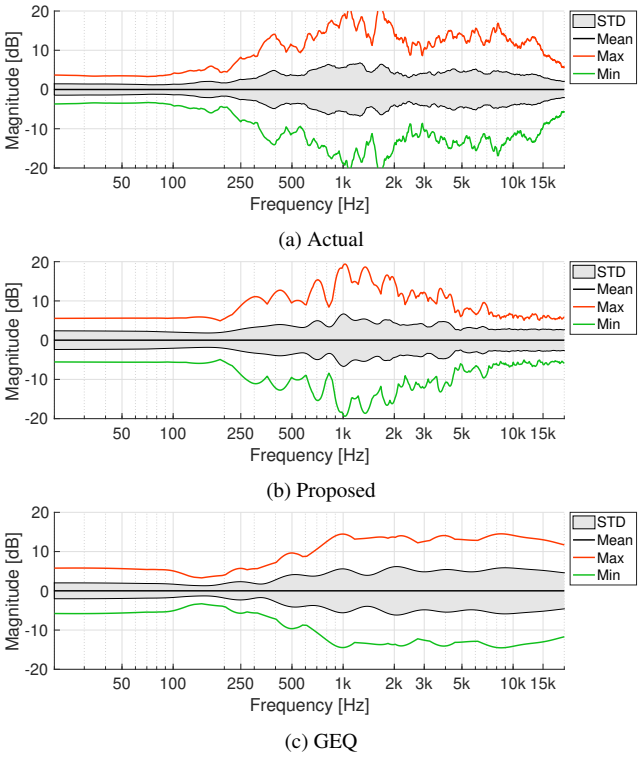


(a) Actual

(b) Proposed

(c) GEQ

Figure 7: *Statistical spectral analysis of the (a) recorded and (b), (c) processed tomtom sounds, cf. Fig. 5.*



(a) Actual

(b) Proposed

(c) GEQ

Figure 8: *Statistical spectral analysis of the (a) recorded and (b), (c) processed knocking sounds, cf. Fig. 5.*

participants reported having normal hearing, and 20 of them had participated in a formal listening test before. The mean age of the participants was 28.3 years with a standard deviation of 2.87. The test was implemented using webMUSHRA [27] and conducted in sound-proofed listening booths at the Aalto Acoustics Lab using Sennheiser HD-650 headphones. The task in the test was to assess the realism of the timbral variations within a single stimulus compared to the variations within the reference item. Each test item was a time-quantized sequence of 20 samples. The total energy of each sample was normalized. The temporal quantization and normalization were applied to eliminate any timing and loudness variations, and to focus the study purely on the timbral variations.

The test included one training page and three experiment pages for each of the tested sound types. The subjects were allowed to adjust the sound level suitably during the training phase. Each experiment contained five stimuli and the hidden reference, which are listed in Table 3 with short descriptions. The order of the experiments as well as the order of each item was randomized for each participant. With 21 participants, 63 data points per test item were gathered. For the reference item (Ref1), we selected a set of 20 recorded samples, and for the anchor, a loop of the first sample in Ref1. The anchor was assumed to receive a low score as it contained no variations. The stimuli included also a second item having real recordings (Ref2). Ref2 was obtained from Ref1 by shuffling the order of the samples. This item was added as a sanity check to find out whether a shuffled set of real recordings would be rated as realistic.

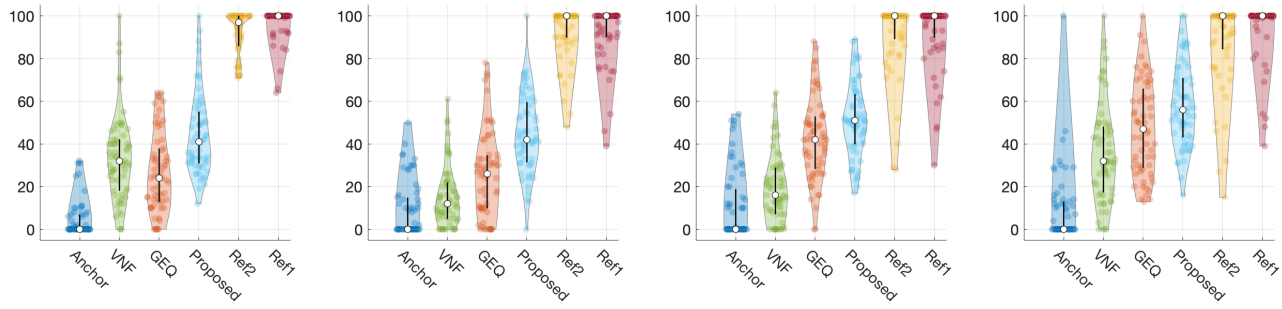
The remaining three stimuli in the listening test were all sequences of sounds filtered with the different methods, as summarized in Table 3. The first recording of Ref1 was used as the source sample for each filtering method. Test item "proposed" corresponded to the proposed variation filter using the parameters given in Table 1 for each sound type, whereas "VNF" was created with a naive VNF without the low-shelf filter or the direct path. Finally, "GEQ" corresponded to the previous method using a randomized GEQ [12].

Table 3: *Items included in the listening test.*

Test item	Description
Ref1	Reference sequence of 20 unique recordings
Ref2	Random permutation of Ref1
Anchor	First sample of Ref1 repeated 20 times
VNF	20 sounds filtered with naive VNF
GEQ	20 sounds filtered with randomized GEQ [12]
Proposed	20 sounds filtered with proposed method

The rating scale was from 0–100 with five text labels, as seen in Fig. 10. At the top, corresponding to a scoring range 80–100 was the label "realistic" followed by "almost realistic" (60–80), "slightly realistic" (40–60), "unrealistic" (20–40), and, finally, at the bottom "very unrealistic" (0–20).

Figure 9 shows the perceptual test results. The hidden reference (Ref1) received a median score of 100 with all tested sound types with a 75% confidence interval within the "realistic" label. Ref2 achieved similar results as Ref1 but with a larger variance. This indicates that distinguishing Ref1 and Ref2 was quite challenging. Additionally, many participants reported finding two items very close to the reference in most of the experiments. The



(a) Hihat (b) Snare (c) Tomtom (d) Door knock

Figure 9: *Perceptual test results for the (a) hihat, (b) snare, (c) tomtom, and (d) door-knock sounds, shown using the violinplot function [26]. The boxplot shape is included as a black line in the center of the violin. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The violin outline shows the kernel density estimation and is overlaid with the data points.*

anchor received a median score of 0 with each sound type with the 75% confidence interval within the "very unrealistic" label, as expected.

Of the compared filtering methods, the proposed method achieved the highest median score, which was within the "slightly realistic" label with each of the sound types. It was rated clearly as more realistic than the anchor with each sound type. However, it was not shown to be statistically better than the previous GEQ method. The GEQ method was the second-best filtering method, based on the median score with all sounds, except in the case of the hihat, where the naive VNF received a higher median score.

The door-knock sound received the widest distribution of scores. It might have been harder to rate the door-knock sounds than the musical sounds of the hihat, snare, and tomtom. The latter group of sounds is commonly heard in music, so there might have been a better consensus on their realism, whereas door-knock sounds might not be listened to with as much focus on detail in everyday life.

Overall, the listening test method itself could be improved. Especially the difficulty of distinguishing between Ref1 and Ref2 is problematic in a MUSHRA test. Moreover, the relatively wide distribution of scores, even for the hidden reference, suggests that the task was very challenging, and there was no clear consensus among test subjects of what is considered a realistic variation in percussive sounds. Finally, a direct comparison against the reference in a MUSRHA test would require the filtered variations to match precisely those of the reference signal to achieve a high score, as indicated by the results of test items Ref1 and Ref2.

6. CONCLUSIONS

A novel algorithm for creating realistic variations of a single percussive source sample was proposed. The new method employs a short and sparse VNF for creating random variations and a first-order low-shelf filter for targeting the variations in the desired frequency range. A scaled version of this processed signal is added to the original sample to change it. The introduced variations are modeled based on a statistical analysis of multiple recordings of the same sound type.

The variations in the magnitude spectrum in real recordings of four sound types— hihat, snare, tomtom, and door knocking

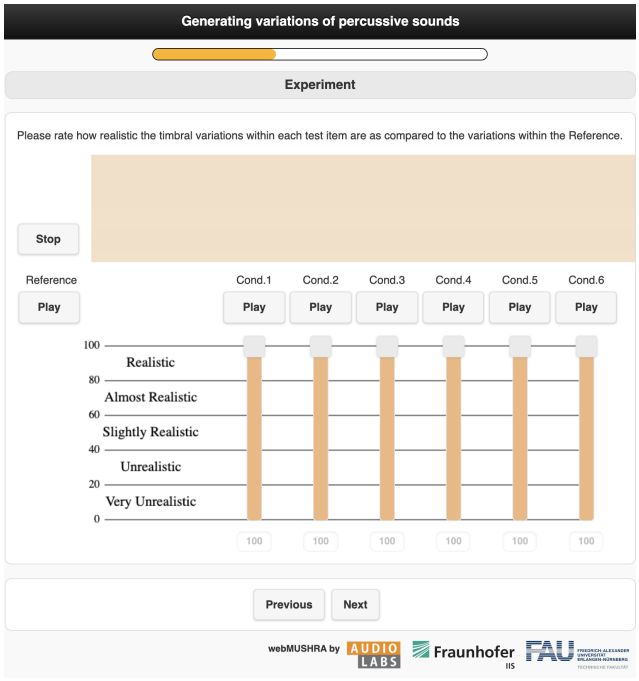


Figure 10: *Graphical User Interface (GUI) of the perceptual test.*

sounds—were analyzed and used as the ground truth. The variations generated with the proposed method and a previous random-equalization method were compared against the ground truth objectively. The proposed method was shown to produce detailed variations with each test sound. However, the previous method matched the range of variations better with some test sounds. A perceptual test was conducted to assess the subjective quality of the generated sounds. The proposed method was shown to produce variations that improved the realism over the repetitive playback of a single sample. Furthermore, the proposed method outperformed the previous method marginally in the listening test while using 77% fewer computational operations.

The proposed method may be used to humanize sampling syn-

thesis of drum and contact sounds, for example. As the algorithm has only a few parameters, it is easy to adjust for use with many different sounds for which only a single recording is available. Further research is needed to understand better the effect of the filter parameters on the perceived variations and to improve the overall audio quality to better match the ground truth perceptually. The method is widely applicable in music, gaming, and virtual reality.

7. ACKNOWLEDGMENTS

This work was supported by the "Nordic Sound and Music Computing Network—NordicSMC", NordForsk project number 86892. Thank you to Leevi Luoto for performing and recording the drum samples and to Janis Heldmann and Juho Liski for helping to set up the MUSHRA test.

8. REFERENCES

[1] J. O. Smith, "Viewpoints on the history of digital synthesis," in *Proc. Int. Computer Music Conf.*, Montreal, Canada, Oct. 1991.

[2] P. R. Cook, *Real Sound Synthesis for Interactive Applications*, AK Peters, Ltd., 2002.

[3] W. Kim and J. Nam, "Drum sample retrieval from mixed audio via a joint embedding space of mixed and single audio samples," in *Proc. Audio Eng. Soc. 149th Conv.*, Oct. 2020.

[4] J. Shier, K. McNally, G. Tzanetakis, and K. G. Brooks, "Manifold learning methods for visualization and browsing of drum machine samples," *J. Audio Eng. Soc.*, vol. 69, no. 1/2, pp. 40–53, Jan. 2021.

[5] K. J. Werner, J. S. Abel, and J. O. Smith III, "A physically-informed, circuit-bendable, digital model of the Roland TR-808 bass drum circuit," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, Erlangen, Germany, Sept. 2014, pp. 159–166.

[6] M. Ilmoniemi, V. Välimäki, and M. Huotilainen, "Subjective evaluation of musical instrument timbre modifications," in *Proc. Baltic-Nordic Acoustics Meeting (BNAM)*, Mariehamn, Aland, June 2004.

[7] B. Schmidt, "Interactive mixing of game audio," in *Proc. Audio Eng. Soc. 115th Conv.*, New York, NY, USA, Oct. 2003, paper 5857.

[8] W. Ahmad, H. Hacihabiboglu, and A. M. Kondoz, "Morphing of transient sounds based on shift-invariant discrete wavelet transform and singular value decomposition," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 297–300.

[9] S. Siddiq, "Real-time morphing of impact sounds," in *Proc. Audio Eng. Soc. 139th Conv.*, New York, NY, USA, Oct. 2015, paper 9407.

[10] J. Wikström, "Improved sound sampling," M.S. thesis, Aalto University, Espoo, Finland, Nov. 2017.

[11] V. Mäntyniemi, R. Mignot, and V. Välimäki, "REMES final report: Realistic machine and environmental sounds for a training simulator to improve safety at work," Aalto Univ. publication series SCIENCE + TECHNOLOGY 16/2014.

[12] D. B. Lloyd, N. Raghuvanshi, and N. K. Govindaraju, "Sound synthesis for impact sounds in video games," in *Proc. Symp. Interactive 3D Graphics and Games (I3D)*, San Francisco, CA, USA, Feb. 2011, pp. 55–62.

[13] C. Zheng and D. L. James, "Toward high-quality modal contact sound," *ACM Trans. Graphics*, vol. 30, no. 4, pp. 1–12, July 2011.

[14] J. Traer, M. Cusimano, and J. H. McDermott, "A perceptually inspired generative model of rigid-body contact sounds," in *Proc. 21st Int. Conf. Digital Audio Effects (DAFx-19)*, Birmingham, UK, Sep. 2019, pp. 136–143.

[15] X. Jin, S. Li, T. Qu, D. Manocha, and G. Wang, "Deep-modal: Real-time impact sound synthesis for arbitrary shapes," in *Proc. 28th ACM Int. Conf. Multimedia (MM'20)*, Seattle, WA, USA, Oct. 2020, pp. 1171–1179.

[16] A. Barahona-Ríos and S. Pauletto, "Synthesising knocking sound effects using conditional WaveGAN," in *Proc. 17th Sound and Music Computing Conf.*, Torino, Italy, Jun. 2020, pp. 450–456.

[17] V. Välimäki, J. Rämö, and F. Esqueda, "Creating endless sounds," in *Proc. 21st Int. Conf. Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, Sep. 2018, pp. 219–226.

[18] M. Karjalainen and H. Järveläinen, "Reverberation modeling using velvet noise," in *Proc. Audio Eng. Soc. 30th Int. Conf. Intell. Audio Environ.*, Saariselkä, Finland, Mar. 2007.

[19] S. D'Angelo and L. Gabrielli, "Efficient signal extrapolation by granulation and convolution with velvet noise," in *Proc. 21st Int. Conf. Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, Sep. 2018, pp. 107–112.

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

[21] S. J. Schlecht, B. Alary, V. Välimäki, and E. A. P. Habets, "Optimized velvet-noise decorrelator," in *Proc. Int. Conf. Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, Sept. 2018, pp. 87–94.

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

[23] V. Välimäki, H.-M. Lehtonen, and M. Takanen, "A perceptual study on velvet noise and its variants at different pulse densities," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1481–1488, Jul. 2013.

[24] V. Välimäki and J. Reiss, "All about audio equalization: Solutions and frontiers," *Applied Sciences*, vol. 6, no. 5, May 2016, paper 129.

[25] N. Werner, S. Balke, F.-R. Stöter, M. Müller, and B. Edler, "trackswitch.js: A versatile web-based audio player for presenting scientific results," in *Proc. 3rd Web Audio Conf.*, London, UK, Aug. 2017.

[26] B. Bechtold, "Violin Plots for Matlab, Github Project," Available at https://github.com/bastibe/Violinplot-Matlab, accessed Jun 3, 2021.

[27] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, "WebMUSHRA—A comprehensive framework for web-based listening tests," . *Open Res. Softw.*, vol. 6, no. 1, pp. 1–8, Feb. 2018.