

## FLUTTER ECHO MODELING

Gloria Dal Santo

Institute of Electrical and Micro Engineering,  
Ecole Polytechnique Fédérale de Lausanne EPFL,  
CH-1015 Lausanne, Switzerland  
gloria.dalsanto@epfl.ch

Karolina Prawda\*, Vesa Välimäki

Acoustics Lab,  
Department of Signal Processing and Acoustics,  
Aalto University,  
FI-02150 Espoo, Finland  
{karolina.prawda, vesa.valimaki}@aalto.fi

### ABSTRACT

Flutter echo is a well-known acoustic phenomenon that occurs when sound waves bounce between two parallel reflective surfaces, creating a repetitive sound. In this work, we introduce a method to recreate flutter echo as an audio effect. The proposed algorithm is based on a feedback structure utilizing velvet noise that aims to synthesize the fluttery components of a reference room impulse response presenting flutter echo. Among these, the repetition time defines the length of the delay line in a feedback filter. The specific spectral properties of the flutter are obtained with a bandpass attenuation filter and a ripple filter, which enhances the harmonic behavior of the sound. Additional temporal shaping of a velvet-noise filter, which processes the output of the feedback loop, is performed based on the properties of the reference flutter. The comparison between synthetic and measured flutter echo impulse responses shows good agreement in terms of both the repetition time and reverberation time values.

### 1. INTRODUCTION

In room acoustics, flutter echo is considered as an unwanted phenomenon and a source of degradation of speech intelligibility, and is usually prevented with careful application of sound absorbing materials. Similarly, in artificial reverberation, enormous efforts have been made to achieve a smooth-sounding decay, devoid of artifacts that can be attributed to fluttering and beating sounds [1, 2]. In this paper, however, we propose a different approach to deliberately synthesize flutter echo in order to use it as an audio effect. Application of flutter echo as audio effect can be of interest to the music industry, as an unusual reverb effect, and in audio technologies for games and virtual reality for the realistic simulation of narrow hallways, alleys, and rooms that present flutter.

Recently, Halmrast conducted an extensive work in modelling the physics of flutter [3]. He described the characteristic middle-frequency timbre of the flutter echo as the result of a gradual bandpass filtering of the broadband impulsive sound generating the repetitions. In ordinary rooms, the energy of the flutter was always found to decay slowest around 1-2 kHz, as a result of low-frequency damping due to increasing source distance and edge diffraction, and of high-frequency damping caused by atmospheric absorption [3].

\* This work was supported by the “Nordic Sound and Music Computing Network—NordicSMC”, NordForsk project number 86892.

Copyright: © 2022 Gloria Dal Santo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, adaptation, and reproduction in any medium, provided the original author and source are credited.

A repetitive sound resembling flutter echo is a common nuisance to velvet-noise-based artificial reverberation [4, 5]. This artifact was problematic in the early velvet-noise techniques, where a single sequence was reused multiple times. It became particularly noticeable when synthesizing interleaved velvet-noise reverbs with too low of a number of insufficiently long velvet-noise sequences [5]. Although the repetitions are usually hidden with an appropriate choice of parameters in reverberation algorithms, the present work exploits this feature in the flutter echo synthesis.

In this paper, we propose a reverberator structure for the simulation of temporal and spectral characteristics of flutter echo in acoustic spaces. In contrast to [3], where room geometry is used to control the reverberator parameters, we propose a method to synthesize flutter from measurements. The model structure consists of a conventional reverberator based on a delay line with feedback and a convolution with a velvet-noise sequence. The paper also discusses methods of obtaining parameters specific to flutter echo from impulse response measurements.

The remainder of this paper is organized as follows. Section 2 describes the temporal and spectral characteristics of the flutter echo, accompanied by an example from real measurements. Section 3 presents a novel algorithm for the synthesis of fluttery late reverberation based on a delay line with feedback and a velvet-noise sequence acting as sparse finite-impulse-response filter. The proposed algorithm has been tested in comparison with a real measurement and the evaluation of its results are discussed in Section 4. Finally, Section 5 offers some concluding remarks.

### 2. DETERMINING FLUTTER ECHO PARAMETERS

Room impulse responses (RIR) of physical spaces are characterized by an increasing echo density related to the subsequent interactions of sound with objects and surfaces in the environment. After a sufficient period of time characterized by low density of reflections, the arriving echoes may be treated statistically. Consequently the sound field can be considered to be a Gaussian process with a power spectrum indicating the size of the space and the absorption of the materials present in it [6, 7]. This approximation motivated the use of random-noise generators in artificial reverberation algorithms [8, 9, 4]. In the presence of flutter echo, however, the late reverberation is non-diffuse and the approximation is no longer applicable.

To better describe the fluttering in RIRs, relevant time parameters need to be identified together with the energy decay. This section describes the methods to determine the flutter echo parameters from a measured RIR.

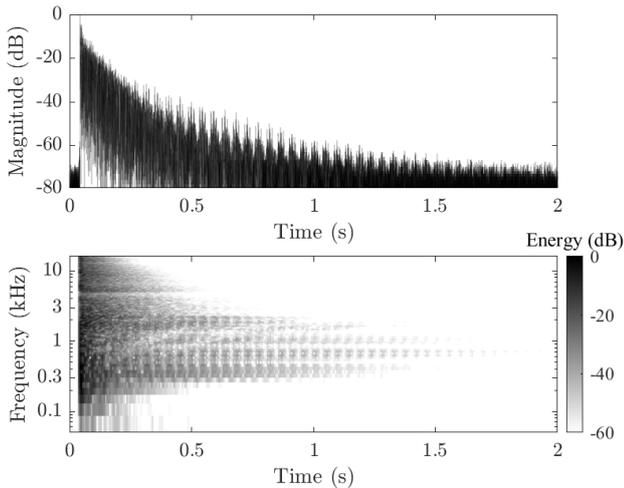


Figure 1: (Top) Measured impulse response of the variable acoustics laboratory Armi with parallel reflective panels and (bottom) its spectrogram, both exhibiting the flutter echo effect.

## 2.1. Repetition Time

The basic principles of sound fluttering between two parallel reflective walls can be straightforwardly described by both geometrical and physical acoustics [10]. Flutter can be excited by an impulsive sound of length  $T$  s, such that  $T$  is much shorter than the time required by a sound wave to travel the distance between the walls. A listener located between the two parallel walls will hear a series of pulses of the original frequencies, each of a duration of  $T$ . The interval between the arrivals of any two pulses will vary as a function of the listener position.

An important parameter of flutter echo is the time that passes between occurrences of flutter, called repetition time, defined as

$$t_r = \frac{2\ell}{c}, \quad (1)$$

where  $\ell$  is the distance between the opposite walls in meters and  $c$  is the speed of sound in meters per second. If the receiver position is in the midpoint between the walls,  $t_r$  reduces to  $\ell/c$  seconds.

Fig. 1 shows an example RIR containing flutter echo. The tail of the late reverberation exhibits a series of pulses that repeats at regular intervals of about  $t_r$  seconds. Similarly, the spectrogram in the bottom pane presents an alternating pattern of high- and low-energy chunks of sound, prominent between about 300 Hz and 2 kHz.

The repetitions produce a perceived tone, referred to as repetition tonality, at the frequency  $f_0 = t_r^{-1} = c/2\ell$ , which coincides with the first axial resonance of the room, along the axis perpendicular to the walls. Accurate estimation and simulation of  $t_r$  are thus of critical importance when reproducing the perceived tonality of flutter.

## 2.2. Autocorrelation Analysis

Traditional pitch- and fundamental-frequency-detection algorithms are based on time-domain autocorrelation. In such methods, the estimated period of the analyzed signal is given by the time difference between the peak at the center of the autocorrelation function

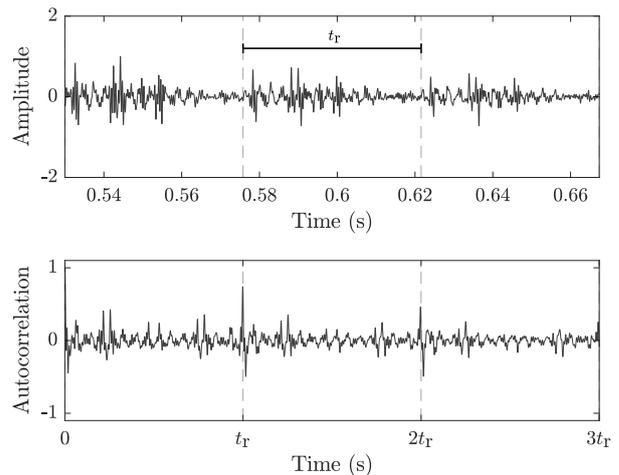


Figure 2: (Top) Normalized section of fluttery late-reverberation of the RIR in Fig. 1 and (bottom) its normalized autocorrelation function.

and its nearest highest peak. For the estimation of the repetition time, we use the autocorrelation function of a short window of the late reverberation tail, after processing the signal with a bandpass filter to isolate the fluttery region.

We obtained good results with a window of 10 times the expected repetition time, starting at a time  $t_L$  from the direct sound. The time  $t_L$  is defined as the moment when the energy of the frequencies outside of the 0.5–1 kHz octave bands have decayed by 30 dB. To further reduce unwanted components outside of the flutter band, we used a bandpass filter with a passband at 0.25–2 kHz and with stopband attenuation of 60 dB. The bottom pane of Fig. 2 shows a section of the autocorrelation function exhibiting a peak at the repetition time  $t_r$  relative to the late-reverberation tail displayed in the top pane.

Another technique for period estimation of fundamental modal frequency was presented by Goto et al. [11]. The method is based on the autocorrelation and power spectral peak selection, which showed robustness to unknown non-stationary noise. For the RIRs analyzed in this work, the repetition time estimated with Goto's method was consistent with the more traditional autocorrelation technique.

## 2.3. Energy Decay Curve

Flutter echo measurements show a stepped decay of the energy decay curve (EDC) in frequency bands that contain audible flutter, indicating non-diffuse features of the sound field [12, 13]. The distance between two successive points of the steepest decay in the EDC correlates with the time spanned by the series of pulses covering a round trip path between the reflective parallel surfaces, resulting in a reliable estimation of  $t_r$ .

The bottom pane of Fig. 3 presents an example of a stepped EDC in the 500-Hz octave band of a RIR containing flutter echo. Time instants of the greater slope were derived from the local maxima of the approximated derivative of the EDC, highlighted in the top pane of Fig. 3 with vertical lines. We selected the tallest peaks that were separated by at least  $0.75 t_r$ .

The values of  $t_r$  estimated from the EDCs showed to be in

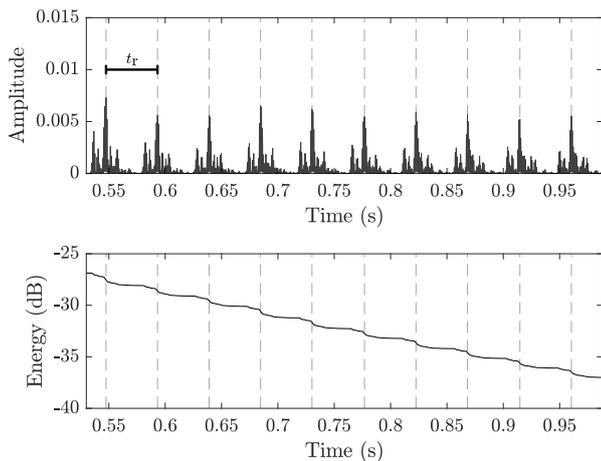


Figure 3: (Bottom) EDC at the 500-Hz octave band relative to the RIR in Fig. 1 and (top) its estimated derivative starting at  $t_L = 535$  ms.

accordance with the traditional autocorrelation analysis. In addition, this method provides a better understanding of  $t_r$  variations in different octave bands, which could be useful for an extended analysis of different echo paths in irregular rooms, for example, those covered with domes.

## 2.4. Ringing Modes

The excitation of partials of  $f_0$ , which for a regular (shoebox) room is usually in the low-frequency range, gives a middle frequency coloration to the RIR. Here, the slowest decaying frequency in the middle frequency range is referred to as flutter tonality. In more realistic scenarios, the full analysis of flutter echo is hindered by other phenomena, such as diffraction and long-ringing room modes.

The frequency of the long-ringing modes, was found from the magnitude of the reference frequency response at time  $2t_L$ . A window of about 500 Hz is shifted on the frequency-domain signal with a 25% overlap. The highest component within each instance of the window is then selected as a candidate frequency. The size of the window  $W$  is to be set depending on the expected number of ringing modes  $k_r$ , defined as the number of peaks of the smoothed reverberation time (RT) curve in the bandwidth  $B_f$  of the fluttery region, as  $W = B_f/k_r$  Hz. The accuracy of the candidate frequencies can be further verified from the RT curve calculated from the energy decay relief (EDR) [7]. The frequency of the long-ringing modes is identified as points of local maxima in the envelope function of the RT curve, determined using spline interpolation over peaks separated by at least  $0.75 f_0$ .

## 2.5. Pulse Shape

Depending on the position of the sound source and receiver, the successive reflections of the source impulses will form a repetitive series of pulses of duration  $t_r$  s, as seen for example in the top pane of Fig. 2. A listener located in the midpoint between the walls will hear one pulse every  $t_r$  s only for the case where the source is also equidistant from the walls. In almost all the other configurations,

the tail of the RIR will be characterized by a repetitive pattern of four pulses [10]. After  $t_r$  seconds from the direct sound, the first four echoes already arrived at the microphone position although they can be hardly heard as they are hidden by other early reflections components.

To produce a good estimation of the initial pulse shape, we decided to use multi-resolution analysis (MRA) of the maximal overlap discrete wavelet transform. Wavelet multi-resolution analysis has already been used in [13] to isolate repeating patterns related to flutter and provide a comparison between measurements and computer simulation results. The frequency components accounting for the shape of the series of pulses are most likely to fall in the flutter range, where the stepped decay in the EDC is perceivable. From this assumption, we decided to use only the level of the MRA that isolates the frequency bands of interest. We chose the Daubechies wavelet family with 11 vanishing moments and projected the RIR onto the level 4 wavelet subspace. The projection of the RIR maintains the original exponential decaying envelope that needs to be removed to avoid overestimation of some of the peaks in the pattern. To this objective, we multiply the resulting projection by the reciprocal of its envelope function.

To reconstruct the initial magnitude response of the series of pulses, a segment  $x_p(t)$  of length  $L = \lfloor t_r/f_s \rfloor$  samples was selected from the flattened projection, where  $f_s$  is the sampling rate in Hz and  $\lfloor \cdot \rfloor$  is the rounding operation to the nearest integer. For simplicity, the starting point  $\tilde{t}_L$  of the segment is defined as the time corresponding to the integer multiple of  $t_r$  closest to  $t_L$ , when the non-fluttery components have already sufficiently decayed. The magnitude of the frequency components has to be adjusted to account for the frequency-dependent decay affecting the signal from the direct sound arrival time to the  $\tilde{t}_L$ . To this end, we use the energy-decay rate derived from the attenuation filter employed in the synthesis algorithm, and whose design is explained in Section 3. The magnitude response is reconstructed as

$$|\tilde{X}_P(t_s, f)| = |X_P(\tilde{t}_L, f)| \cdot 10^{\frac{3(\tilde{t}_L - t_s)}{\Gamma_a(f)}}, \quad (2)$$

where  $\tilde{X}_P(t, f)$  is the estimated frequency response of the series of pulses at time  $t$  and frequency  $f$ ,  $X_P(t, f)$  is the frequency response of  $x_p(t)$ ,  $\Gamma_a(f)$  is the frequency-dependent attenuation in dB produced by the attenuation filter, and  $t_s$  is the onset time of the synthesized flutter echo, defined as the arrival time of the direct sound incremented by  $t_r$  s. In the measured RIR, the first echoes are most likely to appear before  $t_s$  second, unless the sound source is located at the mid point between the reflective walls. This choice of  $t_s$  will thus disregard the echoes arriving during the first  $t_r$  s.

To reconstruct the temporal envelope we use the phase of the reference signal at time  $\tilde{t}_L$ ,  $\varphi_P(\tilde{t}_L, f)$ , as

$$\tilde{x}_p(t_s) = \mathcal{F}^{-1} \left\{ |\tilde{X}_P(t_s, f)| \exp \{ i \varphi_P(\tilde{t}_L, f) \} \right\}, \quad (3)$$

where  $\mathcal{F}^{-1}\{\cdot\}$  is the inverse Fourier transform, and  $i$  is the imaginary unit. The series of pulse estimated from the projection of  $t_r$  seconds of the signal depicted in the top pane of Fig. 2 is shown in Fig. 4 (top), where the vertical dashed lines mark the peaks of the four pulses.

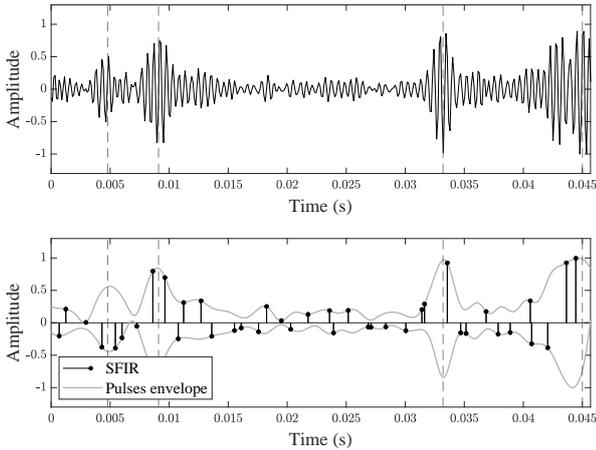


Figure 4: (Top) Projection of the series of pulses of the RIR in Fig. 1 onto the wavelet subspace and (bottom) coefficients of the SFIR filter derived from the velvet-noise sequence for  $t_r = 45.7$  ms, where the temporal envelope of the sequence has been applied to match the estimated shape of the series of pulses.

### 3. FLUTTER ECHO SYNTHESIS

Having estimated the flutter echo parameters from a reference RIR, we can use them in the synthesis algorithm. The proposed synthesis model, shown in Fig. 5, consists of a feedback structure and a sparse finite impulse response (SFIR) filter performing convolution with the velvet-noise sequence. The feedforward path is composed of a bandpass filter  $H_{eq}(z)$ , a delay line, and a ripple filter  $H_r(z)$ . The total delay-line length  $L'$  in the feedforward path is related to the length of the RIR segment containing four flutter echo pulses (cf. Sec. 2.5). The parameter  $L'$  also determines the lengths of the feedforward sections,  $L' = M + R + \varepsilon$ , where  $M$  is the delay introduced by the delay filter,  $R$  is the delay of the ripple loop, and  $\varepsilon$  is the group delay caused by the bandpass filter.  $L'$  defines the repetition time of the synthesized flutter echo.

To reduce the complexity of the system, the synthesis algorithm works at a lower sample rate. The input signal is downsampled by a factor  $K$  by the decimator block before passing through the feedback structure. An interpolator block with an upsampling factor of  $K$  brings back the processed signal to its original sample rate. In this section, the implementation of all the components of the structure is presented.

#### 3.1. Sampling Rate Conversion

The spectrogram and the energy-decay curves of the RIRs studied in this work indicate that flutter echo grows more feeble at frequencies above 2 kHz and becomes nearly inaudible at frequencies above 4 kHz. Due to the biggest chunk of the flutter energy being concentrated below this frequency, the sample rate of the synthesis block was reduced to 8.82 kHz. The decimator block consists of a finite-impulse-response (FIR) anti-aliasing filter followed by a downsampling by factor of  $K = 5$ , which results in a Nyquist limit of 4.41 kHz for the downsampled signal, when the input signal has a  $f_s$  of 44.1 kHz. The decimation filter is implemented efficiently using a polyphase structure whose coefficients are derived from a lowpass FIR filter. We obtained good results with a filter based on

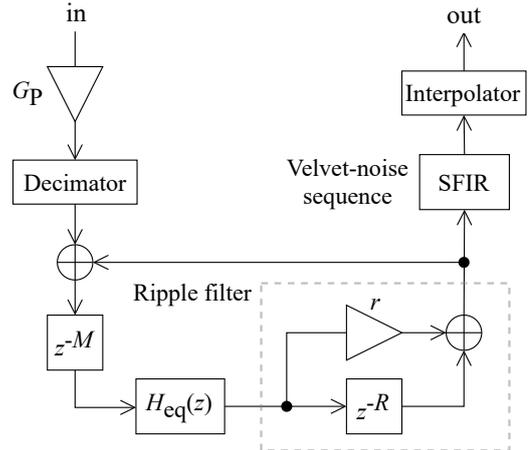


Figure 5: Block diagram showing the structure of the flutter echo synthesis algorithm.

the Kaiser window of order 10 and a shape factor of 5.

Similarly to the decimator block, a polyphase structure is used to implement the upsampling and anti-imaging filtering operations efficiently for the purpose of restoring the original  $f_s$ . The lowpass FIR filter is similar to the one employed by the decimator, with the only difference that the order is increased to 25.

#### 3.2. Energy Decay Rate

The energy decay of flutter can be divided into two trends: an overall downward trend as a function of the distance from the flutter tonality, and a slow decay of partials of the repetition frequency  $f_0$  that are close to the ringing room modes. For the approximation of the overall frequency-dependent RT,  $T_{60}(f)$ , we use a combination of two Butterworth filters, one lowpass and one highpass, to form a bandpass filter with an asymmetric magnitude response and a flat passband. The advantage of using two separate filters is the ability to model the stopband attenuation and the roll-off at high and low frequencies independently.

The design parameters of the Butterworth filters, i.e. the order and the cutoff frequency, are determined by the prototype magnitude response  $G_a(f)$ , derived from the energy-attenuation rate of the reference RIR,  $\Gamma(f) = -60/T_{60}(f)$ , as

$$G_a(f) = \Gamma(f) \frac{L'}{f_s} . \quad (4)$$

where  $L' = \lfloor \frac{L}{K} \rfloor$  is the delay-line length after downsampling. The magnitude response  $G_a(f)$  can be smoothed to reduce the effect of noise at low frequencies. Few manual adjustments of the filter magnitude response might be necessary in case the RT is affected by noise at higher frequencies.

Bayesian decay analysis of the RIRs presented in this work showed that the flutter frequency bands exhibit multiple decay slopes. In particular, two trends can be identified: a steeper decay characterizing the early reflections, and a slower decay associated with the flutter echoes. Consequently, we decided to use the latter slope and use the RT value associated with it to compute  $G_a(f)$ .

In the example in Fig. 6, the longest ringing mode is at around 700 Hz, which coincides with maximum of both: the magnitude

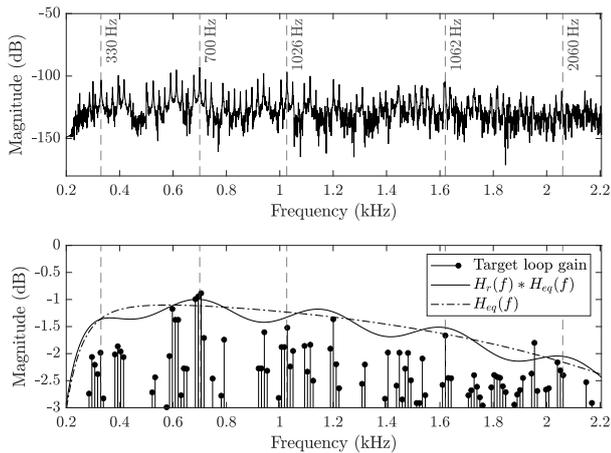


Figure 6: (Top) Normalized magnitude of the RIR in Fig. 1 at time  $2t_L$  and (bottom) target loop gain  $G_a(f)$  for the flutter echo synthesis with peak at 700 Hz, magnitude response of the feedback loop filter, and magnitude response of the parametric equalizer.

response and the target loop gain. The difference in frequency between the adjacent modes, which were detected from the magnitude response, can be approximated as multiples of  $f_0$ . By taking the average of the difference between modal frequencies we found that they differ by about  $k_P f_0$  Hz, for  $k_P = 21$ . To model this behaviour we use an approach similar to [14], where a feed-forward comb filter (ripple filter) was introduced in the harpsichord string model to simulate the differences in the partial decay rates. The variation introduced by the filter follows a sinusoidal decay with its maxima at the slowest-decaying modes. One limitation of this filter is that it over approximates the trend of the decay of intermediate frequencies, which is not perfectly sinusoidal and it is characterized by finer and in-harmonic components. However, the masking effect that the ringing modes have on the perceived tonality allows us to neglect the produced error [15].

In this work, the ripple filter  $H_r(z)$  was designed as a feed-forward comb filter with a feedforward coefficient  $r$  and a delay-line length  $R = L'/k_P$  seconds. The characteristic frequency response of the comb filter introduces symmetric oscillations around the magnitude response of the parametric equalizer  $H_{eq}(z)$ . By applying a scaling factor to the impulse response of  $H_r(z)$  the frequency response can be shifted downward in order to maintain the target maximum gain at the flutter tonality.

### 3.3. Velvet-Noise Sequence

Velvet noise is a discrete sparse sequence in which most samples are zero, whereas the non-zero samples can take values  $\{-1, 1\}$  only [4]. The number of impulses per second, also known as pulse density  $N_d$ , determines the perceived smoothness of the sequences. Its relation to the average distance between impulses  $T_d$  is defined as

$$T_d = \frac{f_s}{N_d}. \quad (5)$$

It was found experimentally that for  $f_s$  of 44.1 kHz and pulse density of at least 2000 impulses per second, the velvet noise is perceived as smoother than the Gaussian white noise [4, 16].

The impulse locations in a velvet-noise sequence are deter-

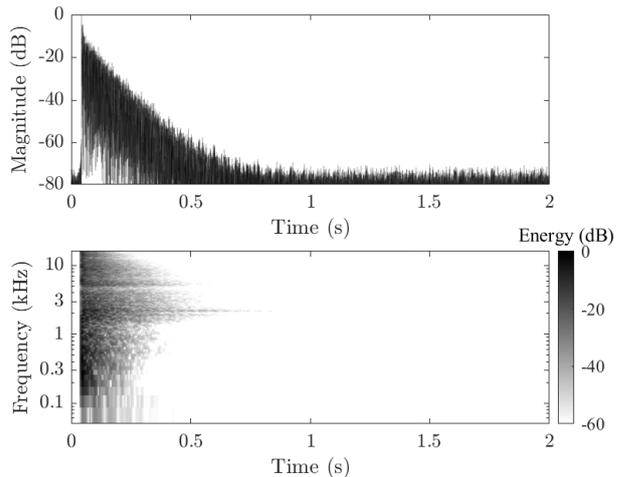


Figure 7: (Top) Measured RIR of the variable acoustics laboratory Arni with random configuration of the reflective panels and (bottom) its spectrogram.

mined as [16]

$$k(m) = \lfloor mT_d + r_1(m)(T_d - 1) \rfloor, \quad (6)$$

where  $m$  is the pulse counter and  $r_1(m)$  is a random value drawn from a uniform distribution on the interval (0,1). The sign of the impulses is also determined with a random number generator [4, 16, 5].

As the last step of the proposed architecture, the output of the feedback structure is processed with an SFIR filter. The coefficients of this filter are taken from a velvet-noise sequence with  $T_d = 20$ . The length of the sequence should coincide with the delay-line length  $L'$  to avoid gaps of silence or overlapping pulses in the output of the system. The temporal envelope estimated from the shape of the series of pulses is applied on the generated sequence to increase similarity to the target flutter. The SFIR coefficients determined by the velvet-noise sequence and pulse shape from the reference RIR are depicted in the bottom pane of Fig. 4.

The early reflections and the non-fluttery parts of the synthetic reverberation were obtained from the measured RIR and used as coefficients for an FIR filter that was applied directly to the input signal. Before mixing the synthesized flutter with the early reflections, its gain is adjusted according to the target average signal power by means of a scaling factor  $G_P$  defined as

$$G_P = \frac{(\sqrt{\widehat{P}_r} - \sqrt{\widehat{P}_e})^2}{\widehat{P}_f}, \quad (7)$$

where  $\widehat{P}_r$ ,  $\widehat{P}_e$ , and  $\widehat{P}_f$  are the average powers calculated from the target RIR, the simulated early reflections, and the synthetic flutter signal, respectively. Similarly to [17], to compute the average power we perform spectral whitening of the RIRs with a linear predictor inverse filter of order 10.

## 4. EVALUATION

In this section, the performance of the proposed method is evaluated based on the reproduction accuracy of the repetition time

Table 1: RT values for measured and synthetic RIRs and the percentage difference in octave frequencies. Values not exceeding the JND of 5.0 % are highlighted.

Center frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
Measured RT (s)	0.39	1.47	2.04	2.05	1.22	0.65	0.48	0.31
Synthesized RT (s)	0.34	1.53	1.97	2.12	1.25	0.62	0.49	0.31
Difference (%)	14.71	<b>3.92</b>	<b>3.55</b>	<b>3.30</b>	<b>2.40</b>	<b>4.84</b>	<b>2.04</b>	<b>0.00</b>

and frequency-dependent energy decay. The flutter echo parameters were based on the reference RIR shown in Fig. 1. In the flutter echo synthesis, the early reflections were obtained from a measured RIR depicted in Fig. 7. Both RIRs were captured in the variable acoustics laboratory Arni at the Acoustics Lab of Aalto University, Finland [18, 19]. The number of the absorptive and reflective elements in the room was the same in both measurements. The RIR in Fig. 1, however, was measured with reflective panels along two parallel walls. In case of the RIR depicted in Fig. 7, the reflective panels were scattered randomly over the walls, so that the flutter echo does not occur.

The results of the flutter echo synthesis are presented in Fig. 8. The proposed algorithm successfully simulates the vertical stripes representing the repetitive pulses in the frequency region of flutter. The harmonic behavior observed in the reference RIR in Fig. 1 is recreated as well. The synthesized audio samples are available online<sup>1</sup>, along with further examples of measured and synthesized fluttery RIRs. The MATLAB implementation of the proposed method can be found in the online repository<sup>2</sup>.

For the estimated RT, the  $T_{30}$  values are used instead of  $T_{60}$  as the signal-to-noise ratio is insufficient to measure the 60-dB decay. The difference between the RT values of the synthesized RIR and the reference values should not exceed 5%, which corresponds to the just noticeable difference (JND) of reverberation [20]. Fig. 9 shows the RT values in octave bands that were calculated from the RIRs using the ITA Toolbox for acoustic measurements [21]. In Table 1, the obtained  $T_{60}$  values are compared to the reference ones. The difference generally does not exceed 5%, and at higher frequencies it is less than 2.1%, proving that the attenuation filter does not leak additional decay beyond the fluttery region. The RIR used for the synthesis of the early reflections has less energy in the low-frequency region when compared to the reference RIR, resulting in a large difference at the 125-Hz octave band.

The distance  $\ell$  between the two parallel walls was 8.1 m, suggesting a  $t_r$  of about 47.2 ms. Applying the methods presented in Sec. 2.1 on the reference RIR, we obtained an average value of  $t_r = 45.78$  ms, with a maximum discrepancy of 0.18 ms across different methods. Using the same methods on the synthesized RIR we obtain an average value of  $t_r = 45.76$  ms, with a maximum difference of 0.06 ms, proving that the proposed method is able to accurately simulate the repetition time of the reference RIR.

### 5. CONCLUSIONS

This paper introduces a new method to synthesize flutter echo as an audio effect. The basic structure of the synthesis algorithm con-

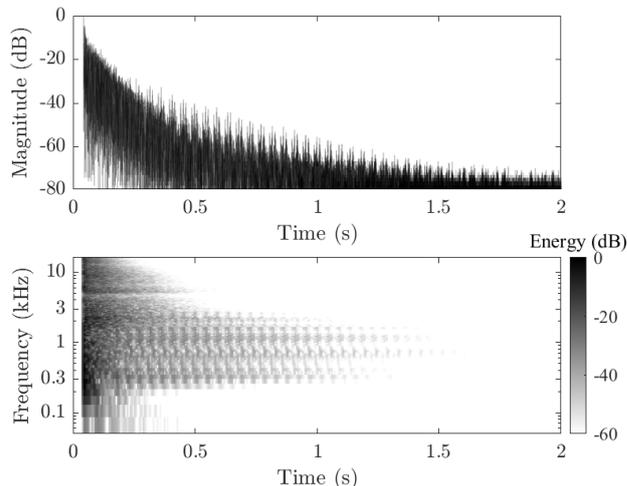


Figure 8: (Top) RIR with the synthetic flutter echo and (bottom) its spectrogram, cf. Fig. 1.

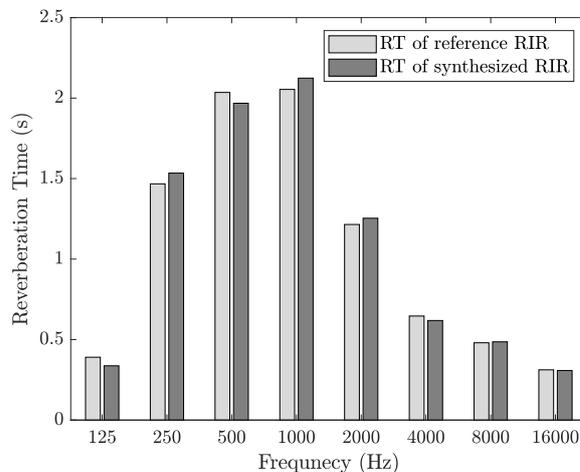


Figure 9: Comparison of RT values of the reference and synthetic RIR.

sists of a feedback structure with two filters: an attenuation filter used to obtain a target decay, and a ripple filter aimed at realistic reproduction of the modal behavior of the flutter. The output of the feedback loop is convolved with a velvet-noise sequence to obtain a dense fluttery sound. The limited frequency range occupied by the flutter components allows to apply sample rate conversion and thus to reduce the computation required by the velvet-noise con-

<sup>1</sup><http://research.spa.aalto.fi/publications/papers/dafx22-flutter-echo/>

<sup>2</sup><https://github.com/gdalsanto/flutter-echo-modeling>

volution. Additional adjustments are introduced by using a temporal envelope on the velvet-noise sequence, so that it resembles the shape of the original series of pulses in the flutter echo.

Comparison of the synthetic and reference RIRs shows a good agreement between the objective parameters. The obtained reverberation time is within one JND of the target values in the flutter range, and the repetition time is consistent with that estimated from the reference RIR.

## 6. ACKNOWLEDGMENTS

This research was conducted at the Aalto Acoustics Lab, when the first author was a visitor there from February to August 2022. The authors would like to thank Aaron Geldert for the editorial suggestions.

## 7. REFERENCES

- [1] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, “Fifty years of artificial reverberation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, July 2012.
- [2] J.-M. Jot and A. Chaigne, “Digital delay networks for designing artificial reverberators,” in *Proc. 90th Conv. Audio Eng. Society*, Paris, France, Jan. 1991.
- [3] T. Halmrast, “Flutter echoes: Timbre and possible use as sound effect,” in *Proc. 18th Int. Conf. Digital Audio Effect (DAFx-15)*, Trondheim, Norway, Nov. 2015.
- [4] M. Karjalainen and H. Järveläinen, “Reverberation modeling using velvet noise,” in *Proc. Audio Eng. Soc. 30th Int. Conf. Intelligent Audio Environments*, Saariselkä, Finland, Oct. 2007.
- [5] V. Välimäki and K. Prawda, “Late-reverberation synthesis using interleaved velvet-noise sequences,” *IEEE/ACM Trans. Audio, Speech, and Language Process.*, vol. 29, pp. 1149–1160, Feb. 2021.
- [6] J.-D. Polack, *La transmission de l’énergie sonore dans les salles (The transmission of sound energy in auditoria)*, Ph.D. thesis, Jan. 1993.
- [7] J. M. Jot, L. Cerveau, and O. Warusfel, “Analysis and synthesis of room reverberation based on a statistical time-frequency model,” in *Proc. Int. Computer Music Conf.*, Thessaloniki, Greece, Sept. 1997.
- [8] P. Rubak and L. G. Johansen, “Artificial reverberation based on a pseudo-random impulse response, part I,” in *Proc. Audio Eng. Soc. 104th Conv.*, Amsterdam, The Netherlands, May 1998, paper no. 4725.
- [9] P. Rubak and L. G. Johansen, “Artificial reverberation based on a pseudo-random impulse response, part II,” in *Proc. Audio Eng. Soc. 106th Conv.*, Munich, Germany, May 1999, paper no. 4900.
- [10] D. Y. Maa, “The flutter echoes,” *The Journal of the Acoustical Society of America*, vol. 13, no. 2, pp. 170–178, Oct. 1941.
- [11] S. Goto, Y. Takahashi, and M. Tohyama, “Autocorrelation analysis in time and frequency domains for passive structural diagnostics,” *Advances in Acoustics and Vibration*, vol. 2013, pp. 1–8, Oct. 2013.
- [12] F. Martellotta, L. Álvarez Morales, S. Girón, and T. Zamarreño, “An investigation of multi-rate sound decay under strongly non-diffuse conditions: The crypt of the Cathedral of Cadiz,” *Journal of Sound and Vibration*, vol. 421, pp. 261–274, May 2018.
- [13] H. Lai and B. Hamilton, “Computer modeling of barrel-vaulted sanctuary exhibiting flutter echo with comparison to measurements,” *Acoustics*, vol. 2, no. 1, pp. 87–109, Feb. 2020.
- [14] V. Välimäki, H. Penttinen, J. Knif, M. Laurson, and C. Erkut, “Sound synthesis of the harpsichord using a computationally efficient physical model,” *EURASIP Journal on Advances in Signal Processing*, vol. 2004, pp. 934–948, June 2004.
- [15] S. A. Gelfand, *Hearing: An Introduction to Psychological and Physiological Acoustics*, CRC Press, Boca Raton, sixth edition, 2018.
- [16] V. Välimäki, H.-M. Lehtonen, and M. Takanen, “A perceptual study on velvet noise and its variants at different pulse densities,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 21, pp. 1481–1488, July 2013.
- [17] V. Välimäki, B. Holm-Rasmussen, B. Alary, and H.-M. Lehtonen, “Late reverberation synthesis using filtered velvet noise,” *Applied Sciences*, vol. 7, no. 5, pp. 483, May 2017.
- [18] K. Prawda, S. J. Schlecht, and V. Välimäki, “Evaluation of reverberation time models with variable acoustics,” in *Proc. 17th Sound and Music Computing Conf.*, Turin, Italy, June 2020, pp. 145–152.
- [19] K. Prawda, S. J. Schlecht, and V. Välimäki, “Robust selection of clean swept-sine measurements in non-stationary noise,” *The Journal of the Acoustical Society of America*, vol. 151, no. 3, pp. 2117–2126, Mar. 2022.
- [20] M. Karjalainen and H. Järveläinen, “More about this reverberation science: perceptually good late reverberation,” in *Proc. 111th Audio Eng. Soc. Conv.*, New York, USA, Sept. 21–24, 2001.
- [21] P. Dietrich, M. Guski, J. Klein, M. Müller-Trapet, M. Pollow, R. Scharrer, and M. Vorlaender, “Measurements and room acoustic analysis with the ITA-Toolbox for MATLAB,” Jan. 2013.