OPEN ACCESS Freely available online
V. Välimäki, L. Fierro, S. J. Schlecht and J. Backman, "Audio Peak Reduction Using Ultra-Short Chirps"
J. Audio Eng. Soc., vol. 70, no. 6, pp. 485–494, (2022 June).
DOI: https://doi.org/10.17743/jaes.2022.0011

Audio Peak Reduction Using Ultra-Short Chirps

VESA VÄLIMÄKI,[1] *AES Fellow*, **LEONARDO FIERRO,**[1]
(vesa.valimaki@aalto.fi) (leonardo.fierro@aalto.fi)
SEBASTIAN J. SCHLECHT,[1,2] *AES Associate Member* **AND JUHA BACKMAN,**[3] *AES Fellow*
(sebastian.schlecht@aalto.fi) (juhabackman@aactechnologies.com)

[1]*Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*
[2]*Media Lab, Department of Art and Media, Aalto University, Espoo, Finland*
[3]*AAC Technologies, Turku, Finland*

Two filtering methods for reducing the peak value of audio signals are studied. Both methods essentially warp the signal phase while leaving its magnitude spectrum unchanged. The first technique, originally proposed by Lynch in 1988, consists of a wideband linear chirp. The listening test presented here shows that the chirp must not be longer than 4 ms, so as not to cause any audible change in timbre. The second method, called the phase rotator, put forward in 2001 by Orban and Foti is based on a cascade of second-order all-pass filters. This work proposes extensions to improve the performance of the methods, including rules to choose the parameter values. A comparison with previous methods in terms of achieved peak reduction, using a collection of short audio signals, is presented. The computational load of both methods is sufficiently low for real-time application. The extended phase rotator method is found to be superior to the linear chirp method and comparable to the other search methods. The practical peak reduction obtained with the proposed methods spans from 0 to about 3.5 dB. The signal processing methods presented in this work can increase loudness or save power in audio playback.

0 INTRODUCTION

Peak reduction is a common operation in audio mastering and playback. Conventionally a dynamic range compressor (DRC) is used to limit the largest parts of the audio waveform in a nonlinear manner. This paper studies two linear filtering techniques to reduce the peak value [1, 2].

In 1988 Lynch [1] published a paper about the reduction of the crest factor, or the peak-to-RMS ratio, of speech signals. He proposed to combine a nonlinear DRC and his new idea, which he called *quadratic phase dispersion*. Lynch briefly described the filter impulse response to be a chirp with frequency increasing linearly.

As his application was speech processing, Lynch set the bandwidth of the chirp to 5 kHz [1]. He mentions having tested chirps of different lengths from 1 ms up to 18 ms, which he reported sounding distracting. Lynch finally settled at a 12-ms chirp for speech. With the dispersion filter alone, he achieved a modest average peak reduction of 1.3 dB using voice data. In combination with a DRC, the method yielded an average reduction of 7.7 dB.

Even less is known about the technique proposed by Orban and Foti, which they call the *phase rotator* [2, 3]. It is a chain of all-pass filters, "typically four poles, all at 200 Hz," Orban and Foti write [2]. Such a filter system has a flat magnitude response and nonlinear phase response. Orban and Foti mention that this processing achieves a peak reduction of 3–4 dB on speech recordings. Simultaneously the audio waveform becomes more symmetric, which is another advantage [2].

Many years later, Parker and Välimäki independently studied audio peak reduction [4]. Their work was inspired by the observation that the output signal of a highly dispersive all-pass filter was often smaller in amplitude than the input. They realized that to be useful in practice, the dispersive filter must not smear the sound too much but should remain inaudible. Parker and Välimäki [4] used three Schroeder all-pass filters [5–7] and varied the delay-line length and feedback coefficient to find out that an average 2.5-dB reduction could be achieved in a small selection of musical instrument sounds [4]. Such a method required testing 100 different parameter settings and selecting the best one for each signal segment.

Belloch et al. continued the work, using a high-performance computer with a graphics processing unit, to

*To whom correspondence should be addressed, e-mail: vesa.valimaki@aalto.fi. Last updated: Mar. 18, 2022

explore millions of parameter combinations [8]. The reduction was slightly improved, but the method remains impractical for real-time playback because of the huge burden of the parameter search. The method may have practical value for offline processing of signals, however, for example, for those stored as warning or gaming sounds.

The topic of this paper, audio peak reduction, is a close relative of other phase processing techniques leading to inaudible or nearly inaudible results, such as group-delay equalization [9–11], decorrelation [12–15], and signal-processing techniques used for upmixing [16]. Chirps or sweeps similar to the ones discussed in this paper are used in measurements [17–20], the short ones especially in the estimation of the properties of time-varying systems [21, 22].

This paper explores the Lynch and Orban–Foti techniques for audio peak reduction. The authors suggest a parameter choice for each method so as to obtain the best reduction for each frame, which may be a short segment of the audio waveform containing a local peak. Parameters are restricted so that their impulse response does not lead to excessive smearing, which could be heard as a chirp, or be reminiscent of the spectral-delay effect [23]. Such restricted chirpy impulse responses are called *ultra-short* to emphasize the fact that they do not sound like a sweep but rather like a click. This also helps reduce the search space of the best parameters for each signal.

This paper is organized as follows. SEC. 1 proposes an enhanced linear chirp method and reports on a listening test to find out the perceptual limits of the chirpiness. SEC. 2 introduces improvements to the phase rotator method. In SEC. 3, the proposed methods are compared with each other and other methods and are evaluated using a database of test sounds. SEC. 4 concludes the paper.

1 LINEAR CHIRP

This section analyzes the chirp method introduced by Lynch [1] and proposes an enhanced method that is suitable for high-fidelity audio sampled at $f_s = 44.1$ kHz.

1.1 Analysis of the Lynch Chirp

Lynch [1] originally defined the linear upward chirp with a quadratic phase as

$$h_{\text{Lynch}}(n) = g \sin\left(\frac{\pi W}{T f_s^2} n^2\right), \quad \text{for } n = 0, 1, ..., L-1, \tag{1}$$

where n is the sample index, g is the gain factor, W is the signal bandwidth in hertz, T is the chirp duration in seconds, and $L = \lceil T f_s \rceil$ is the upward chirp length in samples, where $\lceil \cdot \rceil$ is the ceiling, or the largest integer, function. Fig. 1 shows the waveform of the chirp with parameters $W = 20$ kHz and $T = 12$ ms. The chirp has been scaled to have unit energy by setting $g = 0.0621$. The peak waveform value in Fig. 1(a) is surprisingly small ($\simeq 0.06$), which gives hope that the convolution of this sequence with an audio signal can lead to a notable peak reduction.

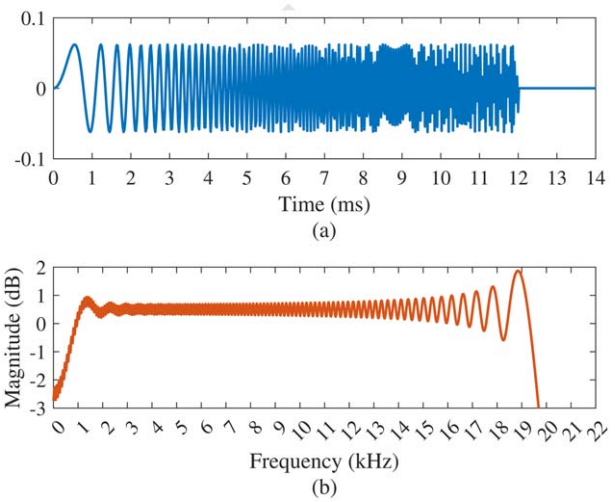

Fig. 1. (a) Waveform and (b) magnitude spectrum of a 12-ms linear chirp from 0 to 20 kHz.

However Fig. 1(b) reveals that the spectrum of the chirp contains some ripple and has irregularities at each end of the audio frequency range: frequencies below 1 kHz are attenuated by almost 3 dB, and there is a 2-dB bump at 19 kHz. The ripple in the frequency domain is the Gibbs phenomenon, which is caused by the abrupt beginning and ending of the signal in the time domain [24]. Using the chirp of Fig. 1(a) as a filter for an audio signal would cause the same changes to its spectrum. The low-frequency attenuation is particularly significant, because it drastically affects the sound quality. Next, modifications to the design, which improve the chirp's performance in the frequency domain, are proposed. A finite-length chirp with a completely flat spectrum cannot be synthesized. However avoiding discontinuities in the frequency response can flatten the spectrum [17, 19].

1.2 Whitened Linear Chirp

The chirp must be created carefully to avoid oscillations in its spectrum. The bump at high frequencies can be corrected by tapering the spectrum smoothly toward the Nyquist limit. Here the suggestion is to first create a full-band linear chirp going from 0 Hz up to the Nyquist limit, i.e., $W = f_s/2$:

$$h(n) = g \sin\left(\frac{\pi}{2 T f_s} n^2\right), \quad \text{for } n = 0, 1, ..., L-1. \tag{2}$$

The wideband chirp $h(n)$ is post-processed in the frequency domain, but its length is not increased much in the time domain. The chirp spectrum $H(f)$ is computed with the discrete Fourier transform (DFT), which should be longer than the chirp length. Zero padding can be applied to extend the sequence to the desired length. Thereafter the response's ripply magnitude spectrum is replaced with a smooth model spectrum $C(f)$:

$$|H(f)| e^{j \angle H(f)} \rightarrow C(f)\, e^{j \angle H(f)}, \tag{3}$$

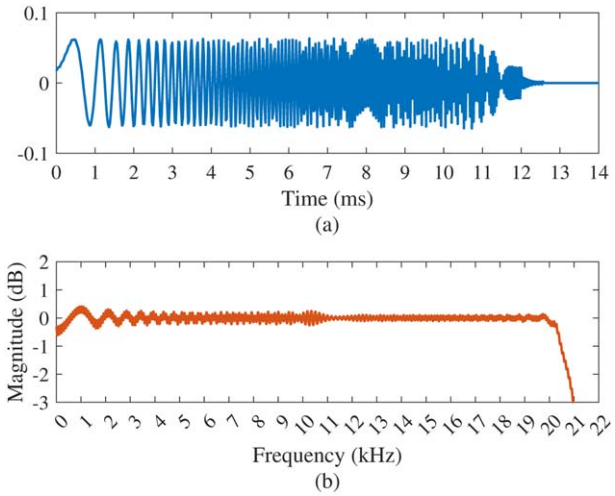

Fig. 2. (a) Waveform and (b) magnitude spectrum of a whitened 12-ms chirp, cf. Fig. 1.

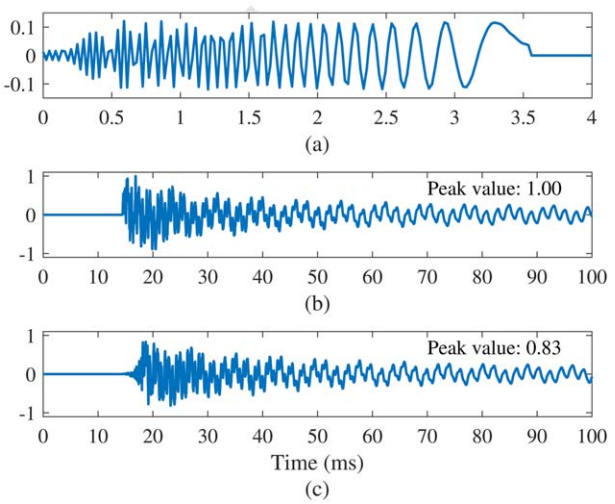

Fig. 3. Example of applying (a) a 3.4-ms downward chirp to filter (b) a mallet sample, which leads to (c) reduction of the signal's peak value. Note the different time scale in (a).

where $|H(f)|$ and $\angle H(f)$ are the magnitude and phase spectra of $h(n)$ and f is the frequency. The model $C(f)$ is unitary between 0 Hz and 20 kHz, and it is attenuated from 1 to 0 above that range using the first quarter of a cosine function. In practice, the model spectrum is combined with the original phase response at both positive and negative frequencies.

Finally the time-domain version of the whitened chirp is obtained with the help of the inverse DFT (IDFT). The flattening process slightly increases the length of the signal, but the enhanced chirp $h_{\text{flat}}(n)$ is obtained by cropping $L + 0.05L$ samples of the IDFT result, i.e., allowing a 5% increase in chirp length at the high-frequency end. This way, the spectral ripple is reduced, and the chirp still remains short.

Fig. 2 shows the waveform and magnitude spectrum of the chirp $h_{\text{flat}}(n)$ whitened using 1,024-point DFT and IDFT processing. The waveform in Fig. 2(a) has been slightly modified in comparison to Fig. 1(a), but above all, the spectral ripple has been attenuated in Fig. 2(b), now restricted within −0.6 and 0.5 dB. In comparison, the linear chirp proposed in [19] of the same length has spectral ripples of 2–3 dB. Shorter chirps down to 1 ms have also been tested, and the ripple always remains within ±0.6 dB, which is considered inaudible and, thus, sufficiently small for hi-fi audio. Even shorter chirps can be used, down to 0.4 ms, provided that an increase in chirp length to 1 ms at the high-frequency end is ensured. Below such a value, the whitening process is unable to counter the ripple effect. At the beginning, or the low-frequency end, of the chirp, extending the length is unnecessary. The DFT/IDFT processing also affects the start of the waveform, however, which can be seen by comparing Fig. 1(a) with Fig. 2(a).

1.3 Peak Reduction Using the Linear Chirp Filter

The samples of the whitened chirp can be used as a finite impulse response (FIR) filter:

$$y(n) = h_{\text{flat}}(n) * x(n), \tag{4}$$

where $x(n)$ and $y(n)$ are the original and processed signal frames, respectively, and the asterisk (*) refers to the convolution operation. The corresponding FIR filter is fairly long, e.g., 529 coefficients when $T = 12$ ms, as in Fig. 2(a). Many experiments with whitened chirps of various lengths were conducted in this study, using both the upward chirp, as the original one proposed by Lynch, and downward chirp, which is obtained by reversing the sequence in time.

Fig. 3 shows a processing example using a downward chirp with a nominal length of 3.4 ms (150 samples, including the 5% extra) to process a mallet percussion instrument sample. The downward chirp has a discontinuity at the end at about 3.6 ms in Fig. 3(a) but is harmless and is only there to flatten the magnitude response of the chirp at the low end. Fig. 3(c) shows that, as a consequence of the convolution with the linear chirp, the peak value of the signal is reduced from 1.0 to 0.83, which corresponds to 1.6 dB. At the same time, the signal waveform has become more symmetric: before filtering, in Fig. 3(b), the minimum and maximum sample values are 1.0 and −0.89, corresponding to a peak-to-peak ratio of 1.12, whereas after filtering, in Fig. 3(c), they are −0.83 and 0.81, and the peak-to-peak ratio is equal to 1.02.

Furthermore the linear chirp causes a few milliseconds of effective delay to the signal, which can be observed by comparing the onset times in Figs. 3(b) and 3(c). For reference, the group-delay estimate of the downward chirp used in this example is presented in Fig. 4(a), which shows that the lowest frequencies undergo the largest delay of about 3.4 ms. On the other hand, the high-frequency range around 20 kHz has the smallest group delay, smaller than about 0.5 ms. In other words, the chirp starts from high frequencies and proceeds approximately linearly downward toward 0 Hz. For comparison, Fig. 4(b) presents the corresponding upward chirp, as originally proposed by Lynch [1].

Recent work on the audibility of group delay showed that the human ear is sensitive to mid and high-frequency varia-

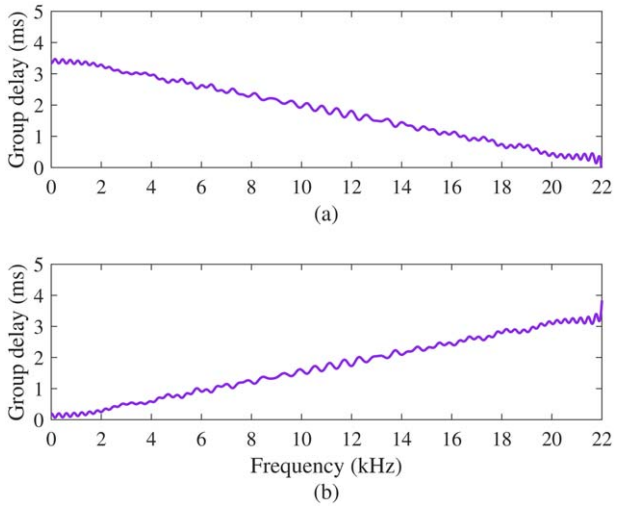

Fig. 4. Group delay of the whitened 3.4-ms (a) downward and (b) upward chirp. The waveform of the downward chirp is shown in Fig. 3(a).

Table 1. Audio items used in the listening test.

Name	Description
UnitImp	A sequence of three synthetic unit impulses.
PinkImp	A sequence of three pink impulses [11].
SynthHH	A sequence of three synthetic hi-hats [11].
Castanets	A sequence of recorded castanets.

tions, while tolerating more phase processing in the low frequencies [11]. This suggests that the downward chirp [see an example in Fig. 4(a)] is a stronger candidate for practical implementation than its time-reversed counterpart, the upward chirp, which is exemplified in Fig. 4(b). Practical limits of the chirp length are studied next.

1.4 Listening Test

This section presents the design and results of a listening test to determine the perceptual limits for ultra-short chirps. The goal is to define the maximum nominal length for the linear chirp before the dispersion becomes audible and reduces the sound quality, i.e., find the longest linear chirp that does not introduce a chirp-like effect in audio signals.

A formal blind listening test was conducted on a selection of 16 experienced listeners. All participants had previous experience with listening tests and reported no hearing impairments or relevant medical conditions. The test software was run on webMUSHRA [25] over a computer running MacOS 10.14.6 using a single pair of Sennheiser HD 650 headphones, inside a soundproof listening booth at the Aalto Acoustics Lab, Espoo, Finland. A set of four short audio samples (1–3 s long, sample rate 44.1 kHz) was selected, inspired by the test reported by Liski et al. [11], which also studied the audibility of phase processing. The test samples are summarized in Table 1.

In each trial, test subjects were presented with one of the audio excerpts as a reference, and they were asked to rate six stimuli on a scale from 0 to 100 with respect to the similarity to the reference sound and the amount of perceived *chirpiness*, which was described to the participants as the audible effect of dispersion in time. Stimuli consisted of the hidden reference and its five different versions filtered with either downward or upward chirps of lengths $T = 2$, 4, 8, 12, or 20 ms, the last of which was considered the low anchor because of the strong audibility of its chirp effect.

The participants were allowed a short training phase before starting the actual test to familiarize themselves with the given concept of chirpiness, get acquainted with the interface, and adjust the loudness to a comfortable level. The scores given during the training phase were not included in the data analysis. During the actual test, participants were presented with a trial with downward chirps and a trial with upward chirps per audio excerpt, each repeated twice for reliability, for a total of 16 trials and 48 different audio stimuli. The processed audio excerpts are available on the companion webpage [26].

Mean opinion scores (MOS) were computed from the subjective ratings to estimate the audibility of the chirp effect. Two subjects (12.5% of the participants) were discarded after an initial screening of the data. The criterion was that any subject rating the reference below 4.6 MOS points, i.e. 90/100, for more than 10% of the trials is disqualified. For 16 trials, in practice, anyone who rated the reference below 90 points more than once was considered unreliable and was discarded.

Box plots of data distribution are displayed in Figs. 5(a) and 5(b), respectively, for the downward and upward chirps. Samples filtered with 2-ms–long chirps are barely distinguishable from the reference, scoring between 4 (Good) and 5 (Excellent) for all audio excerpts. A nominal length of 4 ms makes the chirp slightly more audible, achieving the MOS between 3 (Fair) and 4 for the unit impulse and pink impulse but retaining the higher range for hi-hat and castanets. Moreover, 8-ms–long chirps are ranked between 2 (Poor) and 3. Longer chirps (12 and 20 ms) are largely audible, scoring constantly between 1 (Bad) and 2.

Results show that the audibility of the chirp effect is higher for impulsive sounds, whereas the phase dispersion is partially masked for noisy or reverberant transients, such as the synthetic hi-hat and castanet sounds. This can be taken into account by adjusting the maximum chirp length according to the input audio frame.

The proposed design for the peak-reducing phase dispersion involves both downward and upward chirps with a maximum length of 4 ms to keep them ultra short. The minimum length of the chirps is set to 0.4 ms to guarantee that the magnitude response ripple effect does not exceed the $\pm$1-dB range. A shorter linear chirp is also ineffective in reducing the waveform peak value.

2 PHASE ROTATOR

The phase rotator algorithm briefly mentioned by Orban and Foti [2, 3] is examined next. In this work, an extension that improves the performance of the original method is proposed.

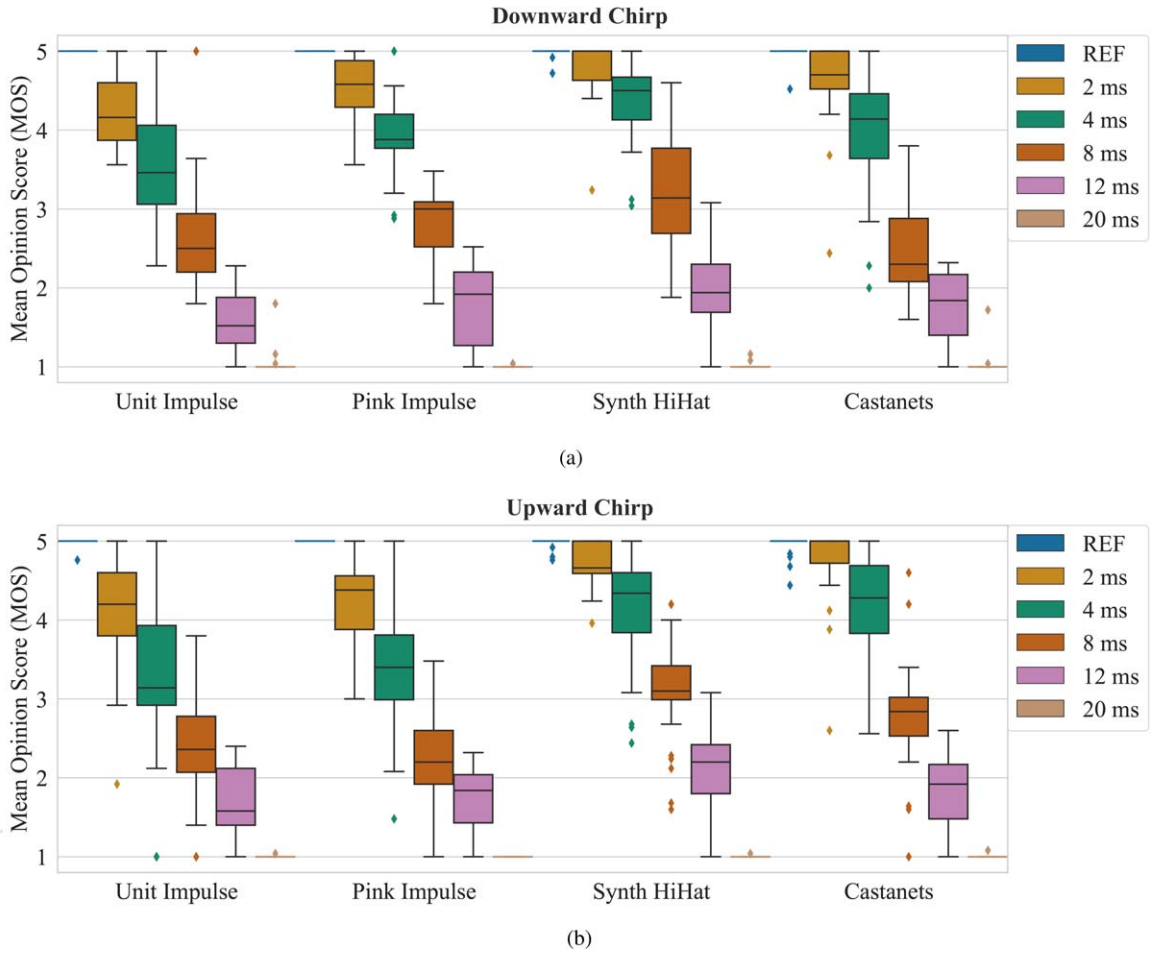

Fig. 5. Mean opinion scores and confidence intervals for (a) downward and (b) upward chirps from the listening test.

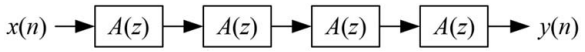

Fig. 6. Phase rotator consists of four cascaded all-pass filters.

2.1 Original Phase Rotator

When Orban and Foti mention a "chain of allpass filters" and "four poles," they refer to four cascaded second-order all-pass filters. The digital second-order all-pass filter has the following transfer function:

$$A(z) = \frac{r^2 - 2r\ \cos(\omega_c)z^{-1} + z^{-2}}{1 - 2r\ \cos(\omega_c)z^{-1} + r^2 z^{-2}}, \tag{5}$$

where $0 \leq r < 1$ is the pole radius and $\omega_c = 2\pi f_c/f_s$ is the pole frequency in radians. To implement a phase rotator, four such filters are cascaded, as shown in Fig. 6, yielding the following transfer function:

$$H_{\text{phr}}(z) = [(A(z)]^4. \tag{6}$$

A synthetic hi-hat signal is filtered with the phase rotator as an example of its usage. Orban and Foti do not reveal the pole radius, so a suitable value was obtained with a manual search. The pole radius of $r = 0.8$, which yielded a good result in this case, is used in Fig. 7. It reduces the peak value of the signal by 1.3 dB. In addition to the reduced peak value, a more symmetric waveform is obtained, since the minimum and maximum sample values, −0.86 and 0.85, respectively, are almost equal after the filtering (peak-to-peak ratio 1.01). The delay caused by the phase rotator to the overall waveform is negligible.

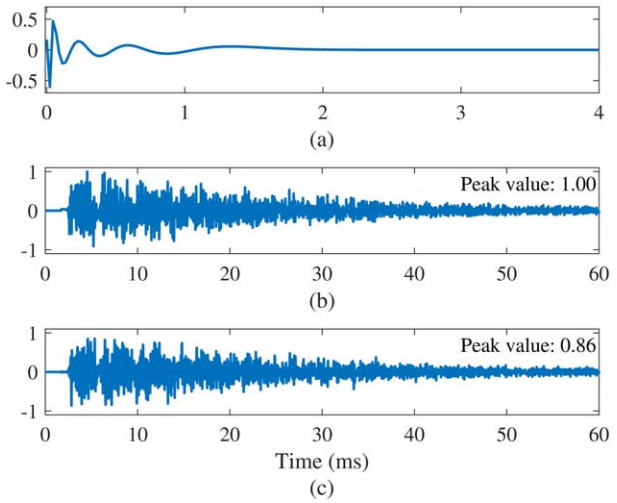

Fig. 7. (a) Phase rotator impulse response ($f_c = 200$ Hz; $r = 0.80$), (b) a hi-hat sample, and (c) the filtered hi-hat sample having a reduced peak value.

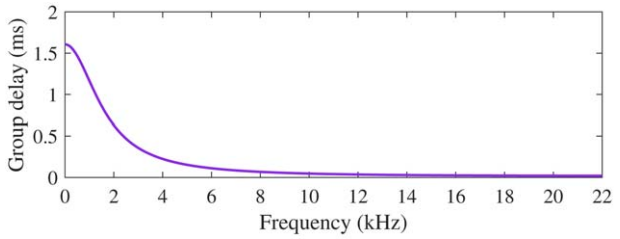

Fig. 8. Group delay of the phase rotator ($f_c = 200$ Hz; $r = 0.80$), whose impulse response is shown in Fig. 7(a).

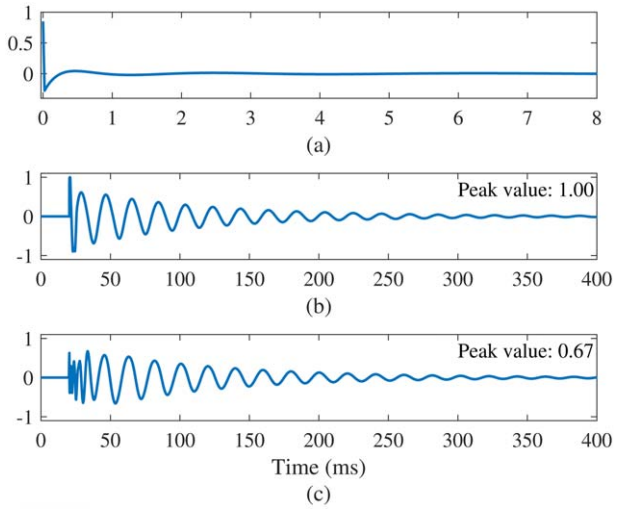

Fig. 9. (a) Extended phase rotator impulse response ($f_c = 40$ Hz; $r = 0.98$) and (b) the unprocessed and (c) processed bass drum sample.

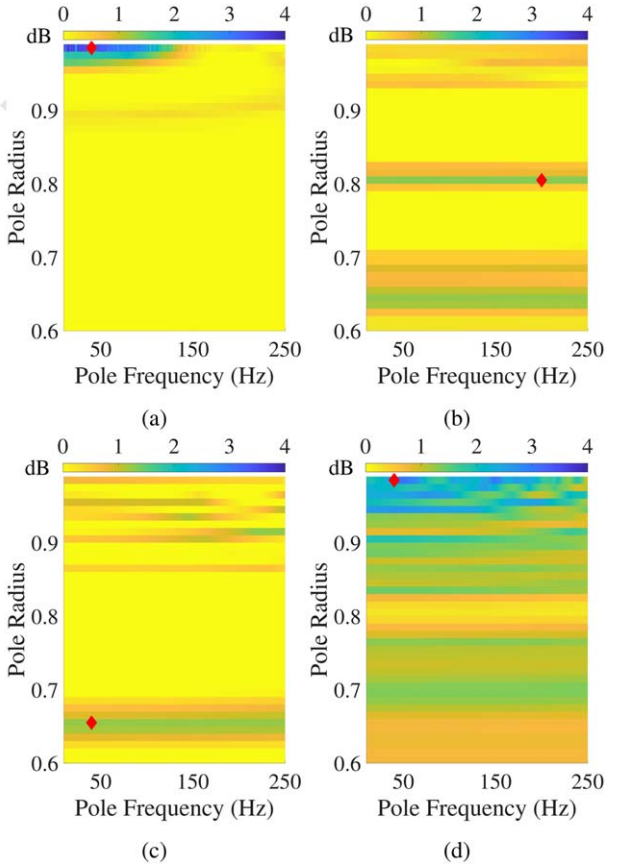

Fig. 10. Peak-reduction heatmap of the extended phase rotator for the (a) bass drum, (b) hi-hat, (c) mallet, and (d) snare drum test signals. The diamond indicates the point of best reduction for each signal.

Fig. 8, which shows the group delay corresponding to the impulse response of Fig. 7(a), reveals that the phase rotator also produces a short chirp as its impulse response. The delay of this phase rotator is about 1.6 ms near 0 Hz and decreases toward higher frequencies. This means that it is a downward chirp: the high frequencies appear first in the impulse response, and the low frequencies follow, as can also be interpreted from Fig. 7(a), where fast oscillations appear in the beginning (at 0 ms and immediately after that) but the response's tail is a wave that slows down with time.

2.2 Extended Phase Rotation Algorithm

Changing the phase rotator parameters ω_c and r was seen to yield a remarkable effect on the reduction performance. Nevertheless it was decided that all four all-pass filters should have the same parameters because otherwise the search space would get expanded in a way similar to Parker and Välimäki's earlier work [4].

Fig. 9 shows an example of processing a bass drum sample when the pole frequency is 40 Hz and pole radius is 0.98, which reduces the peak sample value to 0.67, which is 3.4 dB lower than the original 1.0. For comparison, the all-pass filter parameters used in Fig. 7 do not help reduce the peak value of the bass drum signal but instead increase its maximum value by 0.83 dB. This also shows that a phase rotator with fixed parameters cannot be optimal for all signals.

Fig. 10 shows the performance analysis of the extended phase rotator method over four test signals for the selected search space. The same sound files, all short percussion instrument sounds, have been used previously in the study by Parker and Välimäki [4]: the bass drum (BD), hi-hat cymbal, mallet percussive instrument, and snare drum. Two recurrent patterns are observed in Fig. 10. Either the peak reduction is mainly or solely dependent on the pole radius, hence forming "stripes" in the heatmaps, as in Figs. 10(b) and 10(c), or the optimal reduction concentrates in a low-frequency region near the unit circle, as seen in Fig. 10(a). A mix of the two patterns is observed in Fig. 10(d).

Based on extensive tests with such audio signals, it was decided to allow the pole frequency of the phase rotator to vary between 40 and 200 Hz. On the other hand, the pole radius can be effective in the range from about 0.6 to 0.98. If the poles of the all-pass filters are set closer to 1.0, the phase rotator produces a disturbing change in the signal timbre. As long as the pole radius of the extended phase rotator is kept smaller than about 0.98, its impulse response is an ultra-short chirp, which does not cause a clearly audible chirp-like effect. With the pole radius smaller than 0.6, the effect is diminutive and thus not worth exploring.

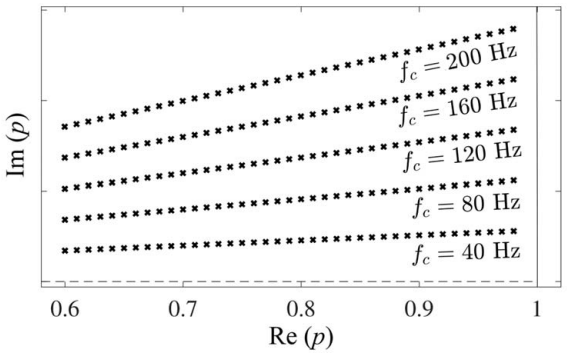

Fig. 11. Search space in the Z plane for the extended phase rotator. Each cross represents one pole location, and there are 200 of them in total.

Table 2. Best parameter values for the extended phase rotator method, cf. Fig. 10. BD = bass drum.

Sound	Pole radius	Pole frequency (Hz)
BD	0.98	40
HiHat	0.80	200
Mallet	0.65	160
Snare	0.98	40

Fig. 11 illustrates the proposed search space for the extended phase rotator that contains five different frequencies and 40 pole radii sampled uniformly between the extreme values. This leads to a search space of 200 parameter sets, which also samples the useful parameters quite densely. Testing this many options per signal frame or transient is reasonable. Additionally a bypass mode must still be implemented, implying no processing, for signals that cannot be reduced, as suggested in earlier works [1, 4].

Fig. 10 proves that including the entire parameter search space is useful, since the optimum parameter value pair (marked by a red diamond for each map) is located at a different point for different test sounds. Table 2 lists the parameter values yielding the largest peak reduction for each test sound according to Fig. 10.

3 EVALUATION

In this section, a comparison of the proposed method improvements with state-of-the-art techniques [4, 8] is conducted using five test sounds used in previous studies and over a dataset of transient sounds.

3.1 Previous Test Sounds

The peak reduction results obtained for five test samples used in previous studies are shown in Table 3. Although not achieving the same performances of the method by Belloch et al. [8], which remains superior for four out of five samples, the improved phase rotator is partially successful because it is always comparable or superior to the method by Parker and Välimäki [4] and achieves the best reduction for the BD excerpt.

Table 3. Comparison of peak reduction capability (in decibels) of four methods on five test signals that have been used in earlier studies [4, 8]. The best reduction for each instrument is shown in bold. BD = bass drum.

Method	BD	Snare	HiHat	Piano	Mallet
Parker [4]	0.5	2.3	1.4	1.3	1.2
Belloch [8]	2.6	**3.2**	**2.5**	**2.3**	**2.7**
Linear Chirp	0.2	2.0	1.3	0.5	1.4
Phase Rotator	**3.5**	2.7	1.3	1.6	1.3

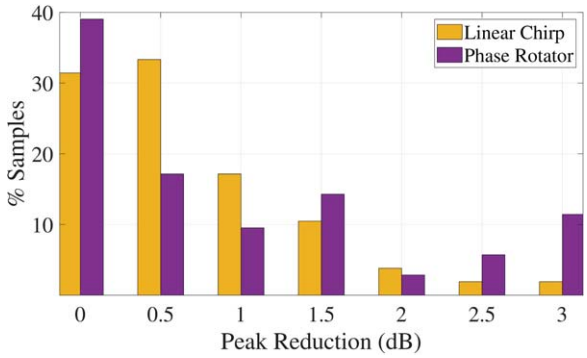

Fig. 12. Peak reduction histogram for the whitened chirp and extended phase rotator methods over the selected database. Values have been rounded to the nearest half-integer.

The whitened chirp method slightly surpasses the phase rotator and Parker–Välimäki methods for the mallet percussion excerpt. It is comparable for the hi-hat cymbal, but it is inferior for the snare drum and fails to achieve a useful reduction for both the BD excerpt and piano sound (see Table 3). All sounds are accessible on the companion website for this paper [26].

3.2 Dataset of Over 100 Sounds

Peak reduction performance for the linear chirp and phase rotator was further tested on a database of 105 different transient samples comprising drums, knocks, speech excerpts, shots, and other impulsive sounds that were compiled for this project. The database also contains precompressed audio and sounds presenting a small crest factor, which are typically hard to reduce.

The results of the evaluation using this dataset are summarized in Fig. 12. The extended phase rotator achieves at least 1 dB of reduction for 43% of the samples and at least 3 dB for more than 10% of them. A 3-dB reduction corresponds to a 50% reduction in signal power.

The performance of the whitened chirp is modest. Although the method provides a reduction of 2 dB or more in some cases, it achieves at least 1-dB reduction only for a third of the samples under test, as can be seen in Fig. 12.

3.3 Implementation Complexity

A comparison of the implementation complexity for the aforementioned methods is reported in Table 4. Parker's method is quite efficient because it only requires filtering the input signal 100 times with three cascaded all-pass fil-

Table 4. Comparison of implementation complexity of peak reduction methods. OPS (operations) refers to the number of multiplications and additions.

Method	OPS/sample	Search space	Usage
Parker [4]	12	Limited to 100	Real-time
Belloch [8]	12	$>10^7$	Offline
Linear Chirp	<379	74	Real-time
Phase Rotator	32	200	Real-time

ters [4]. This can be realized in real time. However the 100 parameter options sample its multidimensional parameter space very sparsely, which leads to unsatisfying results. Belloch's method [8] is otherwise the same but samples the parameters space more densely, leading to tens of millions of options. This is clearly impractical, so this approach is suited to offline processing only.

The linear chirp method is implemented with an FIR filter having 190 or less coefficients, which leads to a maximum of 379 operations per sample (the discrete convolution requires one less addition than multiplication). This corresponds to about 30 times more operations per sample than the three cascaded Schroeder all-pass filters used in the Parker and Belloch methods. However its search space is limited to 74 options, which samples the chirp length quite densely. Thus the chirp method is computationally only slightly more intensive than Parker's method, but is still suitable for real-time processing.

The extended phase rotator method requires 32 operations per sample (four multiplications and four additions for each of its four all-pass filters). Its search space contains 200 options and is thus a little more efficient than the chirp method and much more efficient than Belloch's method. The extended phase rotator is sufficiently efficient for real-time use.

In practice, all four methods would process the input signal in segments containing a single transient or another local waveform peak. The signal filtered with the parameters leading to the best reduction would be selected as the output signal. Processing this linearly flattened signal still with a DRC would be natural, since limiting the signal waveform as much as desired is possible. As indicated in previous studies, the linear peak reduction used prior to a DRC can reduce harmonic distortion [1, 4].

4 CONCLUSION

Two old peak reduction techniques, which have been lost for years and have remained underdeveloped, were studied. This paper first expanded the linear chirp method, originally proposed by Lynch for speech processing in 1988, to wideband audio and reduced its spectral ripple. A formal listening test was conducted, and, based on the results, the chirp length was limited to a maximum of 4 ms to ensure good audio quality. This is much shorter than what Lynch originally proposed. The phase rotator method mentioned by Orban and Foti in their 2001 paper was also extended by investigating its useful parameter space.

The two methods were compared with other known methods. The linear chirp has the smallest search space of all methods because only the chirp length (0.4–4 ms) and direction (up or downward) are varied. However its performance is modest compared with the other methods. Its implementation as a long FIR filter requires more processing power than the other, all-pass filter–based methods. However related sweep methods are very popular in measurements [20], and the ultra-short chirps with a flat spectrum, as described in this work, can find applications in the time-varying system identification, where getting quick snapshots of the system behavior in short time intervals is important [22]. Approximating the linear chirp using all-pass filters would be possible without magnitude-response artifacts and with a potential computational advantage [27, 10].

The comparison revealed that the thus-far little-known phase rotator is a surprisingly simple and useful method. It can achieve, in special cases, a better peak-reduction performance than the massive search technique proposed by Belloch et al. [8], which has a search space several magnitudes bigger. Among all tested cases, the phase rotator was never more than 1.4 dB behind Belloch's method. The phase rotator has the potential to become a widely used peak reduction technique in mobile devices having small loudspeakers.

Both peak reduction techniques discussed in this paper can be applied alone or in combination with a nonlinear compressor to reduce the peak value of the waveform. This can lead to increased loudness, when the waveform is scaled up using make-up gain. The peak reduction methods can be useful in mobile audio devices having a limited output power. Future work includes developing a more sophisticated parameter estimation technique to further optimize the phase rotator technique and studying the psychoacoustics of the phase-based peak reduction by itself and in combination with a conventional DRC.

5 ACKNOWLEDGMENT

The work conducted at Aalto University has been financed in part by AAC Technologies through the Planer project (Aalto University project number 410970). This research is part of the activities of the Nordic Sound and Music Computing Network—NordicSMC (NordForsk project number 86892).

6 REFERENCES

[1] J. T. Lynch, "Reduction of Peak/rms Ratio of Speech by Amplitude Compression and Quadratic Phase Dispersion," *J. Audio Eng. Soc.*, vol. 36, no. 3, pp. 147–152 (1988 Mar.).

[2] R. Orban and F. Foti, "What Happens to My Recording When It's Played on the Radio?" presented at the *111th Convention of the Audio Engineering Society* (2001 Nov.), paper 5469.

[3] R. Orban, F. Foti, and B. Katz, "Radio Ready: The Truth," in B. Katz, *Mastering Audio—The Art and*

the Science, pp. 271–278 (Focal Press, Waltham, MA, 2002).

[4] J. Parker and V. Välimäki, "Linear Dynamic Range Reduction of Musical Audio Using an Allpass Filter Chain," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 669–672 (2013 Jul.). https://doi.org/10.1109/LSP.2013.2263136.

[5] M. R. Schroeder and B. F. Logan, "'Colorless' Artificial Reverberation," *J. Audio Eng. Soc.*, vol. 9, no. 3, pp. 192–197 (1961 Jul.).

[6] S. J. Schlecht, "Frequency-Dependent Schroeder Allpass Filters," *App. Sci.*, vol. 10, no. 1, paper 187 (2020 Jan.). https://doi.org/10.3390/app10010187.

[7] K. J. Werner, F. G. Germain, and C. S. Goldsmith, "Energy-Preserving Time-Varying Schroeder Allpass Filters and Multichannel Extensions," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 465–485 (2021 Jul./Aug.). https://doi.org/10.17743/jaes.2021.0018.

[8] J. A. Belloch, J. Parker, L. Savioja, A. Gonzalez, and V. Välimäki, "Dynamic Range Reduction of Audio Signals Using Multiple Allpass Filters on a GPU Accelerator," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, pp. 890–894 (Lisbon, Portugal) (2014 Sep.).

[9] A. Mäkivirta, J. Liski, and V. Välimäki, "Modeling and Delay-Equalizing Loudspeaker Responses," *J. Audio Eng. Soc.*, vol. 66, no. 11, pp. 922–934 (2018 Nov.). https://doi.org/10.17743/jaes.2018.0053.

[10] J. Rämö and V. Välimäki, "Graphic Delay Equalizer," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 8018–8022 (Brighton, UK) (2019 May). https://doi.org/10.1109/ICASSP.2019.8682949.

[11] J. Liski, A. Mäkivirta, and V. Välimäki, "Audibility of Group-Delay Equalization," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 2189–2201 (2021 Jun.). https://doi.org/10.1109/TASLP.2021.3087969.

[12] G. S. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Comput. Music J.*, vol. 19, no. 4, pp. 71–87 (1995). https://doi.org/10.2307/3680992.

[13] M. Bouéri and C. Kyriakakis, "Audio Signal Decorrelation Based on a Critical Band Approach," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), paper 6291.

[14] E. K. Canfield-Dafilou and J. S. Abel, "A Group Delay-Based Method for Signal Decorrelation," presented at the *144th Convention of the Audio Engineering Society* (2018 May), paper 9991.

[15] S. J. Schlecht, B. Alary, V. Välimäki, and E. A. P. Habets, "Optimized Velvet-Noise Decorrelator," in *Proceedings of the 21st International Conference on Digital Audio Effects (DAFx)*, pp. 87–94 (Aveiro, Portugal) (2018 Sep.).

[16] C. J. Keyes and A. Tan, "Design and Evaluation of a Spectral Phase Rotation Algorithm for Upmixing to 3D Audio," *J. Audio Eng. Soc.*, vol. 68, no. 11, pp. 856–864 (2020 Nov.). https://doi.org/10.17743/jaes.2020.0055.

[17] N. Aoshima, "Computer-Generated Pulse Signal Applied for Sound Measurement," *J. Acoust. Soc. Am.*, vol. 69, no. 5, pp. 1484–1488 (1981 May). https://doi.org/10.1121/1.385782.

[18] D. Griesinger, "Impulse Response Measurements Using All-Pass Deconvolution," in *Proceedings of the AES 11th International Conference: Test & Measurement* (1992 May), paper 11-035.

[19] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An Optimum Computer-Generated Pulse Signal Suitable for the Measurement of Very Long Impulse Responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119–1123 (1995 Feb.). https://doi.org/10.1121/1.412224.

[20] E. K. Canfield-Dafilou and J. S. Abel, "An Allpass Chirp for Constant Signal-to-Noise Ratio Impulse Response Measurement," presented at the *144th Convention of the Audio Engineering Society* (2018 May), paper 10041.

[21] R. Kiiski, F. Esqueda, and V. Välimäki, "Time-Variant Gray-Box Modeling of a Phaser Pedal," in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx)*, pp. 31–38 (Brno, Czech Republic) (2016 Sep.).

[22] A. Wright and V. Välimäki, "Neural Modeling of Phaser and Flanging Effects," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 517–529 (2021 Jul.). https://doi.org/10.17743/jaes.2021.0029.

[23] V. Välimäki, J. S. Abel, and J. O. Smith, "Spectral Delay Filters," *J. Audio Eng. Soc.*, vol. 57, no. 7/8, pp. 521–531 (2009 Jul.).

[24] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Pearson Education, Upper Saddle River, NJ, 2010), 3rd ed.

[25] M. Schoeffler, S. Bartoschek, F.-R. Stöter, et al., "webMUSHRA—A Comprehensive Framework for Web-Based Listening Tests," *J. Open Res. Softw.*, vol. 6, no. 1, paper 8 (2018 Feb.). https://doi.org/10.5334/jors.187.

[26] V. Välimäki, L. Fierro, S. J. Schlecht, and J. Backman, "Audio Peak Reduction Using Ultra-Short Chirps: Companion Webpage," http://research.spa.aalto.fi/publications/papers/jaes-chirp/ (accessed Nov. 12, 2021).

[27] J. S. Abel and J. O. Smith, "Robust Design of Very High-Order Allpass Dispersion Filters," in *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx)*, pp. 13–18 (Montreal, Canada) (2006 Sep.).

THE AUTHORS


Vesa Välimäki


Leonardo Fierro


Sebastian J. Schlecht


Juha Backman

Vesa Välimäki is Professor of Audio Signal Processing and Vice Dean for Research at Aalto University, Espoo, Finland. He received his M.Sc. and D.Sc. degrees from the Helsinki University of Technology in 1992 and 1995, respectively. In 1996, he was a Postdoctoral Research Fellow at the University of Westminster, London, UK. In 2001–2002, he was Professor of Signal Processing at the Pori unit of Tampere University of Technology. In 2008–2009, he was a visiting scholar at the Stanford University Center for Computer Research in Music and Acoustics. His research interests are in signal processing and machine learning applied to audio and music technology. Prof. Välimäki is a Fellow of the AES and IEEE. He was the General Chair of the 14th International Sound and Music Computing Conference in 2017. Currently Prof. Välimäki is the Editor-in-Chief of the *Journal of the Audio Engineering Society*.

•

Leonardo Fierro is a doctoral candidate at Aalto University, Espoo, Finland. He received his M.Sc. degree in Communication Technologies and Multimedia from the University of Brescia, Italy, in 2019. He has been part of the Aalto Acoustics Lab since 2019. He is a Teaching Assistant in Applied Digital Signal Processing and Audio Signal Processing courses at Aalto University. His research interests involve audio time–scale modification, transient modeling, and loudness equalization.

•

Sebastian J. Schlecht is a Professor of Practice for Sound in Virtual Reality at the Acoustics Lab, Department of Signal Processing and Acoustics, and Media Labs, Department of Art and Media, Aalto University, Espoo, Finland. He received the Diploma in Applied Mathematics from the University of Trier, Germany, in 2010 and an M.Sc. degree in Digital Music Processing from the School of Electronic Engineering and Computer Science at Queen Mary University of London, U.K., in 2011. In 2017, he received a Doctoral degree at the International Audio Laboratories Erlangen, Germany, on artificial spatial reverberation and reverberation enhancement systems. From 2012 to 2019, Dr. Schlecht was also external research and development consultant and lead developer of the 3D Reverb algorithm at the Fraunhofer IIS, Erlangen, Germany.

•

Juha Backman is a Senior Technologist at AAC Technologies, Turku, Finland. He received his M.Sc. degree in applied physics and acoustics from Helsinki University of Technology in 1991. In 1994, he joined Nokia Corporation and moved to Microsoft with the corporate merger in 2014, staying there for as long as the audio development for mobile devices continued. After taking the opportunity to explore various industrial applications of acoustics and signal analysis in different companies, he returned to telecommunications audio in 2018. He holds 12 granted patents in mobile phone technology and has published over 60 conference and journal papers. Most of his publications are related to loudspeakers. His AES activities include, besides the publications, serving as the Vice Chairman for the Technical Committee on Loudspeakers and Headphones, being a member and past chairman of the AES Finnish Section, and chairing the 16th and 51st AES International Conferences.