Kahles, Julen; Esqueda Flores, Fabian; Välimäki, Vesa

Oversampling for Nonlinear Waveshaping: Choosing the Right Filters

Published in:
Journal of the Audio Engineering Society

DOI:
10.17743/jaes.2019.0012

Published: 01/06/2019

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

OPEN ACCESS Freely available online
J. Kahles, F. Esqueda and V. Välimäki "Oversampling for Nonlinear Waveshaping: Choosing the Right Filters"
J. Audio Eng. Soc., vol. 67, no. 6, pp. 440–449, (2019 June.).
DOI: https://doi.org/10.17743/jaes.2019.0012

Oversampling for Nonlinear Waveshaping: Choosing the Right Filters

JULEN KAHLES,
(julen.kahles@aalto.fi)

FABIÁN ESQUEDA, AND VESA VÄLIMÄKI, *AES Fellow*
(fabian.esqueda@native-instruments.de) (vesa.valimaki@aalto.fi)

Aalto University, Acoustics Lab, Dept. of Signal Processing and Acoustics, FI-02150 Espoo, Finland

An assessment of filters for classic oversampled audio waveshaping schemes is carried out in this paper, pursuing aliasing reduction. For this purpose, the quality measure of the A-weighted noise-to-mask ratio is computed for test tones covering the frequency range from 27.5 Hz to 4.186 kHz, sampled at 44.1 kHz, and processed at eight-times oversampling. All filters are designed to have their passband contained within a ± 1 dB range and to display a minimum stopband attenuation value of 40 dB. Waveshaping of sinusoids via hard clipping is investigated: the spectral enrichment due to the discontinuities introduced by its nonlinear transfer function maximizes aliasing distortion. The obtained results suggest that linear interpolation equalized with a high shelving filter is a sufficiently good method for upsampling. Concerning decimation, the interpolated FIR, elliptic, and cascaded integrator-comb filters all improve the results with respect to the trivial case. Regarding performance, the cascaded integrator-comb filter is the only tested decimation filter that achieves perceptually sufficient aliasing suppression for the entire frequency range when combined with the linear interpolator.

0 INTRODUCTION

Waveshaping is a classic nonlinear signal processing technique for audio synthesis and effects application [1–5]. Nonlinear processing enriches the spectrum of a signal by introducing harmonic distortion, a fact that can be problematic in digital audio. Waveshaping is generally performed after discretization, thus causing aliasing distortion when the new harmonics exceed the Nyquist limit, reflecting into the baseband spectrum. In audio signals, this produces unpleasant disturbances mainly due to the inharmonious relation of the mirrored spectrum with the original baseband content [6, 7].

Oversampling is, in principle, a straightforward approach for mitigating aliasing distortion [8, 9]. It is traditionally employed in the process of continuous- to discrete-time conversion, where raising the sampling frequency allows using a lower cutoff frequency in the anti-aliasing filter, as it expands the Nyquist interval and distances the spectrum replicas farther away from each other [10]. When dealing with nonlinear processing, oversampling reduces the amount of spectral content being reflected (the higher the oversampling factor, the lower the amount of spectral content being mirrored) and allows to low-pass filter the remaining high-frequency components. Its major drawback comes in the form of an increase of the computational load of the digital signal processing [11], which is directly proportional to the oversampling factor and the order of the low-pass filters used for interpolation and decimation. This aspect limits its applicability in computation-sensitive scenarios, such as in real-time processing [12].

Extensive research has been carried out aimed at mitigating the aliasing caused during the discrete-time synthesis of classic waveforms like those used in analog subtractive synthesis [6, 7, 13, 14]. For the case of nonlinear waveshaping, recent research has proposed the use of bandlimited correction functions [15, 16] and the antiderivative antialiasing method [5, 17]. These techniques outperform oversampling for waveshaping in terms of efficiency.

The interest in oversampling relies on the ease of its implementation and widespread recognition as a standard or baseline method [18, 19]. Additionally, oversampling is a flexible technique that can be used in connection with any nonlinear system, as it does not require knowledge of the antiderivative of the nonlinear shaping function or the location of signal samples affected by the nonlinear mapping, which are some of the requisites for the new antialiasing methods [5, 15–17]. The classic architecture of aliasing reduction for waveshaping is discussed in [18–22].

The theory behind the sampling of bandlimited signals and the generalized sampling theorems have been well understood for years [23]. Along these lines, Vaidyanathan describes the approaches and filter design methods related to multirate signal processing [10]. Nonetheless, to the authors' knowledge, no previous research has been carried out with the objective of evaluating filter types in oversampling schemes and finding the most suited ones for audio applications. Therefore, this work searches for the most suitable filter classes for aliasing reduction in oversampled static audio waveshaping. Both finite impulse response (FIR) and infinite impulse response (IIR) filters are studied.

This paper is organized as follows. Sec. 1 describes the formulation of waveshaping. Sec. 2 presents the studied filters, together with their design parameters. Sec. 3 tackles the evaluation and assessment methodology and gathers the obtained results. Finally, Sec. 4 concludes this paper.

1 FUNDAMENTALS OF OVERSAMPLED WAVESHAPING

The transformation waveshaping causes to the waveform of a signal can be mathematically expressed as

$$y\,[n] = w\,(x\,[n])\,, \tag{1}$$

where $w(\,)$ is a nonlinear waveshaping function that transforms the discrete-time input signal $x[n]$ to the discrete-time output signal $y[n]$ and n is the sample index. This function can be approximated through a polynomial power series:

$$w\,(x) = \sum_{k=0}^{N-1} w_k x^k = w_0 + w_1 x + w_2 x^2 + \cdots + w_{N-1} x^{N-1}, \tag{2}$$

where the coefficients with $k > 1$ are responsible for the nonlinear distortion, expanding the spectrum. In fact, for a trivial sinusoidal input, the exponent of the polynomial accounts for the multiplying factor of the expanded harmonics (being double the value for the squared-term, thrice the value for the third-order term, etc.).

When exceeding the Nyquist limit f_N, which is half of the sampling frequency f_s, the introduced harmonic components reflect, invading the original content in the passband. This scenario is illustrated in Fig. 1, where $X(f)$ is the spectrum of the distorted signal before sampling and $\widehat{X}(f)$ is the spectrum after sampling, displaying aliasing.

Oversampling is traditionally used to improve this scenario. Fig. 2 shows the block-diagram structure of a waveshaping function implemented using oversampling. The nonlinear function $f()$ is preceded by interpolation, effectively upsampling the input sequence by a factor of M, and low-pass filtering it to block components above the original Nyquist limit $f_s/2$, where f_s is the sampling frequency before oversampling. Decimation is featured at the end stage so as to return to the original sampling frequency (M-rate downsampling), where a low-pass filter is placed before it, again suppressing spectral components above $f_s/2$. The focus of this work is to search for the best filters to be used with this architecture.

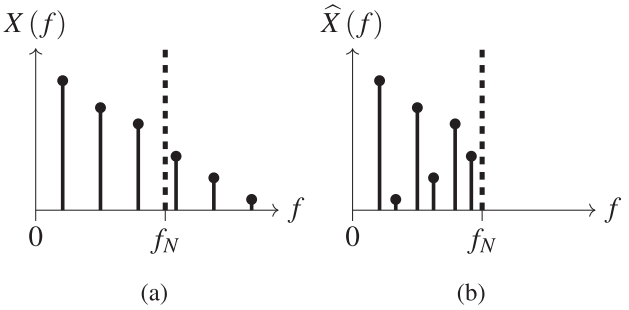

Fig. 1. Aliasing understood as mirroring alongside the Nyquist limit (f_N): (a) the generation of harmonics in the spectrum of a signal that exceed the Nyquist limit (dashed line) leads to (b) their reflection around the Nyquist limit, when sampled.

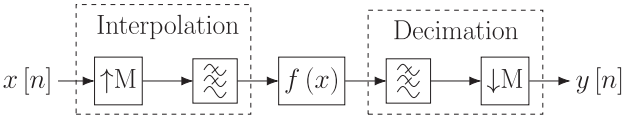

Fig. 2. Discrete-time waveshaping system with oversampling by factor M. Figure adapted from [20–22].

2 CANDIDATE FILTERS

In what follows, the filter classes subjected to testing are presented, as well as the optimization criteria that lead to their characteristic parameters. Additionally, the justification of the filters' choice and their final design values are gathered in this section.

2.1 Design Criteria

The base sampling frequency used in this study is $f_s = 44.1$ kHz and an oversampling factor of $M = 8$ is tested, given that it is generally sufficient for distortion-based audio processing [24]. All filters are designed according to the following criteria:

- The ripple of the cascaded interpolation and decimation filter responses is bounded by ± 1 dB in the passband, which has been selected to range from 0 Hz to 16 kHz. This is achieved by ensuring that the ripple in the individual passband of each interpolation and decimation filter is contained within a magnitude range of ± 0.5 dB.
- The stopband attenuation of the cascaded interpolation and decimation filters has a minimum value of 40 dB. The stopband begins at 28.1 kHz, which is the frequency that aliases to 16 kHz in the downsampling phase (as the Nyquist limit is $f_s/2 = 22.05$ kHz).

The choice of the passband's upper limit is based on the fact that the sensitivity of human hearing diminishes steeply at frequencies higher than 16 kHz [25]. Thus, it is assumed that the tolerance in the filter responses can be relaxed above that frequency.

2.2 Interpolation Filters

An elementary filter used in digital-to-analog converters is the zero-order hold (ZOH) [8], which gets its name from the action of maintaining (holding) a sample value for a

given time interval. It causes one of the simplest low-pass filtering actions in the spectrum, as its impulse response is defined as a discrete-time rectangular pulse of width M samples.

A first-order high-frequency shelving filter, which is a common equalization filter [26], can be used when necessary to flatten the passband, as dictated by the first design criterion. Its transfer function, derived in [26], can be written as

$$H_{HS}(z) = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1}}, \tag{3}$$

with the coefficients defined as

$$b_0 = \frac{\sqrt{G}\tan\left(\frac{\omega_c}{2}\right) + G}{\sqrt{G}\tan\left(\frac{\omega_c}{2}\right) + 1}, \tag{4}$$

$$b_1 = \frac{\sqrt{G}\tan\left(\frac{\omega_c}{2}\right) - G}{\sqrt{G}\tan\left(\frac{\omega_c}{2}\right) + 1}, \tag{5}$$

$$a_1 = \frac{\sqrt{G}\tan\left(\frac{\omega_c}{2}\right) - 1}{\sqrt{G}\tan\left(\frac{\omega_c}{2}\right) + 1}, \tag{6}$$

where G is the filter's linear gain, $\sqrt{G}$ is the linear midpoint gain (which corresponds to the geometric mean of the gains at extreme frequencies), and $\omega_c = 2\pi f_c/f_s$ is the crossover frequency.

Together with a scaling factor g of value 1.059, the shelving filter is used with gain $G = 1.242$ and crossover frequency $f_c = 16$ kHz to accommodate the passband of the ZOH in the magnitude range of ± 0.5 dB at frequencies below 16 kHz. The value of gain G is found using binary search. The high-frequency shelving filter coefficients are derived as $b_0 = 1.071$, $b_1 = 0.345$, and $a_1 = 0.416$. The magnitude responses of the ZOH interpolator without and with the shelving filter being applied are presented in Fig. 3. Note that when the upsampling factor is $M = 8$, the nominal passband gain of the interpolator must be 8.0 (i.e., 18.06 dB) in order for the signal power to be restored after interpolation.

Oversampling schemes with the ZOH perform poorly, as can be seen in Sec. 3. Hence, the linear interpolator [8, 9] is considered next, which obtains its name from its action at the interpolation stage, resulting in the introduction of values that would lie in a line fitted between consecutive non-zero samples. Its impulse response is defined as a discrete-time triangular pulse of width $2M - 1$ samples:

$$h[n] = \begin{cases} 1 - \dfrac{|n|}{M}, & \text{if } |n| < M \\ 0, & \text{otherwise} \end{cases}. \tag{7}$$

Compared to more complex systems, such as cubic or higher order interpolators, the linear interpolator offers a compromise between design and complexity: its easy implementation comes at the expense of distortion in the passband and modest attenuation of the wide sidelobes located in the stopband. Despite its constraints, oversampling schemes yield good results with it (cf., Sec. 3). Given its performance, simplicity, and low cost of implementation, more complex systems are not considered.

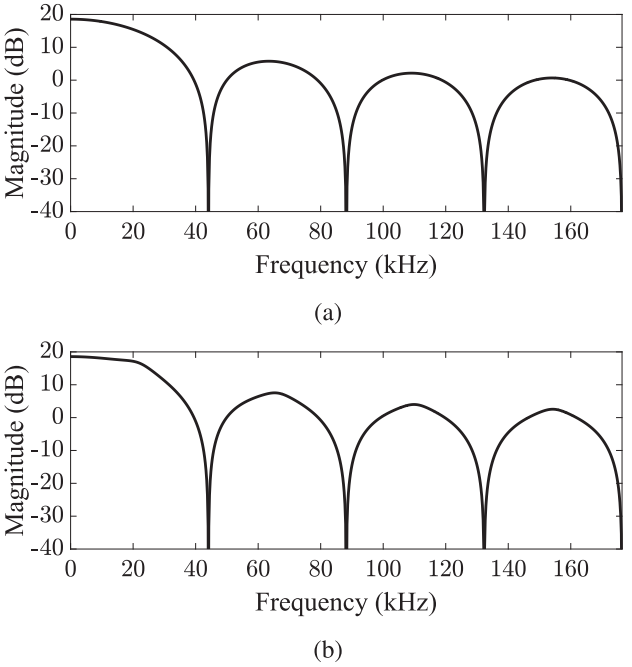

Fig. 3. Magnitude response of the (a) non-processed and (b) equalized ZOH interpolator for the oversampling factor $M = 8$, including the interpolator gain of 18 dB.

A high shelving filter with gain $G = 1.941$, crossover frequency $f_c = 16$ kHz, and a linear gain term $g = 1.059$ is used to equalize the linear interpolation filter to bound the passband in the magnitude range of ± 0.5 dB at frequencies below 16 kHz. The filter coefficients are thus $b_0 = 1.234$, $b_1 = 0.270$, and $a_1 = 0.504$. The magnitude responses of the linear interpolator without and with the shelving filter being applied are shown in Fig. 4.

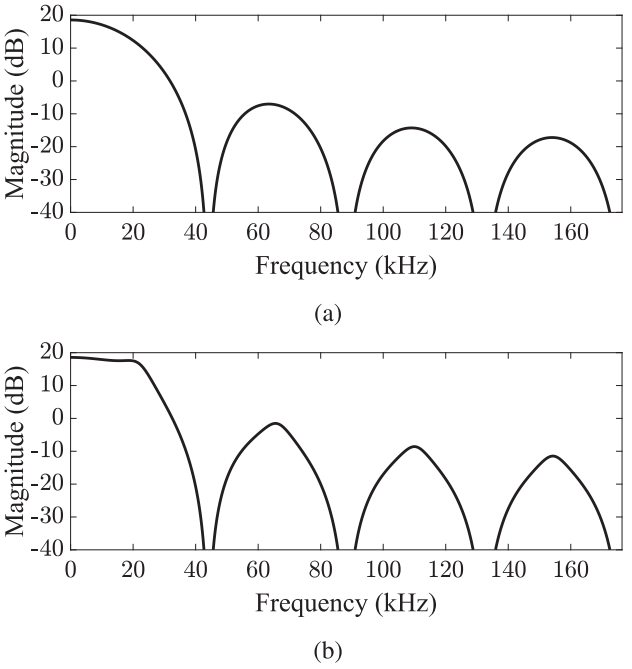

Fig. 4. Magnitude response of the (a) non-processed and (b) equalized linear interpolator ($M = 8$), cf., Fig. 3.

2.3 Decimation Filters

The choice of decimation filters is carried out by selecting those with a promising computational efficiency. They are classified according to the nature of their impulse response as FIR and IIR filters.

2.3.1 FIR Filters for Decimation

Standard FIR filters are known to be rather inefficient for implementing narrow low-pass filters. For this reason, they are not considered in this work. As an alternative, the Interpolated FIR (IFIR) filter [10, 27] is one of the candidate methods studied in this work. It is achieved by cascading an interpolated stretched periodic filter with an image-suppressor, eventually leading to an efficient approximation of a desired narrowband low-pass filter: stretching the cut-off frequency of the desired response relaxes the design specifications.

The IFIR filter is designed with an interpolation factor $\ell = 4$, band edge frequencies $\Delta f = [16, 28.1]$ kHz, maximum ripple $\Delta R = [-19.47, -30]$ dB, and a scaling factor $g = 1.023$ to achieve the passband tunnel of ± 0.5 dB and minimum cascaded stopband attenuation of 40 dB. The orders of the periodic filter and image-suppressor are $N_1 = 11$ and $N_2 = 13$, respectively.

The IFIR filter can be implemented as a multistage stucture in which the image suppressor comes first and runs at the rate of Mf_s and the periodic filter gets downsampled to reduce the number of operations. We decimate in two stages so that downsampling by factor 4 takes place first, which is equal to the interpolation factor of $\ell = 4$ used in the IFIR design. The periodic filter can be then implemented as a polyphase structure: every second output is computed while for every second input sample, the multiplications and additions are not executed, being the samples only run along the delay line. This polyphase process effectively undertakes the decimation by two without any extra operations. Thus, only the image suppressor of the IFIR filter runs at the sample rate Mf_s; the periodic filter runs, in practice, at f_s, which leads to significant computational savings.

Both subfilters of the IFIR structure are linear-phase FIR filters, i.e., their coefficient arrays are symmetric with respect to the center point. This property can also be used for saving more resources by first adding any two delayed input samples that are multiplied by the same coefficient value and then multiplying this sum only once. By applying this, the number of multiplications gets reduced by 50% when there is an even number of coefficients, or nearly 50% when the number of coefficients is odd and the center coefficient does not have an equal pair. Nevertheless, the exploitation of the symmetry does not affect the number of additions in the IFIR filter implementation.

2.3.2 Elliptic IIR Decimation Filter

The linear-phase property of digital filters is usually not required in audio processing [26]. In fact, linear-phase filters are sometimes even considered harmful as their symmetric impulse response can produce pre-ringing and an unnecessary delay [26, 28, 29]. Therefore, minimum-phase IIR filters can be considered for oversampled waveshaping.

The first IIR filter to be studied is the elliptic filter [8, 9], also known as the Cauer or Zolotarev filter, whose derivation is carried out by means of polynomial approximation design, using the elliptic (Cauer) approximation. It offers a quick rejection with the lowest possible polynomial order at the cost of introducing spectral ripple throughout the magnitude spectrum.

For the elliptic IIR filter, the cutoff frequency f_c is selected to be 16 kHz. The passband peak-to-peak ripple R_p and stopband attenuation R_s are chosen to be 1 dB and 28 dB, respectively, and a scaling factor $g = 1.059$ is used so as to both accommodate the passband to the ± 0.5 dB magnitude range and achieve the minimum cascaded stopband attenuation of 40 dB. These requirements can be fulfilled with an elliptic IIR filter of order $N = 3$, which has four parameters in both the numerator and the denominator of its transfer function.

2.3.3 CIC IIR Decimation Filter

The Cascaded Integrator-Comb (CIC) filter [30, 31] is chosen next, which was originally proposed in [30] as an efficient class of narrowband low-pass filters. Its architecture consists of a cascaded set of integrators and comb filters, where the number of integrator and comb sections determines the order of the filter. Moreover, the arrangement of the stages regulates the action of the filter: integrator-comb for decimation, and vice-versa for interpolation. Theoretically, the avoidance of multipliers enables for economic implementations; nonetheless, in practical scenarios (e.g., for stable real-time implementations), two multipliers are placed at the integrator and comb stages, respectively (leaky CIC filter), whose values approach 1 (and are equal to $1 - \varepsilon$, where ε has a value close to 0) [13].

The integrator section functions at the oversampled rate Mf_s, where each of the N integrators is implemented as a one-pole filter with the z-domain transfer function

$$H_{\text{int}}(z) = \frac{1}{1 - (1 - \varepsilon) z^{-1}}. \tag{8}$$

On the other hand, the comb section's transfer function can be described as

$$H_{\text{comb}}(z) = 1 - \left((1 - \varepsilon)^{MD} z^{-MD}\right), \tag{9}$$

where each of the N comb filters introduces a differential delay of MD samples. The total response of the CIC filter is achieved by cascading the integrator and comb sections' responses:

$$H(z) = H_{\text{int}}^{N}(z)\, H_{\text{comb}}^{N}(z) \tag{10}$$

$$= \frac{\left(1 - \left((1 - \varepsilon)^{MD} z^{-MD}\right)\right)^{N}}{\left(1 - (1 - \varepsilon) z^{-1}\right)^{N}}. \tag{11}$$

When decimating, the comb filter presented in Eq. (9) can be placed after downsampling so that it runs at the reduced sampling rate of f_s. This diminishes the computational cost

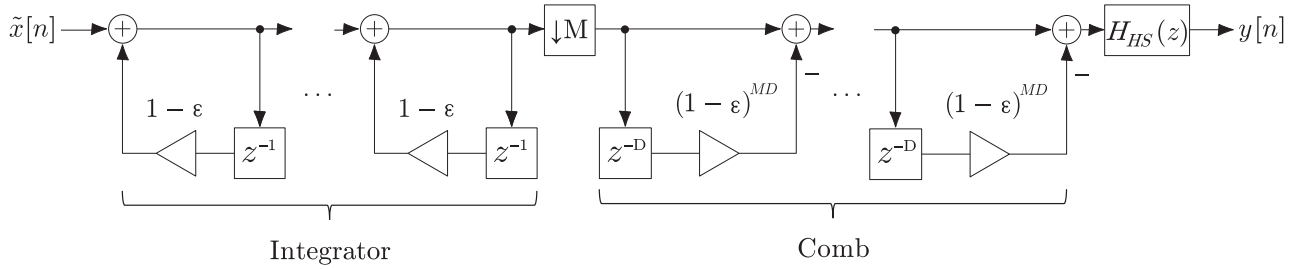

Fig. 5. Architecture of the decimation block when considering the CIC filter, which consists of a leaky CIC decimation filter cascaded with a first-order high shelving filter: $\tilde{x}[n]$ is the signal that enters the decimation filter (after having been interpolated and waveshaped) and $y[n]$ is the resulting output. Figure based on [22, 26, 30, 31].

significantly, as only half of the CIC structure must run at the oversampled rate of Mf_s.

The overall frequency response displays a low-pass filter behavior with nulls at integer multiples of the inverse of the sampling frequency f_s/D, where the differential delay D controls the position of the narrow rejection bands (normally taking values of $D = 1$ or $D = 2$). The null areas land at aliasing bands, and as imaging takes place around them, high mirroring rejection is obtained. In this work, the differential delay of the CIC filter is set to $D = 1$ so that the narrow rejection bands land at exact multiples of the sampling frequency, and the multiplier subtrahend ε is set to 2^{-15}, the smallest number available in 16-bit signed integer format.

The CIC filter is related to discrete B-splines as they are based on the same principle of convolving a basis function by itself [32]. When used for interpolation, discrete B-splines require, in theory, a recursive anticausal prefilter to flatten their magnitude response [33]. However, in this work, a first-order high shelving filter is used to flatten the passband of the CIC filter as it suffices to fulfill the passband constraint. The crossover frequency is set to $f_c = 16$ kHz, the linear gain value $G = 12.710$ is selected, and the scaling factor is chosen to be $g = 3.861 \times 10^{-6}$, both accommodating the passband to the ± 0.5 dB magnitude range and achieving the minimum cascaded stopband attenuation of 40 dB. The obtained filter coefficients are $b_0 = 0.892 \times 10^{-5}$, $b_1 = -0.216 \times 10^{-5}$, and $a_1 = 0.651$. The decimation architecture using leaky CIC filters and the shelving filter is shown in Fig. 5.

2.4 Final Compound Structure

The decimation filter orders and operation counts can be seen in Table 1, where the first-order shelving filtering that the CIC filter undergoes is already taken into account. Based on the general oversampling structure presented in Fig. 2, the pursued overall architecture can be seen in Fig. 6. The decimation filter comprises either the IFIR filter, the elliptic filter, or the CIC filter equalized with a high shelving filter. The last case is further detailed in Fig. 5.

When multiplying the interpolators' and decimators' responses, conceptualizing the waveshaping arrangement as a linear cascading, the passband tunnel gets restricted to a range of ± 1 dB. Fig. 7 shows the combined magnitude responses obtained when cascading the linear interpolator with each of the decimation filters. In all cases, the attenuation at frequencies above 28 kHz is at least 40 dB, and it

Table 1 Number of operations of the tested filters, where ADD and MUL stand for the number of additions and multiplications, respectively, M is the oversampling factor, N represents the filter order of the elliptic and CIC filters, N_1 is the filter order of the IFIR periodic filter, and N_2 is the filter order of the IFIR image suppressor.

Filter type	ADD	MUL	Ops (sum)	Total load
IFIR	$N_1 + MN_2$	$\lfloor \frac{N_1}{2} \rfloor + M \lfloor \frac{N_2}{2} \rfloor + M + 1$	$N_1 + MN_2 + \lfloor \frac{N_1}{2} \rfloor + M \lfloor \frac{N_2}{2} \rfloor + M + 1$	177 ($M = 8, N_1 = 11, N_2 = 13$)
Elliptic	$2MN$	$M(2N + 1)$	$M(4N + 1)$	104 ($M = 8, N = 3$)
CIC	$MN + N + 2$	$MN + N + 3$	$2MN + 2N + 5$	113 ($M = 8, N = 6$)

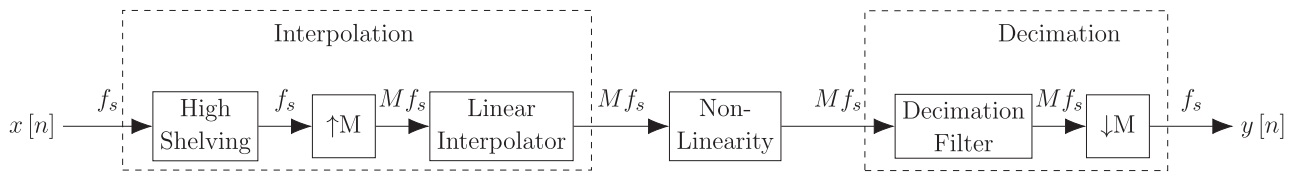

Fig. 6. Oversampling scheme for waveshaping. The decimation block consists of either a single FIR or IIR filter followed by downsampling, or a CIC filter cascaded with a high shelving filter with downsampling taking place between the integrator and comb stages. For the latter case, the resulting decimation configuration is specified in more detail in Fig. 5.

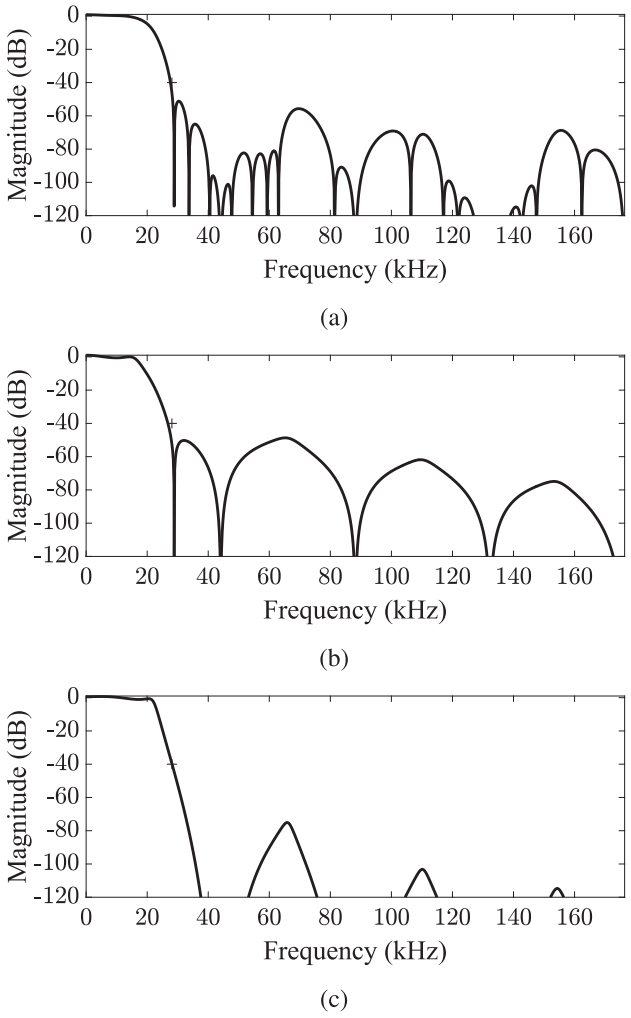

Fig. 7. Cascaded magnitude responses of the linear interpolator with the (a) IFIR, (b) elliptic, and (c) CIC decimation filters for the oversampling factor $M = 8$. The corner of the stopband at 28.1 kHz and at -40 dB is indicated with a cross (+).

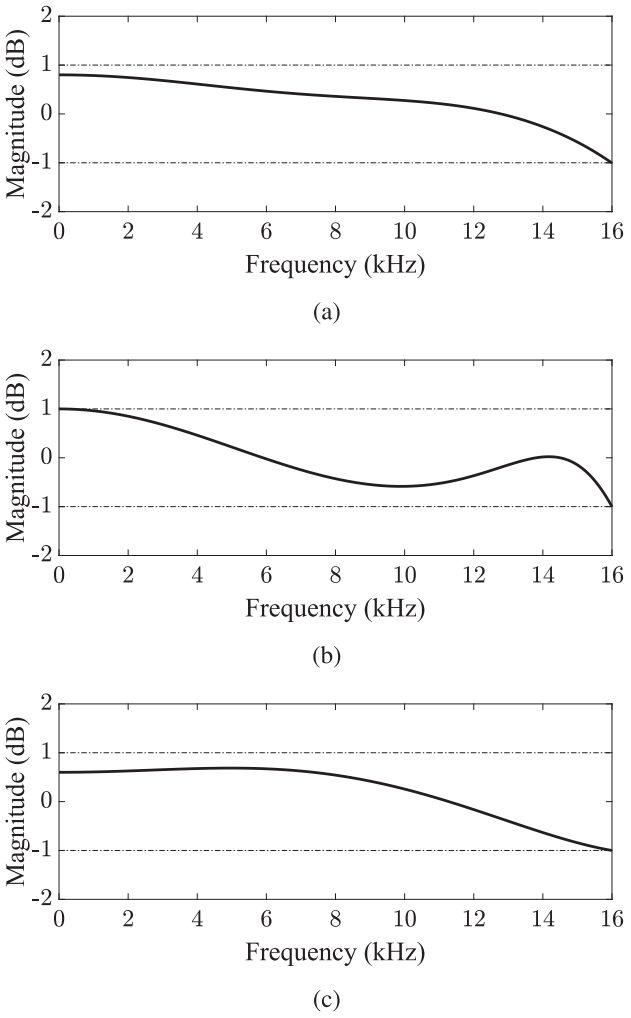

Fig. 8. Passband details of Fig. 7. The ± 1-dB tolerances are indicated with dashed lines.

is seen that at high frequencies, it is generally much better than that. Fig. 8 further zooms into their passbands, where the ± 1-dB restriction can be appreciated.

3 PERFORMANCE COMPARISON

In order to assess the aliasing suppression achieved with oversampling using different filters, synthesized sinusoids covering the full range of fundamental frequencies that a piano can produce (ranging from 27.5 Hz to 4186 Hz [14]) are passed through a hard clipper. For this work, only saturating nonlinearities are considered. Thus, the hard-clipping nonlinear shaping function is selected due to its aggressive spectral-enhancement action on the input signals [15, 16, 34]. The nonlinear transfer function of the hard clipper is shown in Fig. 9, where the threshold level is set to a fairly small value of 0.1 to further increase its spectral enriching.

Even though the signal-to-noise ratio (SNR) is a straightforward and well-known quality measure, it does not generally correlate with the perception of aliasing, hence causing the need for more accurate methods [35–37]. As an alternative, the noise-to-mask ratio (NMR) takes into account psychoacoustic masking as present in human hearing [35]: it is defined as the energy ratio between a filtered version of the distorted signal, containing only the aliased

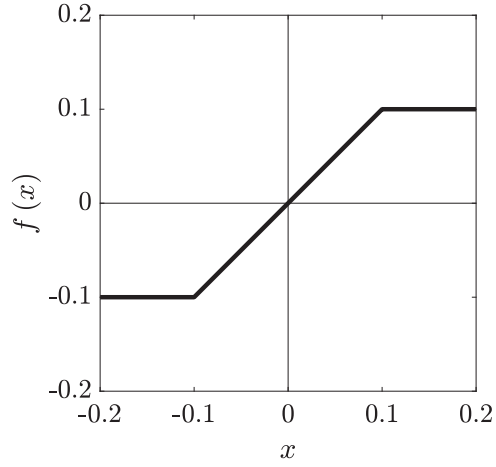

Fig. 9. Hard-clipping waveshaper transfer function with the threshold parameter set to 0.1.

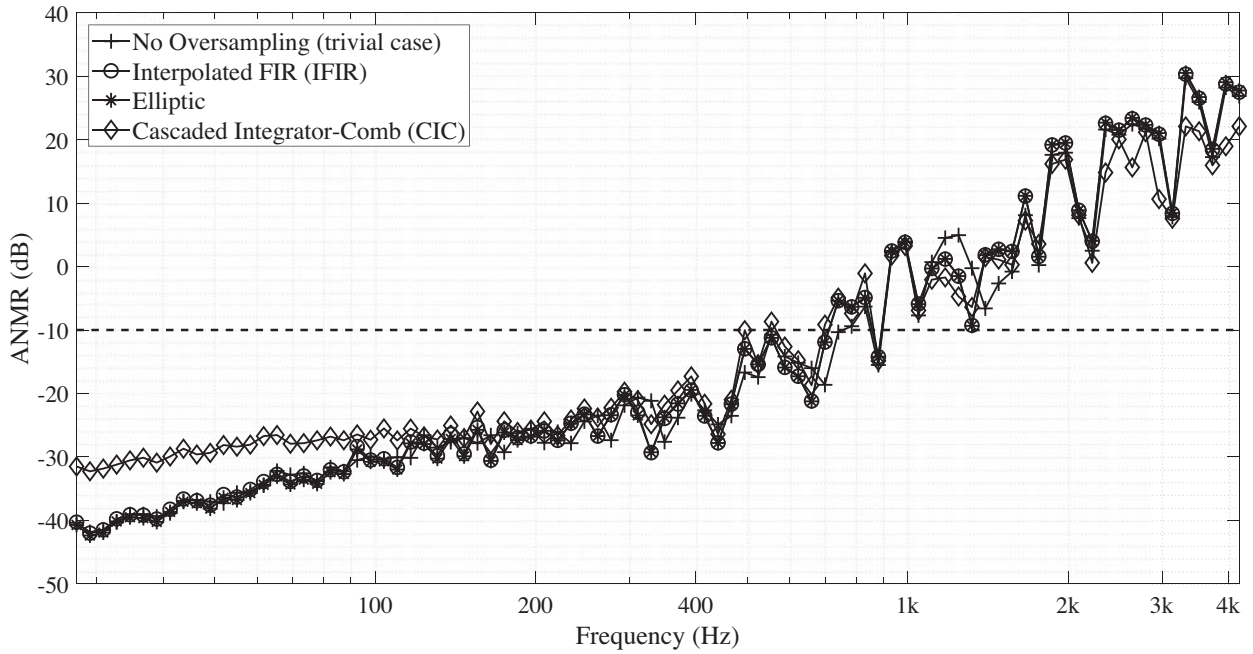

Fig. 10. Performance results in terms of ANMR for audio waveshaping (hard clipping) with the ZOH and an oversampling factor of $M = 8$.

components (i.e., noise), and the simplified masking threshold of the distorted sequence without aliasing [7, 35–37].

In order to obtain a signal containing only the aliased spectral content, a one-second fragment of the distorted sequence is windowed with a Chebyshev window with 120 dB of attenuation [14]. The magnitude and phase values at the integer multiples of the fundamental frequency are used to additively synthesize an ideal sequence, containing only the desired harmonic content; the squared sum of its sample values is equal to its energy (Parseval's theorem):

$$E = \frac{1}{N} \sum_{n=0}^{N-1} |x\,[n]|^2. \tag{12}$$

The aliased spectral content is isolated by subtracting the harmonically synthesized ideal-spectrum signal from the distorted signal, scaling the former so as to ensure an equal magnitude of the first harmonic in both sequences [14].

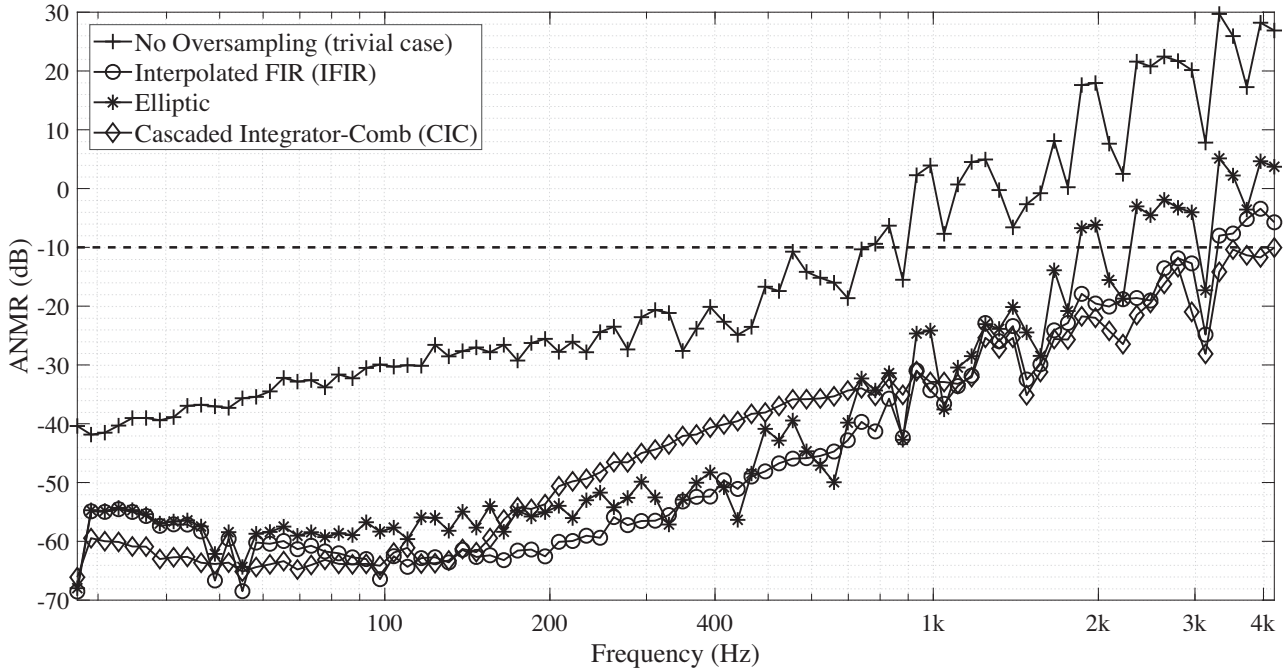

Fig. 11. Performance in terms of ANMR for oversampled waveshaping (hard clipping) with linear interpolation ($M = 8$).

The residual energy value is then obtained via Parseval's theorem.

The NMR was initially designed to flag audible disturbances with positive values and signal inaudibility with negative values, but further listening tests confirmed that a more accurate threshold for separating the regions of audible and inaudible aliasing should be set at -10 dB. In order to fine-tune the results and fit them better to human hearing, A-weighting of the signals can take place before computing the NMR, which leads to the A-weighted NMR (ANMR) [7].

The ANMR outcome with hard clipping for eight-times oversampling with the ZOH and linear interpolator is gathered in Figs. 10 and 11, respectively. From the obtained results, one can see how the linear interpolator outperforms the ZOH given that the latter does not improve the ANMR values with respect to the trivial case, whereas the former does: it is seen in Fig. 10 that the ZOH prevents the oversampling technique from performing well. With a ZOH, aliasing distortion starts becoming audible at fundamental frequencies higher than 554 Hz.

Regarding decimation, all tested filters (IFIR, elliptic IIR, and CIC) improve the results with respect to the trivial case when combined with linear interpolation. When analyzing performance and computational cost in decimation, the CIC filter is the only tested filter that achieves perceptually sufficient aliasing suppression for the whole frequency range (cf., Fig. 11).

4 CONCLUSION

In this work, oversampling has been studied for audio waveshaping via hard clipping, selected due to its vulnerability to aliasing, given that it constitutes an aggressive case of spectral-enriching nonlinear waveshaping. The classic topologies featuring upsampling and decimation for waveshaping have been evaluated for the whole fundamental frequency range of the piano (27.5 Hz to 4186 Hz) at eight-times oversampling, studying a collection of optimized digital filters.

When interpolating, the linear interpolator combined with a high shelving filter improves the trivial case and provides good results, discouraging the use of more complex options. In contrast, the ZOH generates unsatisfactory ANMR values, similar to the ones obtained when applying waveshaping without oversampling, hence causing its dismissal.

Regarding decimation, the results show that all tested decimation filters reduce aliasing when compared to the trivial case (when combined with equalized linear interpolation). In terms of performance, the CIC filter is the best choice among the tested filter structures: it is the only tested decimation filter that achieves perceptually sufficient aliasing suppression for the whole frequency range of the piano.

Oversampling by a factor of $M = 8$ was studied in this work given that it ensures sufficient antialiasing, even when tackling the most aggressive nonlinear waveshapers. With less distorting nonlinearities, it may be perceptually acceptable to use a lower oversampling factor, such as $M = 2$ or $M = 4$. In such cases, the linear interpolation filter combined with a high shelf will probably remain a good choice for upsampling; nonetheless, the order selection, parameter tuning, and comparison of decimation filters will need to be re-evaluated, and thus, its design is left for future work.

5 ACKNOWLEDGMENT

This research was part of the activities of the "Nordic Sound and Music Computing Network — NordicSMC," NordForsk project number 86892. The work of F. Esqueda was supported by the Aalto ELEC Doctoral School.

6 REFERENCES

[1] M. Russ, *Sound Synthesis and Sampling*, 3rd ed. (Focal Press, Oxford, UK, 2008), https://doi.org/10.4324/9780080926957.

[2] M. Le Brun, "Digital Waveshaping Synthesis," *J. Audio Eng. Soc.*, vol. 27, pp. 250–266 (1979 Apr.).

[3] C. Roads, "A Tutorial on Non-linear Distortion or Waveshaping Synthesis," *Computer Music J.*, vol. 3, no. 2, pp. 29–34 (1979 Jun.).

[4] M. Puckette, *The Theory and Technique of Electronic Music* (World Scientific, Singapore, 2007), https://doi.org/10.1142/6277.

[5] J. D. Parker, V. Zavalishin, and E. Le Bivic, "Reducing the Aliasing of Nonlinear Waveshaping Using Continuous-Time Convolution," *Proc. 19th Int. Conf. on Digital Audio Effects (DAFx-16)*, pp. 137–144 (Brno, Czech Republic) (2016 Sep.).

[6] V. Välimäki and A. Huovilainen, "Antialiasing Oscillators in Subtractive Synthesis," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 116–125 (2007 Mar.), https://doi.org/10.1109/MSP.2007.323276.

[7] H.-M. Lehtonen, J. Pekonen, and V. Välimäki, "Audibility of Aliasing Distortion in Sawtooth Signals and its Implications for Oscillator Algorithm design," *J. Acoust. Soc. Amer.*, vol. 132, no. 4, pp. 2721–2733 (2012 Oct.), https://doi.org/10.1121/1.4748964.

[8] A. Oppenheim, A. Willsky, and S. Nawab, *Signals and Systems*, Prentice Hall Signal Processing Series, 2nd ed. (Prentice Hall, Upper Saddle River, NJ, 1996).

[9] D. Manolakis and V. Ingle, *Applied Digital Signal Processing: Theory and Practice* (Cambridge University Press, Cambridge, UK, 2011), https://doi.org/10.1017/CBO9780511835261.

[10] P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, NJ, 1992).

[11] D. Mapes-Riordan, "A Worst-Case Analysis for Analog Quality (Alias-Free) Digital Dynamics Processing," *J. Audio Eng. Soc.*, vol. 47, pp. 948–952 (1998 Nov.).

[12] J. Schimmel, "Audible Aliasing Distortion in Digital Audio Synthesis," *Radioengineering*, vol. 21, no. 1, pp. 56–62 (2012 Apr.).

[13] T. Stilson and J. Smith, "Alias-Free Digital Synthesis of Classic Analog Waveforms," *Proc. Int. Computer Music Conf. (ICMC)*, pp. 332–335 (Hong Kong) (1996 Aug.).

[14] V. Välimäki, "Discrete-Time Synthesis of the Sawtooth Waveform with Reduced Aliasing," *IEEE Signal Process. Lett*., vol. 12, no. 3, pp. 214–217 (2005 Mar.), https://doi.org/10.1109/LSP.2004.842271.

[15] F. Esqueda, V. Välimäki, and S. Bilbao, "Rounding Corners with BLAMP," *Proc. 19th Int. Conf. on Digital Audio Effects (DAFx-16)*, pp. 121–128 (Brno Czech Republic) (2016 Sep.).

[16] F. Esqueda, S. Bilbao, and V. Välimäki, "Aliasing Reduction in Clipped Signals," *IEEE Trans. Signal Process*., vol. 64, no. 20, pp. 5255–5267 (2016 Oct.), https://doi.org/10.1109/TSP.2016.2585091.

[17] S. Bilbao, F. Esqueda, J. D. Parker, and V. Välimäki, "Antiderivative Antialiasing for Memoryless Nonlinearities," *IEEE Signal Process. Lett*., vol. 24, no. 7, pp. 1049–1053 (2017 Jul.), https://doi.org/10.1109/LSP.2017.2675541.

[18] J. Pakarinen and D. T. Yeh, "A Review of Digital Techniques for Modeling Vacuum-Tube Guitar Amplifiers," *Computer Music J.*, vol. 33, no. 2, pp. 85–100 (2009 Jun.), https://doi.org/10.1162/comj.2009.33.2.85.

[19] B. De Man and J. D. Reiss, "Adaptive Control of Amplitude Distortion Effects," presented at the *AES 53rd Int. Conf.: Semantic Audio* (2014 Jan.), conference paper P2-9.

[20] H. Thornburg, "Antialiasing for Nonlinearities: Acoustic Modeling and Synthesis Applications," *Proc. Int. Computer Music Conf. (ICMC)*, pp. 66–69 (Beijing, China) (1999 Oct.).

[21] J. Schattschneider and U. Zölzer, "Discrete-Time Models for Nonlinear Audio Systems," *Proc. 2nd COST G-6 Workshop on Digital Audio Effects (DAFx-99)*, pp. 9–12 (Trondheim, Norway) (1999 Dec.).

[22] P. Dutilleux, K. Dempwolf, M. Holters, and U. Zölzer, "Nonlinear Processing," *DAFX: Digital Audio Effects*, 2nd ed., pp. 101–138 (Wiley, Chichester, UK, 2011), https://doi.org/10.1002/9781119991298.ch4.

[23] M. Unser, "Sampling—50 Years After Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587 (2000 Apr.), https://doi.org/10.1109/5.843002.

[24] D. T.-M. Yeh, *Digital Implementation of Musical Distortion Circuits by Analysis and Simulation*, Ph.D. thesis, Stanford University, Stanford, CA (2009 Jun.).

[25] K. Ashihara, "Hearing Thresholds for Pure Tones Above 16 kHz," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. EL52–EL57 (2007 Sep.), https://doi.org/10.1121/1.2761883.

[26] V. Välimäki and J. D. Reiss, "All About Audio Equalization: Solutions and Frontiers," *Appl. Sci.*, vol. 6, no. 5, p. 129 (2016 May), https://doi.org/10.3390/app6050129.

[27] Y. Neuvo, D. Cheng-Yu, and S. Mitra, "Interpolated Finite Impulse Response Filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 3, pp. 563–570 (1984 Jun.), https://doi.org/10.1109/TASSP.1984.1164348.

[28] C. H. Slump, C. G. M. van Asma, J. K. P. Barels, W. J. A. Brunink, F. B. Drenth, J. V. Pol, D. Schouten, M. M. Samsom, and O. E. Herrmann, "Design and Implementation of a Linear-Phase Equalizer in Digital Audio Signal Processing," *Proc. Workshop on VLSI Signal Process*., pp. 297–306 (Napa Valley, CA) (1992 Oct.), https://doi.org/10.1109/VLSISP.1992.641062.

[29] A. Mäkivirta, J. Liski, and V. Välimäki, "Modeling and Delay-Equalizing Loudspeaker Responses," *J. Audio Eng. Soc.*, vol. 66, pp. 922–934 (2018 Nov.), https://doi.org/10.17743/jaes.2018.0053.

[30] E. Hogenauer, "An Economical Class of Digital Filters for Decimation and Interpolation," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 2, pp. 155–162 (1981 Apr.), https://doi.org/10.1109/TASSP.1981.1163535.

[31] R. Lyons, *Understanding Digital Signal Processing*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ, 2004).

[32] D. Babic and M. Renfors, "Efficient Structure for Conversion Between Arbitrary Sampling Rates Power," *IEEE Signal Process. Lett.*, vol. 12, no. 1, pp. 1–4 (2005 Jan.), https://doi.org/10.1109/LSP.2004.838193.

[33] M. Unser, "Splines: A Perfect Fit for Signal and Image Processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38 (1999 Nov.), https://doi.org/10.1109/79.799930.

[34] F. Esqueda, H. Pöntynen, J. Parker, and S. Bilbao, "Virtual Analog Models of the Lockhart and Serge Wavefolders," *Appl. Sci.*, vol. 7, no. 12, p. 1328 (2017 Dec.), https://doi.org/10.3390/app7121328.

[35] K. Brandenburg, "Evaluation of Quality for Audio Encoding at Low Bit Rates," presented at the *82nd Conv. of the Audio Eng. Soc*. (1987 Mar.), conv. paper 2433.

[36] J. Herre, E. Eberlein, H. Schott, and C. Schmidmer, "Analysis Tool for Real Time Measurements Using Perceptual Criteria," presented at the *AES 11th Int. Conf.: Test & Measurement* (1992 May), conf. paper 11-021.

[37] K. Brandenburg and T. Sporer, "NMR and Masking Flag: Evaluation of Quality using Perceptual Criteria," presented at the *AES 11th Int. Conf.: Test & Measurement* (1992 May), conf. paper 11-020.

THE AUTHORS

Julen Kahles

Fabián Esqueda

Vesa Välimäki

Julen Kahles was born in Baden-Baden, Germany, in 1994. He received the B.Sc. degree in telecommunication technology engineering from the Public University of Navarre, Pamplona, Spain, in 2016, and the Professional Degree in classical music performance from the Professional Music Conservatory "Pablo Sarasate" of Navarre, Pamplona, Spain, in 2013. In January 2019, he received the M.Sc. degree in computer, communication and information sciences from Aalto University, Espoo, Finland.

•

Fabián Esqueda was born in Mexico City, Mexico, in 1989. He received the B.Eng. degree in electronic engineering with music technology systems from the University of York, York, UK, in 2012, and the M.Sc. degree in acoustics and music technology from the University of Edinburgh, Edinburgh, UK, in 2013. In 2014–2018 he was a doctoral candidate in the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland. In June 2018, he received his Doctor of Science in Technology degree from Aalto University with a dissertation on aliasing suppression techniques for nonlinear audio signal processing and virtual analog modeling of music circuits. He has authored several publications on aliasing suppression techniques, virtual analog modeling of nonlinear circuits, and gray-box modeling of time-varying audio circuits.

•

Vesa Välimäki is Professor of audio signal processing at Aalto University, Espoo, Finland. He is the Vice Dean for Research in the Aalto University School of Electrical Engineering. He received his M.Sc. in Technology and Doctor of Science in Technology degrees in electrical engineering from the Helsinki University of Technology (TKK) in 1992 and 1995, respectively. His doctoral dissertation dealt with fractional delay filters and physical modeling synthesis. In 1996, he was a postdoctoral research fellow at the University of Westminster, London, UK. In 2001–2002, he was the professor of signal processing at the Pori unit of the Tampere University of Technology, Pori, Finland. In 2006–2007, he was the head of the TKK Laboratory of Acoustics and Audio Signal Processing. In 2008–2009, he was a visiting scholar at the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA. His research interests include digital filter design, nonlinear audio processing, sound synthesis, and artificial reverberation algorithms. Prof. Välimäki is a Fellow of the AES and a Fellow of the Institute of Electrical and Electronics Engineers (IEEE). He is a Life Member of the Acoustical Society of Finland. He was the general chair of the 11th International Conference on Digital Audio Effects DAFx-08 in 2008 and the 14th International Conference on Sound and Music Computing SMC-17 in 2017.