Antiderivative Antialiasing for Memoryless Nonlinearities

Stefan Bilbao, Senior Member, IEEE, Fabián Esqueda, Graduate Student Member, IEEE, Julian D. Parker, and Vesa Välimäki, Fellow, IEEE

Abstract—Aliasing is a commonly-encountered problem in audio signal processing, particularly when memoryless nonlinearities are simulated in discrete time. A conventional remedy is to operate at an oversampled rate. A new antialiasing reduction method is proposed here for discrete-time memoryless nonlinearities, which is suitable for operation at reduced oversampling rates. The method is based on higher order antiderivatives of the nonlinear function used. The first antiderivative form of the new method is equivalent to a technique proposed recently by Parker et al. The second and third order cases of the new method offer considerable improvement over the first antiderivative method, improving the signal-to-noise ratio in comparison to a straightforward implementation. The proposed methods can be implemented with fewer operations than oversampling and are applicable to discrete-time modeling of a wide range of nonlinear system.

Index Terms—Aliasing, harmonic distortion, nonlinear systems, signal denoising, signal processing algorithms.

I. INTRODUCTION

Aliasing is a fundamental problem in nonlinear signal processing. When a digital signal undergoes a nonlinear operation, its bandwidth is expanded, leading to the spurious mirroring of components back to the baseband. Such aliasing behavior is particularly problematic in audio applications. This letter proposes a general approach to aliasing suppression suitable for the simulation of memoryless nonlinearities in discrete time.

A commonly used general method to reduce aliasing in memoryless nonlinearities is oversampling [1], [2], [3], [4], [5], [6], [7]. In audio applications, the input signal is typically upsampled by a factor of 8 or 16, for example, using an appropriate interpolation filter. When the nonlinear function is applied at the oversampled rate, distortion components will appear at frequencies well above the original Nyquist limit. Under a downsampling/lowpassing operation, such components are suppressed, and aliasing will be negligible. The main disadvantage of oversampling is the increase in the operation count, which is proportional to the oversampling factor. Additionally, the interpolation and decimation filters add to the workload per sample.

Other approaches to antialiasing have been proposed for memoryless nonlinearities. Schattenschneider and Zölzer [8] introduced a harmonic mixer model for polynomial nonlinearities of finite order. In the harmonic mixer, the nonlinearity is divided into a parallel structure in which each branch has a nonlinear function of a single order only and the branch input signals are lowpass filtered to assure that no aliasing can occur [8], [5]. Fernández-Cid and Casajús-Quirós modified this method to a filterbank form [9]. Thornburg suggested that a memoryless nonlinear function could be approximated with a lower-order polynomial to reduce aliasing [1]. However, these methods only apply for polynomial functions, but many of the useful nonlinear functions, such as the hard clipper [10], [2], [11], [12] and saturating functions [13], [14], [15], [16], [17], [18], cannot be approximated well with a polynomial. Recent work has applied bandlimited correction functions commonly used in oscillator synthesis to reduce the aliasing introduced by hard clipping and rectification [19], [20], [21].

Antialiasing methods of a fundamentally different character have been recently proposed by Parker et al. [22]. This method is based on approximating the underlying continuous-time input signal with a piecewise linear function, applying a nonlinear function to it, and convolving the resulting signal with the continuous-time impulse response of a lowpass filter.

In this letter, we present a new discrete-time aliasing reduction method for memoryless nonlinearities, based on discrete differentiation of higher order antiderivatives. The first-order method is equivalent to that in [22]. However, the use of higher antiderivatives leads to increasing levels of aliasing suppression; such methods are distinct from e.g. the second-order method presented in [22]. Such methods are applicable regardless of the particular form of the nonlinear function.

The proposed idea of differentiating antiderivatives is related to previous antialiasing synthesis methods called the differential polynomial waveform [23], [24], [25], [26], [27], and integrated wavetable/sampling synthesis [28], [29]. In these synthesis methods, the waveform is first integrated w.r.t. time one or more times, either analytically [23], [25], [26] or using an integrating digital filter [28], [29], [30], and is finally differentiated as many times. This helps remarkably to reduce aliasing in signal synthesis, but such methods are not directly applicable to nonlinear processing of arbitrary signals.

This letter is organized as follows. Sec. II discusses the context of this work and the first-order antiderivative antialiasing method [22], which is the starting point for this work. Sec. III derives the new method using higher order antiderivatives of the nonlinear function. Sec. IV evaluates the proposed method.
in terms of signal-to-noise ratio and compares it with the
trivial, oversampled, and first-order methods using a hard-
clipping and a hyperbolic tangent function as examples. Sec. V
concludes this letter and lists ideas for further work.

II. BACKGROUND

This letter deals with memoryless nonlinearities of the form

$$y(t) = F_0 \left( (x(t)) \right).$$

Here, $x(t)$ is an input signal, and $y(t)$ is an output signal; both
are assumed defined for $t \in \mathbb{R}$. $F_0$ is a real-valued mapping,
assumed continuous, but not necessarily differentiable, and in
most cases of interest is nonlinear. Two typical examples of
memoryless nonlinear mappings are the saturator, defined by

$$F_0(x) = \frac{1}{2} (|x+1| - |x-1|),$$

and the soft-clipping nonlinearity, defined by

$$F_0(x) = \tanh(x).$$

Fig. 1 shows the input–output relationships for these functions.

In the discrete setting, consider a real-valued input sequence
$x^n$, for $n \in \mathbb{Z}$. Such a sequence could represent samples of
the continuous function $x(t)$, for $t = nT$, where $T$ is a sample
period (and $f_s = 1/T$ is the sample rate), or could be entirely
synthetic. A direct approach to discrete-time emulation of (1)
is to simply compute an output sequence $y^n$ as

$$y^n = F_0 \left( x^n \right).$$

As is well-known, such a trivial implementation generates
aliased components, of strength depending on the smoothness
of the mapping $F_0$, and the amplitude of the input signal $x^n$
[1, 19], [20]. The usual approach to reducing aliasing is to
operate at an oversampled rate, as discussed in Sec. I.

A. First-Order Antialiasing

In a recent paper, Parker et al. presented a novel algorithm
for the reduction of aliasing in discrete-time memoryless
nonlinearities, and suitable for operation at a non-oversampled
rate [22]. It takes on a particularly simple form:

$$y^n = \frac{F_1^n - F_1^{n-1}}{x^n - x^{n-1}}.$$  \hspace{1cm} (5)

Here, $F_1^n = F_1 \left( x^n \right)$ represents the first antiderivative of $F_0$
evaluated at $x^n$. The approximation (5) is arrived at after
a number of steps. In particular, the input sequence $x^n$ is
assumed drawn from samples of a piecewise linear underlying
function $x(t)$, which is then convolved with a box function of
one sample duration, and then resampled. The convolution
operation mentioned above requires the evaluation of the
antiderivative of $F_0$, leading directly to the form in (5).

In the case of the saturator (2) and the soft-clip nonlinearity
(3) the antiderivatives can be given, respectively, as

$$F_1(x) = \frac{1}{4} \left[ (x+1)^2 \text{sgn}(x+1) - (x-1)^2 \text{sgn}(x-1) - 2 \right],$$

where $\text{sgn}(-)$ is the sign function, and

$$F_1(x) = \ln \left( \cosh(x) \right).$$

III. HIGHER-ORDER ANTIDERIVATIVE ANTI_ALIASING

Though it is derived using signal processing considerations
in [22], one observation that can be made about the antialiasing
method (5) is that it represents an approximation to

$$y(x) = \frac{dF_1}{dx} \text{ or } y(t) = \frac{dF_1/dt}{dx/dt}.$$  \hspace{1cm} (8)

It is natural to examine extensions of the method (5), based on repeated antidifferentiation. In particular, consider the extension to $p$th order of (5):

$$y(x) = \frac{d^p F_p}{dx^p} = D^p F_p \quad \text{where} \quad D = \frac{1}{dx/dt} \frac{d}{dt},$$

where $F_p$ is the $p$th antiderivative of $F_0$ to within a polynomial
of degree $p-1$. It is important to point out that except
in special cases, such as the saturator in (2), these are not
available in closed form.

A. A Numerical Method

The key operation in (9) is repeated composition with the
differential operator $D$. In order to construct a discrete-time
approximation to the higher order forms in (9), it is useful
to take an approach based on operator composition through
discrete time approximations to $D$. For an arbitrary sequence
$g^n, n \in \mathbb{Z}$, define unit forward and backward shifts $e_+$ and
e_-, and associated first difference operations $\delta_+$ and $\delta_-$ as

$$e_\pm g^n = g^{n \pm 1} \quad \text{and} \quad \delta_\pm = \pm (e_\pm - 1).$$

Given an input sequence $x^n$ and a sequence $q^n = q(x^n)$, for
some nonlinear mapping $q$, define the operators $D_-$ and $D_2$,
through their action on the time series $q^n$ as

$$D_- q^n = \frac{\delta_- q^n}{\delta_- x^n} \approx D q(x^n)$$

and

$$D_2 q^n = \frac{2}{(e_+ - e_-)} x^n \delta_+ D_- q^n \approx D^2 q(x^n).$$

$D_-$ is a one-sided approximation to $D$, and $D_2$ is a centered
approximation to $D^2$.

Supposing that $p = 2m + r$, where $m = \lfloor p/2 \rfloor$ and $r = \text{mod} \ (p, 2)$, then a discrete-time approximation to (9) may be
written, in operator form, as

$$y^n = e^n D_- D_2^n F_p,$$
where \( F^n_m = F^m_n(x^n) \) and where the \( m \) sample unit delay \( e^m_n \) is used to render the approximation causal. When written explicitly to orders \( p = 1, 2, \) and \( 3, \) the approximation yields

\[
\begin{align*}
y^n &= \frac{F^n_1 - F^{n-1}_1}{x^n - x^{n-1}}, \\
y'^n &= \frac{2}{x^n - x^{n-2}} \left( \frac{F^n_2 - F^{n-1}_2}{x^n - x^{n-1}} - \frac{F^{n-1}_2 - F^{n-2}_2}{x^{n-1} - x^{n-2}} \right),
\end{align*}
\]

and

\[
\begin{align*}
y^n &= \frac{1}{x^n - x^{n-2}} \times \\
&\left[ \frac{2}{x^n - x^{n-2}} \left( \frac{F^n_3 - F^{n-1}_3}{x^n - x^{n-1}} - \frac{F^{n-1}_3 - F^{n-2}_3}{x^{n-1} - x^{n-2}} \right) \right. \\
&\left. - \frac{2}{x^{n-1} - x^{n-3}} \left( \frac{F^{n-1}_3 - F^{n-2}_3}{x^{n-1} - x^{n-2}} - \frac{F^{n-2}_3 - F^{n-3}_3}{x^{n-2} - x^{n-3}} \right) \right]
\end{align*}
\]

(16)

It can be seen that (14) is equivalent to (5), the method presented in [22]. However, (15) and (16) are its novel extensions.

B. Precision and Ill-Conditioning

The operations \( D_n \) and \( D_2 \) both include divisions by differences of signal values; as such, there is the risk of loss of precision or division by zero when the denominator is small. Special approximations are thus necessary under such conditions.

The discrete-time approximation (13) consists of a sequence of operations of the form \( D_n \) or \( D_2. \) Consider a given function \( G_0(x) \) and its first and second antiderivatives \( G_1(x) \) and \( G_2(x) \), as well as the sequences \( G^0_n = G_0(x^n), G^1_n = G_1(x^n), \) and \( G^2_n = G_2(x^n). \)

At time steps \( n \) for which \(|x^n - x^{n-1}| \leq \epsilon\), for some threshold value \( \epsilon\), the following approximation to \( D_n\), obtained through Taylor expansion, may be used:

\[
D_n G^n = G_0(x^n - 1), \quad x^n - 1 = \frac{1}{2} (x^n + x^{n-1}).
\]

(17)

This corresponds to the rule used in [22].

At time steps \( n \) for which \(|x^{n+1} - x^n| \leq \epsilon\), for some threshold value \( \epsilon\), the following approximation to \( D_2 \) may be used:

\[
D_2 G^n = \frac{2}{\Delta} \left( G_1(\bar{x}^n) + \frac{G^n_2 - G_2(\bar{x}^n)}{\Delta} \right),
\]

(18)

where

\[
\bar{x}^n = \frac{1}{2} (x^{n+1} + x^n), \quad \Delta^n = \bar{x}^n - x^n.
\]

(19)

When \(|\Delta^n| \leq \epsilon\), the further approximation

\[
D_2 G^n = G_0 \left( \frac{1}{2} (\bar{x}^n + x^n) \right)
\]

(20)

may be employed.

IV. Evaluation

The antialiasing characteristics of the proposed methods are easily observed in the case of a sinusoidal input. Fig. 2(a) shows the magnitude spectrum of a 1661-Hz sinewave (note G#6) with peak amplitude 10 under trivial hard clipping. A sample rate \( f_s = 44.1 \) kHz was used for this example. As a reference, Fig. 2(b) shows the spectrum of this signal upsampled by a factor of 6 prior to clipping. To bypass the effects of interpolation/decimation filters, the input signal was synthesized on the target rate \( f_s = 264.6 \) kHz. The nonlinearity clearly introduces high levels of aliasing distortion throughout the spectrum which can be reduced using oversampling.

Figs. 2(c)–(e) show the magnitude spectrum of the 1661-Hz sinewave processed using the first-order antiderivative method.
presented in [22] and the proposed extension (13) to second and third order. The methods were implemented using two-
times oversampling, i.e. $f_s = 88.2$ kHz. In particular, aliasing is suppressed more at low frequencies, which is advantageous in audio, because at frequencies below the first harmonic, the audibility of disturbances is limited only by the hearing threshold whereas auditory masking makes high-frequency disturbances between harmonics inaudible [31].

Furthermore, the performance of the proposed methods was evaluated by measuring the caused increase in signal-to-noise ratio (SNR) for a set of sinusoidal inputs. The SNR was defined as the power ratio between the desired part of the signal, and the components generated by aliasing. Following the approach described in [23], an ideal alias-free version of each test signal was synthesized using Fourier analysis and additive synthesis. This bandlimited signal was subtracted from the aliased signal to separate the harmonics from the aliased components. The squared sum of the signal containing only the aliases was used as the noise part in the SNR calculation. Since in audio applications the SNR is only meaningful at audible frequencies, all signals were lowpass filtered up to 16 kHz prior to the SNR evaluations. As before, algorithms (14)–(16) were implemented using an oversampling factor of 2 (i.e. $f_s = 88.2$ kHz). Given the high computational costs of evaluating the hyperbolic tangent function and its antiderivatives at every time step, these were stored in a lookup table (LUT) and read using cubic Lagrange interpolation. Table size was optimized by exploiting the symmetry of these functions. Each LUT was generated by evaluating each nonlinear function at a thousand points between 0 and 10. With this implementation, the error relative to analytical values was found to be in the range of $10^{-15}$.

Figs. 3(a) and (b) show the SNRs for sinusoidal signals with fundamental frequencies between 1 and 10 kHz under hard clipping (2) and soft saturation (3), respectively. As a reference, trivial audio rate processing and oversampling by factor 6 are also shown. For hard clipping, the second- and third-order forms increase SNR by approx. 15 and 30 dB w.r.t. oversampling by factor 6 [cf. Fig. 3(a)]. In the case of the soft-clipper [cf. Fig. 3(b)], the third-order method outperforms oversampling by 6 for high fundamental frequencies. At lower fundamentals the SNRs lie above 96 dB, which can already be considered sufficient for 16-bit audio (CD quality) [32].

The third-order form requires one LUT read and two divisions per output sample. The remaining calculations correspond to previous evaluations and can be stored in memory. If the cost of performing an LUT read as comparable to that of a division, and operating at twice the audio rate, the method requires 6 LUT reads per output sample. This is equivalent to the cost of 6-times oversampling. In practice the additional cost of interpolation and decimation filters required for oversampling must also be taken into account. As the costs of these filters are proportional to oversampling factor, the methods here are more efficient than pure oversampling.

V. Conclusion and Further Work

A new approach to antialiasing for memoryless discrete-
time nonlinearities has been demonstrated in this letter. As has been shown, it is of a general character, more efficient than pure oversampling, and does not depend on the particular type of nonlinearity, or on a simplification of its functional form. It is presented here as a family of methods of increasing order $p$ of antiderivative in the nonlinearity, leading, ultimately, to an increasing degree of aliasing suppression. As is natural, computational cost also scales with the order $p$. There remain many open questions and avenues of future research.

One particular series of discrete approximations to (9) is given in (13): it has the features that a) it maintains the nested structure of the underlying equation, and b) for a given order $p$, is minimal in terms of the number of signal values used to compute an approximation at a given time step, which is $p + 1$. An inherent characteristic of this family of methods is that of spectral shaping of the output; though aliasing is suppressed, there can be some attenuation of the signal in the high frequency range. Additional linear filtering is one option in this case, as suggested in [33], [34] for antialiased oscillators. Given that neither property a) nor b) is necessary in the approximation of (9), the expansion of the scope of the nested structures presented here could prove useful in finding antialiasing methods for which such attenuation is reduced.

Only the single memoryless nonlinearity has been discussed here. A logical extension will be to multiple such nonlinearities in a feedback setting, as it is currently one of the main applications of virtual analog modeling in audio [35], [36], [37], [38], [12]. A major new consideration will be the determination of numerical stability conditions for such antialiasing methods, and will form the basis for future investigations.

VI. Acknowledgment

The main part of this work was conducted in June 2016, during a visit by F. Esqueda, J. D. Parker, and V. Valimäki to the University of Edinburgh.