Audible Aliasing Distortion in Digital Audio Synthesis

Jiri SCHIMMEL

Dept. of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Purkynova 118, 612 00 Brno, Czech Republic

schimmel@feec.vutbr.cz

Abstract. *This paper deals with aliasing distortion in digital audio signal synthesis of classic periodic waveforms with infinite Fourier series, for electronic musical instruments. When these waveforms are generated in the digital domain then the aliasing appears due to its unlimited bandwidth. There are several techniques for the synthesis of these signals that have been designed to avoid or reduce the aliasing distortion. However, these techniques have high computing demands. One can say that today's computers have enough computing power to use these methods. However, we have to realize that today's computer-aided music production requires tens of multi-timbre voices generated simultaneously by software synthesizers and the most of the computing power must be reserved for hard-disc recording subsystem and real-time audio processing of many audio channels with a lot of audio effects. Trivially generated classic analog synthesizer waveforms are therefore still effective for sound synthesis. We cannot avoid the aliasing distortion but spectral components produced by the aliasing can be masked with harmonic components and thus made inaudible if sufficient oversampling ratio is used. This paper deals with the assessment of audible aliasing distortion with the help of a psychoacoustic model of simultaneous masking and compares the computing demands of trivial generation using oversampling with those of other methods.*

Keywords

Music, signal processing algorithms, signal synthesis.

1. Introduction

An audio signal synthesis for digital simulation of electronic musical instruments uses quite different methods than synthesizers designed for measurement purposes or than high frequency range generators like direct digital frequency synthesis [1], etc. A harmonic synthesis [2], wavetable synthesis [3], and SN (sinusoid plus noise) models [4] are used, for instance.

Oscillators generating periodic signals are the basic sources of sound in an audio synthesizer. They are used in both linear and non-linear methods of virtual analog synthesis, i.e. in additive, subtractive, modulation, and wave-shaping syntheses [2]. The type of waveform generated by the oscillator determines the harmonic content of the sound, which affects the resulting timbre.

The basic waveforms are usually triangle, sawtooth, and square (rectangular with 50% duty cycle) waveforms since they are rich in harmonic components [2]. However, the problem is that they have an infinite Fourier series and, therefore, there will be aliasing when they are trivially generated using methods adopted from analog synthesizers [5], see Fig. 1. Coefficients of the Fourier series of these waveforms decrease with their order (square and sawtooth waveforms) or with the square of their order (triangle waveform); even coefficients of square and sawtooth waveform are zero [3], [6]. We assume that direct component of all waveforms is zero and magnitude is not normalized.
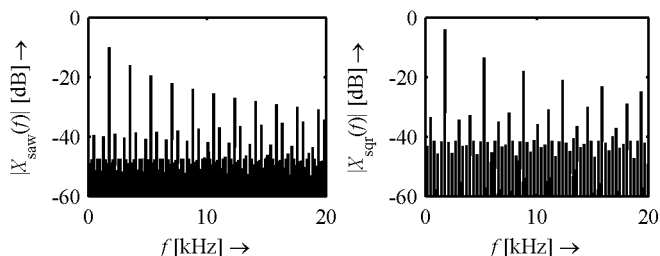

Fig. 1. Sawtooth and square waveforms in frequency domain with obvious aliasing components.

In this paper, the term "baseband" is used for the frequency band from zero to the half of the sampling frequency used in the digital audio processing system in which the synthesis is implemented. The harmonics of the examined waveform that belong to the first period of the spectrum in the baseband will be referred to as "harmonic components". Higher harmonics that belong to replicas of the spectrum in the baseband (i.e. those causing the aliasing distortion) will be denoted as "aliasing components".

Audible aliasing distortion logically reduces the quality of the audio signal. When we deal with an audio quality assessment, subjective methods should be used preferably. However, these methods need a lot of time and many listeners. Also financial costs are not negligible. That is why the objective methods are used almost exclusively for the quality assessments in the telecommunication applications, for example the latest ITU–T P.863 P.OLQA method [7]. For the same reasons, an objective method PEMO-Q that

uses a model of auditory perception was introduced to the audio quality assessment of music as well [8]. However, this method focuses mainly on the assessment of audio codecs. The aim of this research is to find the objective method that can be used for rapid assessment of the audio quality reduction caused by aliasing distortion during development of algorithms of audio signal synthesis and non-linear processing (e.g. tube simulation effects, distortion effects, etc.).

The paper is organized as follows: the following section summarizes the state of the art in the area and section 3 deals with the synthesis of trivially generated classic waveforms and with the reduction of aliasing distortion using oversampling. Section 4 describes psychoacoustic models used for analysis, and section 5 presents the simulation results. The last section discusses the results and outlines the future work.

2. State of the Art

There are several techniques that can be used to avoid or reduce the aliasing distortion: oversampling, additive synthesis, and wavetable methods [2], [3], Discrete Summation Formulas (DSF) [9], Differentiated Parabolic Waves (DPW) [10], and successive integration of Band-limited Impulse Train (BLIT) [11]. These techniques were compared in [12] from the point of view of aliasing distortion and computational complexity in real-time audio signal synthesis. The results show that the BLIT method is the best from the aliasing distortion point of view but it has much higher computing demands in comparison with the other tested methods. However, interesting results were found for trivially generated waveforms when sufficient oversampling was used. The aliasing distortion was not high and the computational complexity was comparable to the modified BLIT method that uses the sum of windowed sinc functions (BLIT-SWS) [11].

Spectral components produced by aliasing can be masked with harmonic components and thus made inaudible. To the author's knowledge, there is no paper or study dealing with the application of psychoacoustic models of simultaneous masking to the spectrum of the waveform to determine which aliasing components are masked by the harmonic components. There is a study of the aliasing effect in [13] that uses auditory spectrum estimation for baseband components to determine the audibility of aliasing components. The paper primarily deals with real-time tube amplifier emulation using the digital triode model. The auditory spectrum estimation is presented only for the amplifier output signal when a sinusoid input signal with a frequency of 1 kHz is used.

3. Synthesis Using Oversampling

One method of synthesis of classic periodic waveforms with reduced aliasing distortion is trivial generation

using a sampling frequency N times higher than the sampling frequency of the output signal, followed by N-fold downsampling [3], [11]. In the analog domain, the classic synthesizer waveforms are generated as follows [11]: the sawtooth waveform is typically implemented as an integrator that is reset when a threshold value is reached. The integrator slope controls the frequency. The square waveform is implemented by running a sawtooth waveform through a comparator. The duty cycle of the resulting wave could be controlled via an offset amount. The triangle waveform is implemented by a variety of methods, for example using full-wave rectification of a sawtooth.

An implementation of these methods in the digital domain is outlined in Fig. 2: modulo counter is followed by a comparator and absolute value computation. If we suppose that the conditional-update operation takes two instruction cycles and no hardware support for modulo operation is available, then the synthesis of the sawtooth waveform $x_{saw}(n)$ needs three instructions per sampling frame, the synthesis of square waveform $x_{sqr}(n)$ needs a total of five instructions, and the synthesis of triangle waveform $x_{tri}(n)$ a total of seven instructions.
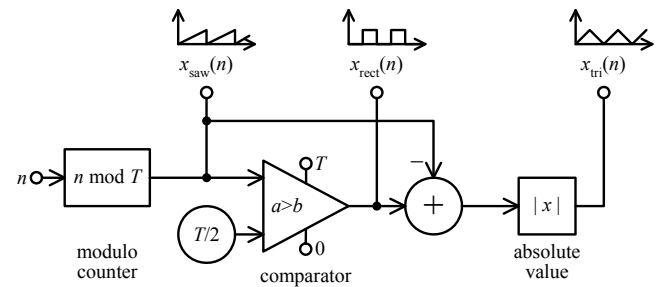

Fig. 2. Trivial generation of classic periodic waveforms in digital domain.

The Fourier series of these waveforms is infinite, so the signal still exhibits aliasing distortion even if oversampling is used. However, the amplitudes of the harmonics of the series are decreasing with the order of the harmonic. We can compute the amplitudes of aliasing components if we know the sampling frequency f_S and oversampling ratio N. The number of the last harmonic component l in the baseband is

$$l = \text{floor}\left(\frac{f_S}{2f}\right). \qquad (1)$$

The number of the first harmonic m, which produces aliasing distortion (aliasing component with the highest frequency in the baseband) is

$$m = \text{floor}\left(\frac{(2N-1)f_S}{2f}\right) + 1 \qquad (2)$$

where f is the fundamental frequency of the waveform. The difference Δf between the last harmonic component in the baseband and the aliasing component with the highest frequency in baseband is

$$\Delta f = Nf_S - (l+m)f. \qquad (3)$$

Due to the decreasing amplitude of the higher harmonics of the examined waveforms, the aliasing components might be masked with harmonic components because aliasing components have lower amplitudes. The expectation that the masking effect will manifest itself gets higher when the amplitude difference between nearby spectral components increases and the frequency distance decreases. This statement follows from the knowledge of the shape of masking patterns [14]. The smallest amplitude difference ΔA is between the lth and the mth harmonics and it can be express as a ratio of Fourier series coefficients b_l and b_m in dB

$$\Delta A = 20 \log_{10} \frac{b_l}{b_m} \qquad (4)$$

which is −6 dB/oct for the square and sawtooth waveforms, and −12 dB/oct for the triangle wave. The worst-case situation occurs when only one harmonic component is present in the baseband; it means when the fundamental frequency of an oscillator is higher than $f_S/4$. This happens for the f^6 tone when the sampling rate is 44.1 kHz, and for the g^6 tone when the sampling rate is 48 kHz. The highest tone that can be played in MIDI-controlled equal-temperament sound synthesizers is coincidentally g^6 [15].

4. Psychoacoustic Model

One can utilize the psychoacoustic model of simultaneous masking to distinguish which aliasing components are masked with harmonic ones and which are audible. Unfortunately, psychoacoustic models describing the human auditory system are still under research; there are many recommendations for various applications.

4.1 Spreading Function

Several spreading functions have been suggested as an approximation of excitation patterns in the bark scale [14]. Fig. 3 shows the most frequently used spreading functions: 2-slope triangle spreading function derived from narrow-band noise masking data [16], modified Schroeder spreading function [17] used for perceptual coding applied to speech signals, and MPEG Psychoacoustic Model 1 and 2 [18]. The last one does not take the level dependence of excitation patterns into consideration. These spreading functions are used in real-time applications instead of the former auditory filterbanks suggested e.g. in [19].
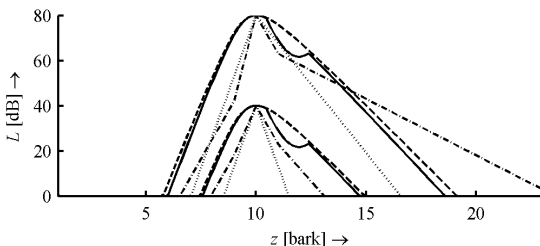

Fig. 3. Spreading functions for 40 dB and 80 dB maskers: 2-slope triangle (dotted), MPEG Model 1 (dash-and-dot), Schroeder (dashed), and MPEG Model 2 (solid).

4.2 Masking Curve Offset

The peak of the masking curve is shifted down from the masker level by an amount ΔL_m that depends on the minimum changes in the excitation pattern that we can detect. However, there is no general agreement about this amount [16]. The peak of the spreading function should be about 6 dB below the level of the masker according to [14] and 16 dB according to [20]. An offset of 10 dB is used in the auditory spectrum estimation in [13]. Later works use different amount ΔL_{mT} for the tone-like signal masking a noise-like signal, and ΔL_{mN} for the noise-like signal masking a tone-like signal, for example according to [21]

$$\Delta L_{mT} = 14.5 + z(f), \ \Delta L_{mN} = C \qquad (5)$$

where $z(f)$ is the critical band rate, and C varies between 3 dB and 6 dB depending upon the experimental data, or according to [18]

$$\Delta L_{mT} = 6.025 + 0.275 z(f), \Delta L_{mN} = 2.025 + 0.175 z(f). \ (6)$$

Harmonic components are tone signals and we assume that aliasing components in the baseband are noise-like signals. Fig. 4 compares the amount of spreading function downshift for the tone-like signal masking a noise-like signal as a function of critical band rate.
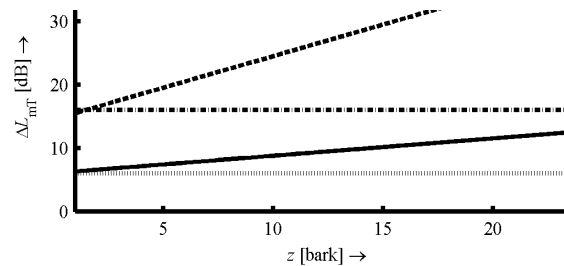

Fig. 4. Comparison of offset between peak of the spreading function and the masker level: Zwicker (dotted), Moore (dash-and-dot), Jayant (dashed), and MPEG Model (solid).

4.3 Masked Threshold

All spectrum components will take part in the masking effect, not only the nearest one. Combining the individual masking curves into a global one is also still under research; various psychoacoustic models use different methods. MPEG psychoacoustic models use the summation of intensities of individual masking curves [18]. Some studies in the literature use a non-linear summation model in which the resulting masked threshold of two masking curves is higher than the sum of individual thresholds [16]. The maximum value of overlapping curves is used in [13] as well as in Dolby AC-2 and AC-3.

4.4 Model Application

At first, the amplitudes and frequencies of harmonic and aliasing components are computed from a given sampling frequency f_S, oversampling ratio N, and fundamental

frequency of the signal f. Aliasing components are computed until their amplitude is less than –80 dB, which is for $k > 10^4$ for the sawtooth and square waveforms, and for $k > 100$ for the triangle waveform. When the amplitudes of spectral components are known, the sound pressure level (SPL) is computed for each spectral component, using a modified computation of MPEG Model 1 SPL [18]

$$L_k = 96 + 10 \log_{10}\left(\left|\frac{b_k}{2}\right|^2\right) \qquad (7)$$

where b_k are coefficients of Fourier series of the waveform. As the next step, the threshold in quiet is computed [22]

$$T(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-15}f^4 \qquad (8)$$

and applied to the signal spectrum. Frequency mapping to critical band rate $z(f)$ is then performed [14]

$$z(f) = 13\arctan(0.76f/1000) + 3.5\arctan\left((f/7500)^2\right). \qquad (9)$$

After that, masking curves are computed for all harmonic components using the spreading function $L(L_K, z)$, downshifted from the masker level, and combined into a masked threshold

$$MT_\mathrm{H}(z) = \sum_{k=1}^{l} L(L_k, z) - \Delta L_\mathrm{mT} \qquad (10)$$

$$MT_\mathrm{H}(z) = \max\{L(L_k, z) - \Delta L_\mathrm{mT}\} \text{ for } k \in \langle 1, l\rangle \qquad (11)$$

where ΔL_mT is the offset for the tone-like signal masking a noise-like signal, and l is given by (1). This masked threshold is stored and computation continues with the aliasing components until all remaining spectral components are lower than the masked threshold

$$MT_\mathrm{A}(z) = MT_\mathrm{H}(z) + \sum_{r=m}^{n} L(L_r, z) - \Delta L_\mathrm{mT} \qquad (12)$$

$$MT_\mathrm{A}(z) = \max\{MT_\mathrm{H}(z), L(L_r, z) - \Delta L_\mathrm{mT}\} \text{ for } r \in \langle m, n\rangle \qquad (13)$$

where m is given by (2), and n is determined iteratively.

5. Audible Aliasing Distortion

In a preliminary work published in [23], three criteria were used to evaluate audible aliasing distortion. The ratio of the energy of the harmonic components and the energy of the aliasing maskers has proved to be the most effective criterion for an objective assessment of audible aliasing distortion. This ratio is defined as

$$SAMR = 10\log\left(\frac{\sqrt{\sum_{k=1}^{l} b_{\mathrm{HC}k}^2}}{\sqrt{\sum_{r=m}^{n} b_{\mathrm{AM}r}^2 - MT_\mathrm{H}(z_r)}}\right) \qquad (14)$$

where $b_{\mathrm{HC}k}$ is the spectral coefficient of the kth harmonic component, and $b_{\mathrm{AM}r}$ is the spectral coefficient of the rth

aliasing masker in the audible frequency range. $MT_\mathrm{H}(z_r)$ is the value of the masked threshold of harmonic components for $z_r = z(f_r)$, where f_r is the frequency of the rth aliasing masker. Aliasing components that are below the masked threshold are not included in the computation of *SAMR*.

Fig. 5 illustrates the computation of the masked threshold and the *SAMR* value. It shows the spectrum of a sawtooth waveform with 8-fold oversampling and its masked threshold computed using MPEG model 1. The masked threshold of harmonic components and the non-masked aliasing components are shown as well.
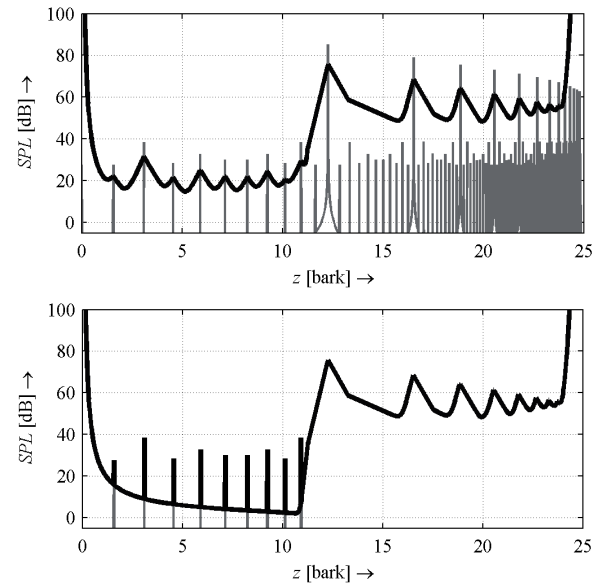

Fig. 5. Spectrum of sawtooth waveform with 8-fold oversampling and its masked threshold $MT_\mathrm{A}(z)$ computed using MPEG model 1 (top); masked threshold of harmonic components $MT_\mathrm{H}(z)$ computed using MPEG model 1 and non-masked aliasing components AMr (bottom).

5.1 Sequence of Testing Tones

The spreading function descends slowly from the masker frequency towards the higher frequencies (see Fig. 3) so that the lower harmonic components support the masking of all higher aliasing components. Less masking generally occurs at the higher fundamental frequencies of the waveform but it also depends on the ratio of the tone pitch and the sampling frequency, see (3). A sawtooth waveform was used as the testing signal because the amplitudes of its higher harmonics decrease more slowly than the amplitudes of the higher harmonics of the triangle waveform and the signal contains all the harmonics, not only the odd ones. Fig. 6 shows the number of aliasing maskers (audible aliasing components) as a function of the tone pitch of trivially generated sawtooth waveform using 8-fold oversampling. The MPEG psychoacoustic model 1 was used to compute the masked threshold. Local minima in Fig. 6 correspond to the fundamental frequencies, which are close to the integer proportion of the sampling frequency, i.e. close to the harmonic components and thus below their masking curve.
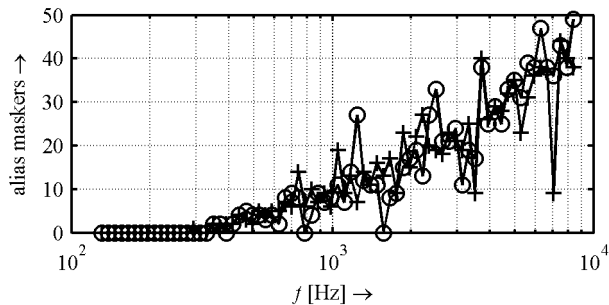
Fig. 6. Number of aliasing maskers as a function of tone pitch of trivially generated sawtooth waveform using 8-fold oversampling for a sampling frequency of 44.1 kHz (circle symbols) and 48 kHz (cross symbols).

Once we start dealing with audio signal synthesis, we should consider maximum reasonable fundamental frequency. The range of MIDI notes is from C_2 to g^6 [15], which means 1223 harmonics in the audible range for the C_2 tone (16.35 Hz) and only a single harmonic for e^6 (10.548 kHz) and higher notes. The timbre of the sawtooth waveform is similar to the timbre of stringed instruments and the timbre of the square signal is similar to that of reed instruments. The e^6 tone (the lowest tone with only a single harmonic in the audible range) is outside the range of these instruments.

We should also consider the efficiency of the additive synthesis when only several harmonics are in the audible range. Fig. 7 shows the number of operations n_{ops} for generating one sample frame of sawtooth waveform using the additive synthesis as a function of the fundamental frequency of the tone f. A look-up table method with a simple linear interpolation of samples with five operations per sample frame is suggested to generate the sine waveform for additive synthesis [3]. Horizontal lines in Fig. 7 denote the number of operations per sample frame of trivially generated sawtooth waveform, using oversampling ratios from 4 to 64, in steps of 4. The use of a second-order IIR filter in the direct form II is assumed in downsampling (9 operations per sample frame).
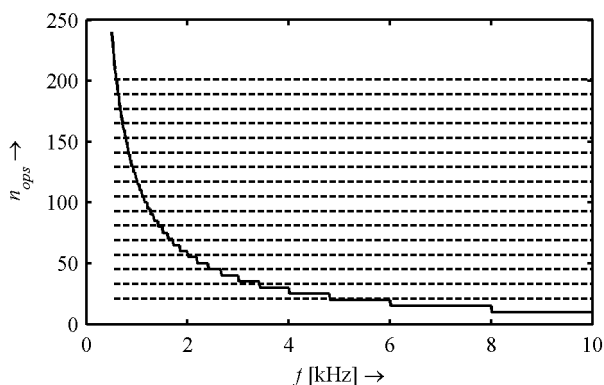

Fig. 7. Number of operations for generating one sample frame of sawtooth waveform using additive synthesis and trivial generation with different oversampling ratio as a function of fundamental frequency of the tone.

Fig. 7 shows the following: in comparison with additive synthesis, trivial generation of the waveform is of lower computational complexity for fundamental frequencies lower than 4800 Hz when 4-fold oversampling is used, and of lower computational complexity for frequencies lower than 3424 Hz when 8-fold oversampling is used, etc. Accordingly, Tab. 1 shows the maximum pitch of the tone suitable for trivial generation with a given oversampling ratio.

oversampling	4x	8x	16x	32x	64x
f_S = 44.1 kHz	c#^{5}	g^4	a#^{3}	c^3	c^2
f_S = 48 kHz	d^5	g#^{4}	h^3	c#^{3}	d^2

Tab. 1. Maximum pitch of the tone suitable for trivial generation.

It can be also seen from Fig. 6 that all aliasing components are masked below a certain pitch of the tone. Tab. 2 shows the highest tones for which all aliasing components are masked when MPEG psychoacoustic model 1 is used. However, the results slightly vary for different models and so the pitch of tone c^1 (261.62 Hz) was chosen as the minimum pitch of the testing tone for all sampling frequencies and oversampling ratios.

oversampling	4x	8x	16x	32x	64x
f_S = 44.1 kHz	c#^{1}	f^1	f^1	a#^{1}	a^2
f_S = 48 kHz	d^1	d^1	g#^{1}	h^1	c#^{2}

Tab. 2. Highest tones for which all aliasing components are masked when MPEG psychoacoustic model 1 is used.

5.2 Simulation Results

As presented in [23], *SAMR* depends on the psychoacoustic model used, mainly on the masking curve offset. Therefore, two models of simultaneous masking were used for the audible aliasing distortion assessment: MPEG psychoacoustic model 1 (model A) and a model that uses the 2-slope triangle spreading function with an offset of 16 dB according to [20], and the maximum value of overlapping curves according to (11) and (13) for computing the masked threshold (model B). For all tones in the testing sequence, both models have no aliasing maskers above the fundamental frequency. There is virtually a constant number of aliasing components below the fundamental frequency of each testing tone but their level gets lower when the oversampling ratio increases.

Fig. 8 shows the *SAMR* of trivially generated sawtooth waveform as a function of the fundamental frequency of the testing tone for 16-fold oversampling. Similar results are obtained when model B is used. The lines in Fig. 8 are regression lines which fit the data in the least squares sense (solid line for f_S = 44.1 kHz, dashed line for f_S = 48 kHz). Regression lines for different oversampling ratios and the sampling frequency 48 kHz are compared in Fig. 9 together with regression lines obtained when model B is used.

There are no results for 64-fold oversampling in Fig. 9 since all aliasing components are masked for both models as well as for both sampling frequencies.
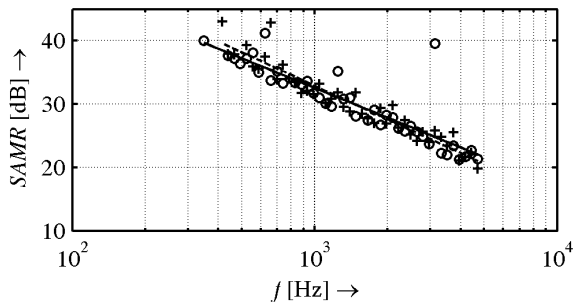
Fig. 8. *SAMR* of sawtooth waveform: 16-fold oversampling, sampling frequency 44.1 kHz (circle symbols) and 48 kHz (cross symbols), model A.
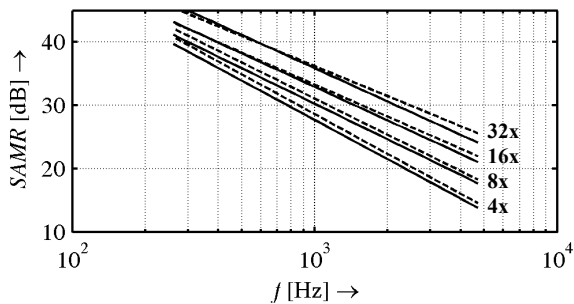

Fig. 9. *SAMR* of sawtooth waveform as a function of the fundamental frequency of the tone for the sampling frequency 48 kHz and different oversampling ratios when model A (solid) and model B (dashed) are used.
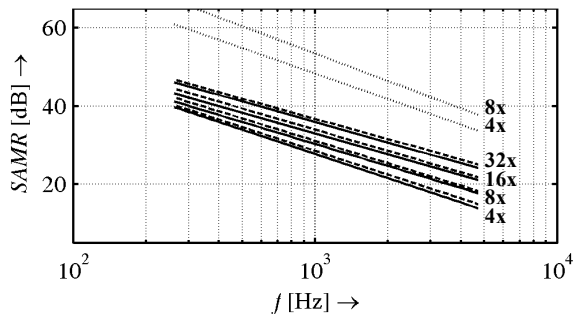

Fig. 10. *SAMR* of sawtooth (solid), square (dashed), and triangle (dotted) waveform as a function of fundamental frequency of the tone for a sampling frequency of 48 kHz and different oversampling ratios when model A is used.

The simulations were performed for both square and triangle signals. Fig. 10 shows the resulting regression lines. One can see that the *SAMR* values for the square waveform are only about 1 dB higher and those for the triangle waveform are more than 20 dB higher than the values for the sawtooth waveform. All aliasing components are masked for oversampling ratios higher than 8.

6. Conclusion

According to the simulation results, there should be no audible aliasing distortion for tones lower than or equal to those listed in Tab. 2. For tones higher than or equal to those listed in Tab. 1 the additive synthesis using a look-up table should be used instead of trivial generation of wave-

forms because the additive synthesis has a lower distortion and lower computational complexity for these tones. The assessment of the audible aliasing distortion using *SAMR* in the tone range bounded by these limits is shown in section 5.2. Tab. 3 shows the audible aliasing distortion and the number of operations n_{ops} needed to generate one sample frame of sawtooth waveform using trivial generation with oversampling, and compares them with the number of operations of the BLIT method. The values for the BLIT method are taken from [12]. The *SAMR* values in Tab. 3 are based on model B. The upper tone is determined according to Tab. 1. The *SAMR* values on the right are computed for this tone.

Method	n_{ops}	$SAMR(a^1)$ [dB]	upper tone	$SAMR$ [dB]
8-fold oversampling	33	37	g#^{4}	21
16-fold oversampling	57	38	h^3	27
32-fold oversampling	105	∞	c#^{3}	33
64-fold oversampling	201	∞	d^2	∞
BLIT	235	∞		∞

Tab. 3. Comparison of methods of generating sawtooth waveform.

A hybrid synthesis that uses the trivially generated waveforms with 64-fold oversampling for low frequencies and the additive synthesis for high frequencies is comparable with the BLIT method from the point of view of computational complexity and audible aliasing distortion. When a lower computational complexity is required, a lower oversampling ratio can be used in view of the maximum audible aliasing distortion.

The analysis presented in this paper can be also used in the design of non-linear digital audio effects. However, in comparison with most of the non-linear audio effects, the trivial generation of waveforms has the advantage of the amplitudes of the output signal spectral components being expressed analytically. In order to validate the simulation results and improve the parameters of the models, a subjective evaluation of audible aliasing distortion has to be performed. Finding a correlation between the objective evaluation using *SAMR* and the subjective evaluation will be the subject of future work.

Acknowledgements

References

[1] STOFANIK, V., BALAZ, I. Frequency stability improvement in direct digital frequency synthesis. *Radioengineering*, 2004, vol. 13, no. 2, p. 17 – 19.

[2] RUSS, M. *Sound Synthesis and Sampling*. Focal Press, 1996.

[3] PUCKETTE, M. *The Theory and Technique of Electronic Music*. World Scientific Publishing Co. Pte. Ltd., 2007.

[4] TURI NAGY, M., ROZINAJ, G. An analysis/synthesis system of audio signal with utilization of an SN model. *Radioengineering*, 2004, vol. 13, no. 4, p. 51 – 57.

[5] OPPENHEIM, A. V., SCHAFER, R. W., BUCK J. R. *Discrete-Time Signal Processing*. 2nd edition. Prentice-Hall, Inc., 1998.

[6] OPPENHEIM, A. V., WILLSKY, A. S., NAWAB, S. H. *Signals & Systems*. 2nd edition. Prentice-Hall, Inc., 1997.

[7] *Recommendation ITU–T, P.863 (ex P.OLQA): Perceptual Objective Listening Quality Assessment*, 2010.

[8] HUBER, R., KOLLMEIER, B. PEMO-Q, A new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech, and Language Processing*, Nov. 2006, vol. 14, no. 6, p. 1902-1911.

[9] MOORER, J. A. The synthesis of complex audio spectra by means of discrete summation formulae. *Journal of the Audio Engineering Society*, 1975, vol. 24, p. 717 – 727.

[10] VALIMAKI, V. Discrete-time synthesis of the sawtooth waveform with reduced aliasing. *IEEE Signal Processing Letters*, 2005, vol. 12, no. 3, p. 214 – 217.

[11] STILSON, T.S. Efficiently-variable non-oversampled algorithms in virtual-analog music synthesis. *Ph.D. dissertation*, Dept. Electrical Engineering, Stanford University, 2006.

[12] SCHIMMEL, J., SERRA, J. P. Comparison of real-time syntheses of band-limited periodic digital audio signals. In *Proceedings of 6th International Conference on Teleinformatics ICT 2011*. Brno, BUT, 2011. p. 197 – 202.

[13] PAKARINEN, J., KARJALAINEN, M. Enhanced wave digital triode model for real-time tube amplifier emulation. *IEEE Transaction on Audio, Speech, and Language Processing*, 2010, vol. 18, no. 4, p. 738 – 746.

[14] ZWICKER, E., FASTL, H. *Psychoacoustic – Facts and Models*. 2nd updated edition. Springer-Verlag, 1999.

[15] MIDI Manufacturers Association, Japan MIDI Standard Committee *The Complete MIDI 1.0 Detailed Specification*, doc. version 96.1, 1997.

[16] BOSI, M. GOLDBERG, R. E. *Introduction to Digital Audio Coding and Standards*. Kluwer Academic Publishers, 2003.

[17] SCHROEDER, M. R., ATAL, B. S., HALL, J. L. Optimizing digital speech coders by exploiting masking properties of the human ear. *Journal of Acoustic Society America*, 1979, vol. 66, no. 6, p. 1647 – 1652.

[18] ISO/IEC 11172-3:1993 *Information technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s - Part 3: Audi*o.

[19] PATTERSON, R. D. Auditory filter shapes derived with noise stimuli. *Journal of Acoustic Society America,* 1976, no. 59, p. 640 to 654.

[20] MOORE, B. C. J. Masking in the human auditory system. In *Collected Papers on Digital Audio Bit-Rate Reduction*. GILCHRIST, N., GERWIN C. (ed.) AES, 1996, p. 9 – 19.

[21] JAYANT, N., JOHNSTON, J., SAFRANEK, R. Signal compression based on method of human perception. *Proc. of IEEE*, 1993, vol. 81, no. 10, p. 1385 – 1422.

[22] TERHARDT, E. Calculating virtual pitch. *Hearing Res.*, 1979, vol. 1, p. 155 – 182.

[23] SCHIMMEL, J. Objective evaluation of audible aliasing distortion in digital audio synthesis. In *Proceedings of 34th International Conference on Telecommunications and Signal Processing TSP 2011*. Budapest (Hungary), 2011. p. 343 – 347.

About Author ...

Jiri SCHIMMEL was born in Brno, Czech Republic, in 1976. He received his M.Sc. and Ph.D. degrees in Electronics and Communications in 1999 and in Teleinformatics in 2006. He is currently an assistant professor at the Department of Telecommunications of the Faculty of Electrical Engineering and Communication, Brno University of Technology, Czech Republic. His research is focused on acoustics, multichannel digital audio signal processing, and software and hardware development for real-time audio signal processing systems. He is a member of the AES and IEEE.