Publication VII

J. Parker and V. Välimäki. Linear Dynamic Range Reduction of Musical Audio using an Allpass Filter Chain. *IEEE Signal Processing Letters*, Vol. 20, No. 7, pp. 669–672, July 2013.

Linear Dynamic Range Reduction of Musical Audio using an Allpass Filter Chain

Julian Parker and Vesa Välimäki, *Senior Member, IEEE*

Abstract—The reduction of signal dynamic range through limiting of peak amplitude is an important process in modern audio signal processing, mainly for loudness maximisation. Traditional processes are non-linear, and can produce significant distortion of the processed signal. In this paper we present a new linear technique that reduces the peak amplitude of transient signals using *golden ratio* allpass filters. The system is applied to test signals consisting of both isolated musical sounds and mixed musical audio. The average reduction of the peak amplitude of the musical passages considered is 2.5dB. The system can be applied alongside non-linear methods, to reduce the distortion associated with a particular reduction in peak amplitude.

Index Terms—Audio systems, digital filters, digital signal processing, dispersion, music.

I. INTRODUCTION

LIMITING is an important process in modern musical audio engineering. It refers to a system, usually non-linear, that constrains peaks in an audio signal to a certain maximum value. Limiting is generally used for one of two purposes. The first and original use is as a final element in the audio chain that prevents the output signal from exceeding the optimal range of the system, and hence causing clipping or overdrive. In recent years, limiting has been used as a way of reducing the peak amplitude of a signal relative to its RMS amplitude. By following the limiter with a gain element, RMS amplitude and hence the perceived loudness of a signal can be maximized. Limiting has hence become a popular tool used by music producers and sound engineers to make their recordings and mixes louder than a competitor's [1].

Limiting has traditionally been approached via non-linear techniques. The most common approach is to employ a variant of a standard dynamic-range compressor [2] designed to allow a short attack time and a high compression ratio. Time-independent non-linearities such as the hard-clipping function have also been used as a form of limiting [1]. As these techniques are non-linear, they produce new frequency components, both harmonic and in-harmonic. This can adversely effect the sound quality of a recording [1].

In this work we present a method for reducing the peak to RMS amplitude ratio (also known as the crest factor) of audio signals via the application of allpass filters. The standard allpass filter has a flat frequency response, but a non-flat phase response. Allpass filters have been used within audio signal processing for warped filter design [3] and emulating dispersive wave propagation [4]. By altering the phase of the signal, we can smear the energy of peak or transient in the

J. Parker and V. Välimäki are with the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland e-mail: first-name.lastname@aalto.fi

signal over a larger period of time, without any alteration of the frequency content. If this period of time is kept small enough – below around 4ms in the case of an impulse and longer in the case of more complex signals [5], the change is inaudible. This potentially allows the peak to RMS amplitude ratio and hence the loudness of a signal to be improved without any degradation in sound quality. Previous work has considered the related problem of minimizing the peak factor of a given power-spectrum both in the general case [6], [7] and as applied to wavetable synthesis [8], [9]. A similar approach has been taken to the dynamic range reduction of speech signals [10], utilizing a sinusoidal model. However, these methods are intended for signals which are approximately harmonic and steady-state, and are less appropriate when attempting to reduce the peak amplitude of transient signals.

Section II examines the simplest case – dynamic range reduction of a single impulse, and derives the optimal allpass filter for this purpose. Section III describes a structure for extending this process to more general musical signals, both isolated instrument sounds and mixed music, and presents an evaluation of the method. Section IV concludes.

II. ALLPASS FILTERING AS DYNAMIC RANGE COMPRESSION OF AN IMPULSE

The genesis of the idea presented in this paper was the realization that the allpass filter, as a structure that conserves energy but spreads that energy over time, is by its nature a structure that reduces the dynamic range of transient signals.

Starting with the difference equation of the general first-order allpass filter with coefficient g, we have:

$$y(n) = gx(n) + x(n-1) - gy(n-1). \qquad (1)$$

If we evaluate the response of this difference equation to a unit impulse at $n = 0$, we have:

$$y(0) = g, \quad y(1) = 1 - g^2, \quad y(2) = g^3 - g,$$
$$y(n) = (-1)^n g^{n-2}(1 - g^2). \qquad \forall\, n > 0$$

Given that $|g| < 1$ in order for the filter to be stable, we can already see that the peak amplitude of the filter has been decreased, and therefore dynamic range compression has taken place. Continuing the analysis:

$$|y(n)| = g^{n-2}(1 - g^2), \qquad\qquad \forall\, n > 0$$
$$\implies |y(n)| \propto g^{n-2} \implies |y(n+1)| < |y(n)|. \qquad \forall\, n > 0$$
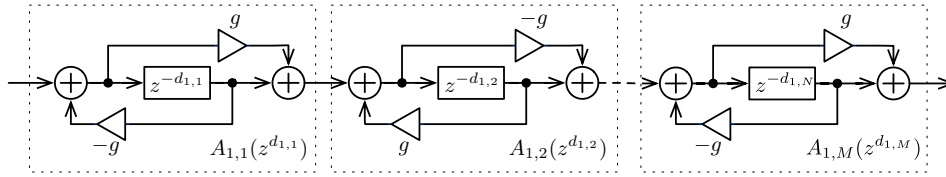
Hence, we can conclude that the maximum peak in the impulse

Fig. 1.   Block diagram showing allpass chain structure.

response occurs at $y(0)$ or $y(1)$. This results in three cases:

$$|y(0)| > |y(1)| \implies g^4 - 3g^2 + 1 < 0,$$
$$|y(0)| < |y(1)| \implies g^4 - 3g^2 + 1 > 0,$$
$$y(0) = y(1) \implies g^4 - 3g^2 + 1 = 0.$$

Therefore the impulse response has three different regions of behavior, depending on the value of the coefficient g. Solving the above quadratic equation in g^2, we find that:

$$g^2 = (3 \pm \sqrt{5})/2 \implies g = (\pm 1 \pm \sqrt{5})/2. \tag{2}$$

These solutions can be recognized as the *golden ratio*, $\phi \approx 1.618$, its inverse, $\Phi \approx 0.618$ and their negative equivalents. Since we have the requirement that $|g| \leq 1$ in order for the filter to be stable, we discard the $\pm\phi$ solutions, and are left with the $\pm\Phi$ solutions. We can then characterize the behavior of the impulse response in the three different regions of g. When $|g| < \Phi$, $y(1)$ is the peak and $|y(1)| > \Phi$. When $1 > |g| > \Phi$, $y(0)$ is the peak and $|y(0)| > \Phi$. Finally, when $|g| = \Phi$, $|y(0)| = |y(1)| = \Phi$. Therefore, whilst peak amplitude reduction occurs for all allowable coefficient values, the greatest reduction is achieved when the allpass coefficient is equal to $\pm\Phi$. This property has been mentioned in previous work [11], without proof.

III. EXTENSION TO GENERAL SIGNALS

A. Isolated Transient Signals

The above analysis only tells us about the impulse response of the filteri.e. its reaction to the ideal transient signal. The response of the filter to extended transients, or to the mixture of transient and steady-state signals found in most audio, is complicated. In some situations, the first-order allpass with Φ coefficient presented above can in fact increase the dynamic range of a signal. The most obvious example is the pathological case where the input to the filter is its time-reversed impulse response. In this case, the filter will concentrate all the energy of the input into a single sample.

The behavior of the filter in response to more complex signals is most easily examined by considering its action as a frequency-dependent delay line. Assuming that total signal energy is the same, a transient whose frequency components all arrive concurrently and in-phase will generally (especially in the case of harmonic signals) have a large peak amplitude, and a transient whose frequency components arrive at slightly different times or out-of-phase will tend to have a lower peak amplitude. It then becomes clear that to reduce the peak amplitude of a general transient, we need a filter that conserves energy but leaves its frequency components arriving at slightly different times. Given that the arrival times of different frequency components may differ between transients, we need to vary the structure somehow in order to reduce

the peak amplitude of a particular transient. Experimentation shows that varying the allpass coefficient, g, is not sufficient — in fact, the best results for many signals are still obtained when $g = \Phi$, although they are not necessarily a reduction in peak amplitude. In order to produce good results for a wide variety of transients, we need more degrees of freedom.

A logical extension is to replace the unit delay in the first-order allpass with a longer delay-line, creating a 'stretched allpass filter' [12]:

$$A(z^d) = \frac{g + z^{-d}}{1 + gz^{-d}}, \tag{3}$$

where d is the delay-line length. This process aliases the phase-delay and group-delay curves of the filter, creating a comb-like shape. In contrast to variation of g, there is no clear region of d values which produce better results than others – the best value is purely dependent on the nature of the input signal. Hence, it is a better choice than g as a free parameter.

Another extension is to cascade M allpass filters in series, each with different internal delay-line lengths. This structure is shown in Figure 1. The sign of the coefficient g is alternated between cascaded sections, in order to avoid the appearance of a large group-delay peak at dc or Nyquist, which can be easily audible. The transfer function is given by:

$$A(z) = \prod_{k=1}^{M} \frac{(-1)^{k+1}g + z^{-d_k}}{1 + (-1)^{k+1}gz^{-d_k}}, \tag{4}$$

where the d_k are the chosen delay-line lengths for each section.

The combination of these allpass sections gives a complicated phase and group-delay response, and can be varied considerably by changing the internal delay-line lengths. The problem then becomes finding a particular combination of delay-line lengths which minimise the peak amplitude of a particular transient, whilst producing an impulse response that is short enough that the spreading of the transient is inaudible. One option would be an optimisation process, where the
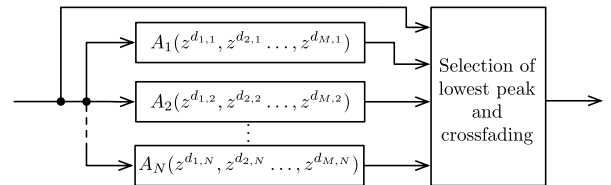

Fig. 2.   Block diagram showing parallel configuration of allpass chains.

structure is repeatedly tested and altered until a local minima of peak amplitude is produced. Instead, we propose a system containing N of the systems described in Figure 1 in parallel. The delay-line lengths in each of the parallel structures are chosen at random from a list of appropriate values. The

input signal is processed through all of these parallel filters simultaneously, and the one that produces the output with the lowest peak amplitude is chosen as the output of the system. This structure is shown in Figure 2. This approach has the significant advantage over an iterative optimisation based system of being highly suited for parallelization on modern CPUs and GPUs [13]. A large value of N is therefore possible whilst maintaining a low computational load. Alongside the parallel allpass filters, the unfiltered signal is considered. This guarantees that the overall system will never increase peak-amplitude, even in the case where all the individual random allpass filters increase the peak amplitude. This situation is rare with transient signals, but quite possible if the signal contains a large steady-state component – especially one which has a noise-like spectrum.

1) Delay-line length selection: Initially, it was assumed that the delay-line lengths should be selected from a list of values that minimise co-incidence between the impulse responses of the first-order allpass filters in the chain, and hence co-incidence between peaks in their group-delay responses. This would be achieved by using mutally prime values (or better, prime n-tuplets). However, experimentation showed that given the fairly restrictive bound on delay-line length (generally < 40 samples at a sampling rate of 44.1kHz) necessary to keep the maximum group delay within the inaudible range, this approach yielded too few possible delay-line lengths. Better peak-amplitude reduction was obtained when all integer delay-line lengths from 1 up to a maximum d_{max} were allowed.

2) Results: Figure 3 shows the results of the above process when applied to a number of isolated clean musical sounds. The sounds consist of single drum hits recorded from a Roland TR-808, a single piano note played at C3, and a single synthesised mallet-like sound, played at C3. All input signals have a sample rate of 44.1kHz, and are normalised so that their peak amplitude is 1. An allpass chain length of $M = 3$ was used, with $d_{max} = 30$. The number of parallel structures is $N = 100$. Peak and RMS amplitude values for the unprocessed and processed signal are given in Table I. As can be seen, the structure appears to be successful in reducing the peak amplitude by a moderate amount, whilst leaving the RMS amplitude unaffected. Careful listening reveals no audible difference between the processed and unprocessed signals in all cases apart from the bass drum. In the bass drum case, the initial transient of sound is slightly altered in timbre. This can be corrected whilst still producing a reduction in peak amplitude by reducing the value of d_{max} to 25. The source samples and their processed versions are available at the accompanying website[1].

3) Combination with non-linear processing: The proposed method gives a modest reduction in peak amplitude. In practice, this reduction may not be enough to produce the desired increase in loudness. It is therefore natural to combine the process with a standard non-linear method of limiting. The peak reduction provided by the allpass system will reduce the amount of work that the non-linear process needs to do, and hence lower the amount of distortion introduced. For

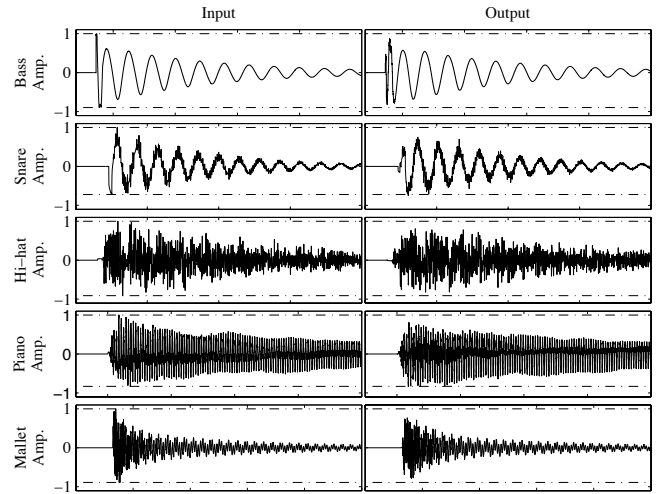[1]Supplementary material available at http://ieeexplore.ieee.org.

Fig. 3. Waveforms of a number of isolated musical sounds, before and after being processed. The horizontal dotted lines show the positive and negative peaks of the waveform before processing.

evaluation, we simulate this situation. First, we normalize the input samples to have an RMS amplitude of -5 dBFS (where 0 dBFS is ± 1). This is achieved by calculating the RMS amplitude of the signal in a 1000 sample rectangular window centered on the peak amplitude location. We know the difference between this value and the desired normalized level, and hence we can calculate the amplification necessary. The resulting signal will have samples exceeding ± 1.

We then produce two different test output signals. One is processed by hard-clipping function, which restricts the maximum signal value to ± 1. The other is processed by the allpass chain structure, followed by the hard-clipper. We measure the difference between the frequency content of both of the limited samples and the unprocessed samples. This is done by calculating the Euclidean 2-norm between their normalised magnitude spectra.

We can compare the calculated spectral difference for both of the samples, and hence evaluate how the distortion has been reduced. The last column of Table I shows the reduction in spectral difference given by the allpass filter chain for each input signal, quoted as a percentage. A significant reduction in distortion results, with the process having the most success on strongly harmonic signals (like the bass drum or piano), and least success on noisy signals like the hi-hat.

B. Mixed Musical Audio

When processing longer sections of mixed musical audio instead of individual sounds, it is unlikely that a single allpass

TABLE I
TABLE SHOWING RMS AMPLITUDES, PEAK AMPLITUDES AND DISTORTION REDUCTION PERCENTAGES FOR SEVERAL SIGNALS BEFORE AND AFTER PROCESSING. DISTORTION REDUCTIONS PERCENTAGES ARE CALCULATED SEPERATELY, AS DESCRIBED IN SEC. III-A3.

Sound	RMS(in)	RMS(out)	Peak(in)	Peak(out)	Dist. Reduc.
Bass	0.12	0.12	1.0	0.86	85%
Snare	0.11	0.11	1.0	0.75	41%
Hi-hat	0.12	0.12	1.0	0.82	12%
Piano	0.20	0.20	1.0	0.83	48%
Mallet	0.96	0.96	1.0	0.79	29%

chain will be appropriate for reducing the peak amplitude of the whole signal. Instead, we must segment the audio somehow, and apply the best allpass chain for each section. The processed segments should then be joined in as seamless a manner as possible to produce the final output.

1) Segmentation: The simplest approach to segmenting the audio would be to divide it regularly. However, this may present problems if segment boundaries fall too close to the position of a transient. A more advanced solution is to segment the audio based on the position of transients. A variety of methods are available for transient detection [14]. For the purpose of this work, we detect transients using a simple scheme. The input signal is passed through a standard envelope detection algorithm, and transients are identified as points where the derivative of this envelope takes on a large positive value. A new segment is started at an approximate number of samples (e.g. 500) before the detected transient, adjusted to lie as near to a zero-crossing as possible. The segment lasts until the next detected transient. This technique can be applied during off-line or real-time processing. In the real-time case, the signal must be buffered to allow the segment start to precede the detected transient.

2) Recombination: Once the individual segments have been processed, they must be recombined into an output signal. This could be achieved by switching to the output of the allpass chain which is producing the lowest peak amplitude for the segment under consideration. However, this will produce a small glitch on the boundary between sections. This artifact can be suppressed by crossfading between the outputs of the two allpass chains. The crossfade time should be kept short, as during the crossfade the combination of the signals will produce a filtering effect. In practice, a crossfading time of 1 ms (44 samples at the 44.1-kHz sample rate) is sufficient.

3) Results: To test the proposed method in practice, the process described above was applied to three 16-second long passages of mixed musical audio consisting of drums and synthesized chord and bass sounds. The passages are mixed from the instrument tracks without any dynamics processing. Fig. 4 (a) shows a scatter plot comparing the peak amplitudes of individual segments of the signals, before and after processing. There are 385 segments in total. The data is presented in this way as an analogy to the familiar method of presenting the input-output gain relationship of a non-linear compressor [2]. Fig. 4(b) shows a histogram of segment peak amplitude reductions. The RMS amplitude of the segments is unaltered by the process. After processing, the three passages have absolute peak amplitude values of 0.67, 0.81 and 0.76 respectively compared to 1.0 before processing. The average reduction of the peak amplitude of the passages is hence 2.5dB. Processed and unprocessed musical passages are available at the accompanying website[2], to verify transparency.

Experiments with mixed musical content that have already undergone heavy dynamic range compression or limiting and hence already have a low peak-to-RMS amplitude ratio show that these signals do not see any improvement through processing by this method.
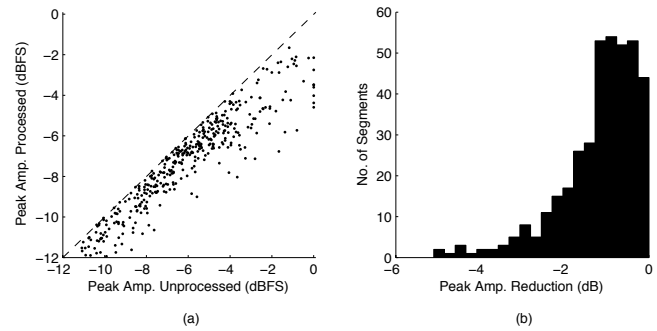
[2]http://ieeexplore.ieee.org

Fig. 4. *(a)* Scatter plot showing peak amplitudes of audio segments, before and after being processed. *(b)* Histogram of segment peak amplitude reductions.

IV. CONCLUSIONS

In this work we have presented a novel method of reducing the peak amplitude of transients in an audio signal by the application of a parallel set of allpass filter chains with golden-ratio coefficients and randomly generated delay-line lengths. Testing indicates that this method allows some reduction of peak amplitude of both isolated and mixed musical audio without coloration of the signal. This reduction is not large enough to replace non-linear limiting or compression methods, but can serve as a useful compliment to them, reducing the distortion associated with these methods by a noticeable amount. Ideally the process would be applied to sub-mixes within a larger musical mix or to the entire mix before any other heavy compression or limiting has taken place.

REFERENCES

[1] E. Vickers, "The loudness war: Do louder, hypercompressed recordings sell better?," *J. Audio Eng. Soc*, vol. 59, no. 5, pp. 346–351, May 2011.
[2] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital dynamic range compressor design: A tutorial and analysis," *J. Audio Eng. Soc.*, vol. 60, no. 6, pp. 399–408, Jun. 2012.
[3] A. Härmä et al., "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc*, vol. 48, no. 11, pp. 1011–1031, Nov. 2000.
[4] V. Välimäki, J. Parker, and J. S. Abel, "Parametric spring reverberation effect," *J. Audio Eng. Soc*, vol. 58, no. 7/8, pp. 547–562, Jul. 2010.
[5] J. M. Kates and K. H. Arehart, "Multichannel dynamic-range compression using digital frequency warping," *EURASIP J. App. Sig. Process.*, vol. 18, pp. 3003–3014, 2005.
[6] M. Schroeder, "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," *IEEE Trans. Information Theory*, vol. 16, no. 1, pp. 85–89, Jan. 1970.
[7] A. Van den Bos, "A new method for synthesis of low-peak-factor signals," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 35, no. 1, pp. 120–122, Jan. 1987.
[8] A. Horner and J. Beauchamp, "A genetic algorithm-based method for synthesis of low peak amplitude signals," *J. Acoust. Soc. Am.*, vol. 99, pp. 433–443, 1996.
[9] A. Horner, "Low peak amplitudes for wavetable synthesis," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 4, pp. 467–470, Jul. 2000.
[10] T. F. Quatieri and R. J. McAulay, "Peak-to-RMS reduction of speech based on a sinusoidal model," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 273–288, Feb. 1991.
[11] D. Griesinger, "Impulse response measurements using all-pass deconvolution," in *Proc. AES 11th Int. Conf.*, May 1992.
[12] V. Välimäki, J. S. Abel, and J. O. Smith, "Spectral delay filters," *J. Audio Eng. Soc*, vol. 57, no. 7-8, pp. 521–531, Jul. 2009.
[13] L. Savioja, V. Välimäki, and J. O. Smith, "Audio signal processing using graphics processing units," *J. Audio Eng. Soc*, vol. 59, no. 1-2, pp. 3–19, Jan. 2011.
[14] J. P. Bello, L. Daudet, et al., "A tutorial on onset detection in music signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, Aug. 2005.