14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A TIME DOMAIN SYSTEM FOR TRANSIENT ENHANCEMENT IN RECORDED MUSIC

Graziano Bertini, Massimo Magrini, Tommaso Giunti

ISTI CNR, Pisa
Via Moruzzi 1, 56100 Pisa, Italy
phone: +39 050 3153125, fax: +39 050 3152810, e-mail: graziano.bertini@isti.cnr.it
web: www.isti.cnr.it

ABSTRACT

The set of audio treatment methods commonly used in the mastering process of commercial music tend to increase the loudness perception of the final audio product, obtaining a "fat" sound that can be played with sufficient quality on low level audio devices, such as small radios or computer loudspeaker. The side effect of these audio processes are the loss of transients and dynamic variations, with a resulting "flat" sound. The widely diffused compressed audio formats (mp3, wma) introduce further degradation of the recorded music. In this paper we describe a method for time domain transient enhancement of recorded music, which can be easily implemented with low-cost Digital Signal Processors in stand alone device or included in HW (iPod like) or SW (Winamp like) audio player applications.

1. INTRODUCTION

Even if digital technologies brought us very reliables, multi-featured and cheap audio systems the overall quality of reproduced sound is worse than it was ten years ago, for many sociological reasons. The "high fidelity" is no longer a primary target, and it is harder than in the past for a HI FI enthusiast to get equipment and devices for producing a better quality in the reproduced sound.
The aim of the method described here is an attempt to recover the original quality of the recorded sound, enhancing the transients lost in today's standard mastering processes and compressions. The paternity of this effect is attributed to Gabriel Biagiotti, a DJ from Florence: during a DJ set he noted that by manually moving the faders of a graphic equalizer according to the movements of the led bars of a graphic analyzer he produced a nice dynamic effect, enriching the music especially when played at lower volumes. Having noted that the dance floor audience enjoyed this trick, he rapidly imagined a device (HW/SW) that could automate this manual process. Later he realized that this process is also interesting for HI-FI listeners at home or in mobile situations (portable mp3 players). Thus, he contacted the I.S.T.I. C.N.R. of Pisa, for a preliminary study of the project. The first version, basically realized on the base of perceptive rules, has been subsequently improved and implemented also on a floating point DSP platform. The final version of the prototype has been named ARIA [1]. In the final recording we have no exact information regarding the mastering operation, so we have to estimate it by analyzing the signal itself. So, before describing the method we need to give an overview of the various audio treatments used in the mastering phase.

2. BRIEF DESCRIPTION OF MASTERING PROCESS

Even if some of these audio treatments could be made using analog devices (as was the case before the advent of digital technologies) in this paper we will consider only the digital one.
We consider that the final mix of the recorded music is available in digital format on a disk. After the final mixing, a set of standard processing is usually performed before the real mastering of the CD/DVD. The proper name of this phase is *premastering* (the real mastering phase is performed in the CD/DVD plants), but everyone knows it as *mastering* [2]. Below we give a brief description of the main operations made during the mastering, which give the most evident alterations vs the original mix.

2.1 Compression

Compression is a process that manipulates the dynamic range of an audio signal. Compression is used in sound recording and live sound reinforcement fields to improve the perceived quality of audio. (This should not be confused with audio data compression, which reduces the data size of digital audio signals.). A compressor (HW or SW based) reduces the dynamic range of an audio signal if it becomes louder than a set threshold. The amount of gain reduction is usually determined by a ratio control. That is, with a ratio of 4:1, if the input level is 4 dB over the threshold, the gain will be reduced so that the output level will only be 1 dB over the threshold. Compressors usually have controls to set how fast the compressor responds to changes in input level, known as attack, and how quickly the compressor returns to no gain reduction once the input level falls below the threshold, known as release. Because the compressor is reducing the gain (or level) of the signal, the ability to add a fixed amount of make-up gain at the output is provided so that an optimum level can be used. Standard compression is used on the single tracks before mixing, while on the mixed signal a multiband compression is preferred.

Multiband compressors can act differently on different frequency bands. It is as if each band has its own compressor with its own threshold, ratio, attack, and release. That allows a diversified action, avoiding that a peak on the lower audio bands lead to an unwanted compression on the higher bands.
Having a good compression could have a positive effect on the SNR, especially when the music is radio-transmitted.
Also, having a louder sound is often considered an advantage in commercial competition. However, adjusting a multiband compressor requires some sense of style and professional skills. This is because the constantly changing spectral balance between audio bands may have an equalizing effect on the output, by dynamically modifying the frequency response.

2.2 Equalization

Equalization is the process of modifying the frequency envelope of a sound. Etymologically, it means to correct, or make equal, the frequency response of another audio device. The term "equalizer" is sometimes applied to audio filters in general, though strictly speaking not all audio filters are equalizers. Usually the spectrum in commercial music is cut at 16 kHz in the upper bound, and on 35/40 Hz in the lower bound. In this way inaudible frequencies does not contribute to increasing the overall volume, leaving the media dynamic range to audible bands.
In the final mixing/mastering phase equalization generally tends to make the spectrum as flat as possible. A slight dump on the high frequencies, starting from 8 kHz is often used in some genres in order to makes the sound softer. A gentle boost of the frequencies around 100 Hz is also very common, especially in pop or dance music.

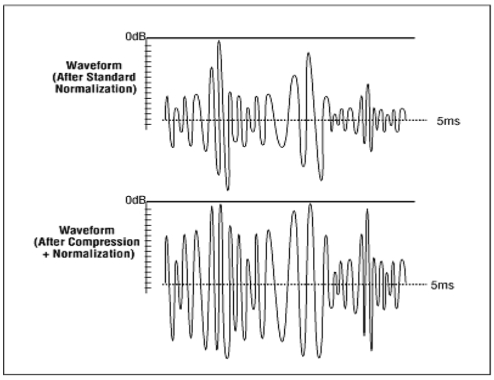


Figure 1 – Maximization effect

2.3 Maximization/Limiting and other processing

This part of the mastering involves the use of a special processing often called Look Ahead Limiting. This processing in some ways is similar to high ratio compression [2]: it acts on the dynamics of the signal, reducing peaks when they occur. The difference basically resides in the use of a delay in the signal, before its dynamic reduction. Instead, the envelope detector uses the non-delayed signal as input, so that it start to act before the peak occurs, with a better behaviour than a simple limiter. Other common processings, whose description is beyond the scope of this paper are enhancing (for artificially creating higher harmonics), dithering (a technique for reducing quantization error, when doing the final 16 bit master on CDs, for example), and the rather infamous clipping-like technique called shred.

2.4 Final considerations

After all these processings the resulting sound seems much louder (Fig.1), and can be played with quite good quality also on small speakers (radios, TV, PC boxes etc.).
As Bob Katz said in Mastering Audio [2] we should note that the perceived loudness difference between the 1990 and 1999 CDs is greater than 6 dB, though both peak to full scale. Listening to a 2005 pop music CD, we can easily see that the all the meter lights come on, and remain there the whole time. The average level of popular music compact discs continues to rise (Fig.2). Popular CDs with this characteristic are becoming increasingly prevalent, coexisting with discs that have a beautiful dynamic range and impact, but whose loudness (and distortion level) is far lower. There are many technical, sociological and economic reasons for this chaos.

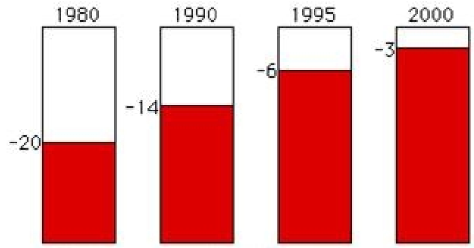


Figure 2 – Change in loudness perception from 1980

On the other hand this overcompression, even if makes the sound louder, leads to a flat, poorer sound, with fewer transients. In addition, nowadays most recorded music is listened to in compressed audio formats (mp3, wma). The resulting sound is even worse. So, there is a need of a method for attempting to automatically restore the original sound of the final mix. We should note that the compression applied to individual tracks cannot be restored correctly, because the compression information is lost when the tracks are mixed together. Anyway, the compression applied to each track could be considered part of the "art" product, a result of the choices made during the production of the records, so it would not be altered. Instead, it is possible to make an attempt to restore part of transients lost during the compression and limiting of the whole mix.

3. ARIA ALGORITHM

The proposed method consists, firstly, in detecting two envelopes with different response time, which give us a dynamic representation of the amplitude level of the audio spectrum sub bands. We need to split the whole spectrum into several sub-bands because the attack time of the faster

envelope is strictly correlated to the compression attack time used in the mastering phase, and it depends on the distribution of the energy in the audio spectrum. Then an index of instantaneous variability, representing the shape of the transient is obtained by comparing these two envelopes and it is used to control the dynamic of the signal.
Note that here we are not interested in detecting the transients for event detection [3], (there are better methods for this, for example based on Principal Components Analysis method) but for a rapid audio processing purpose.

3.1 Block diagram

The algorithm could be applied in stereo or, more generally, multi channel systems. Here, for simplicity's sake, we will describe only one processing channel.

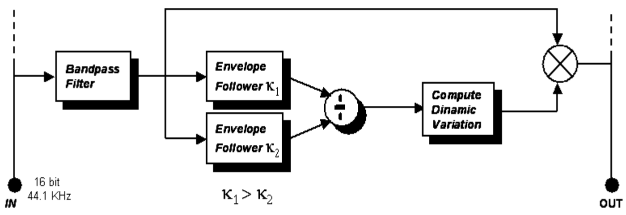


Figure 3 – ARIA Block diagram

Firstly, the signal is divided into different audio bands (Fig.3). In order to have a complete coverage of the audio spectrum we used 10 pass-band filters (Fig.4): each band has a centre frequency double the previous. An ideal pass band filter could be made with a cascade of two Butterworth filters: one high pass and one low pass.

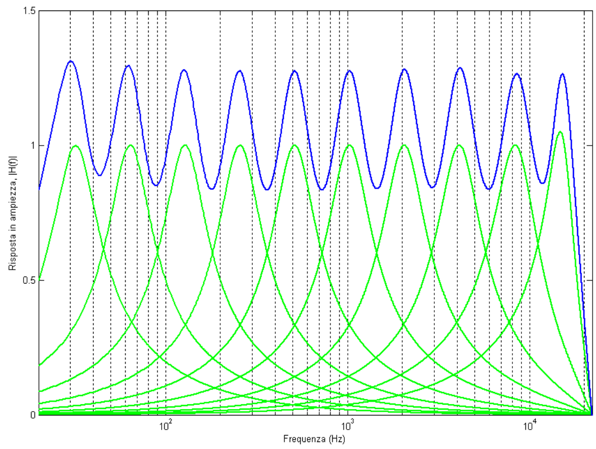
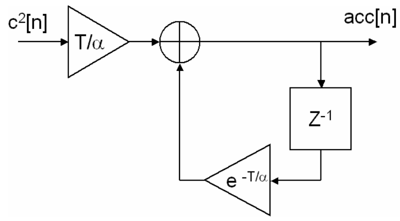


Figure 4 – Overlapped bandpass filters

In our implementation we decided to replace the couple of Butterworth filters with a bell-shaped biquad filter, a bit more optimized for real time applications. From each filter output then we have to extract two envelopes, with different response time. In order to perform this operation we can build a simple envelope follower squaring the sample, and filtering the result with a simple, first order, low pass filter, as shown in the following diagram.



If T is the sampling period and $\alpha = 1/(2\pi f_c)$, with f_c is the cutoff frequency, defining :

$$K = \frac{T}{\alpha} < 1$$

For small values of K we can say that:

$$e^{-K} \cong 1 - K + \frac{K^2}{2} \qquad (1)$$

So, given the samples *c[n],* for band *i* we can compute the current envelope *acc[n]* in this way:

$$acc_i[n] = K \cdot c_i^2[n] + (1 - K + \frac{K^2}{2}) \cdot acc_i[n-1]$$

The K constant is directly linked to cutoff frequency of the envelope follower filter: lower values allow cutting higher frequencies in the envelope signal, leading to a smoother envelope.
A bigger K makes the envelope follower more reactive, and less smooth. The K value have to be carefully chosen, with a compromise between smoothness and fast response. The seconds stage simply computes another envelope, exactly in the same way but with a constant K2 < K1, obtaining a slower response time (Fig.5).

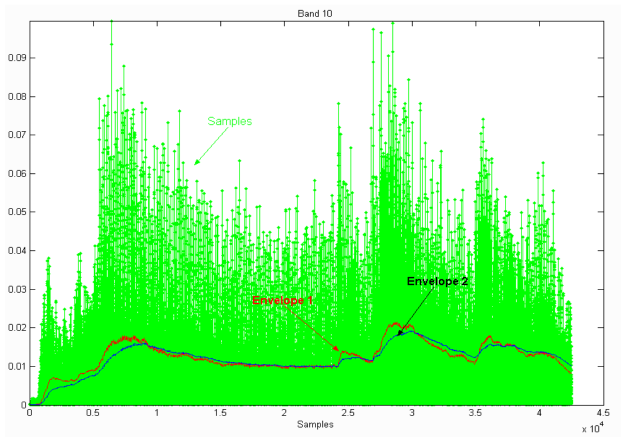


Figure 5 – The two detected envelopes

For the higher bands the cutoff frequency is far from the envelope cutoff, and the (1) could be simplified as:

$$e^{-K} \cong 1 - K$$

Thus we have

$$acc_i[n] = K \cdot c_i^2[n] + (1-K) \cdot acc_i[n-1]$$

To obtain a better response we could consider two sets of K coefficients: one for the attack and one for the release part of the envelope. This is because for most of instrumental sounds the attack part is much faster than the release.

Another strategy for computing the envelope is based on the moving average method. So, the current envelope acc[n] could be expressed as

$$acc_i[n] = \frac{1}{M} \cdot \sum_{k=0}^{M-1} c_i^2[n-k]$$

Where M is the length of the average window.
The envelope obtained with this method is much smoother and precise, especially in the lower bands (Fig. 6).

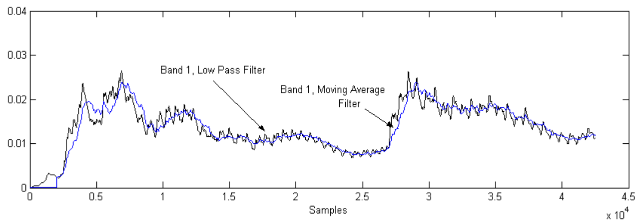


Figure 6 – Envelope with moving average

Basically, filtering the squared samples signal with the mobile average is like convolving it with a rectangular pulse having 1/M as value.
Frequency response is:

$$H(f) = \frac{\sin(\pi f M T)}{M \cdot \sin(\pi f T)}$$

with $B_{-3} \cong 0{,}3/(M \cdot T)$

So, the M parameter is inversely linked to the envelope follower response speed.
A bigger M makes the envelope smoother and faster, while using a small M we can detect faster transitions. With this method it is more difficult to have an efficient technique for diversifying the attack and release time. Nevertheless, the overall results are quite good on both phases. As in the previous type of follower we can compute two different envelopes, a fast and a slow one, using a M2>M1 (Fig.7).

In our implementation we decided to use the moving average method for the fast envelope, more critical for the overall quality, and the lowpass method for the longer one.

When an attack transient occurs, the faster envelope will increase more quickly than the other. Instead, in the release phase, the faster envelope will go down rapidly, faster with respect to the other envelope.

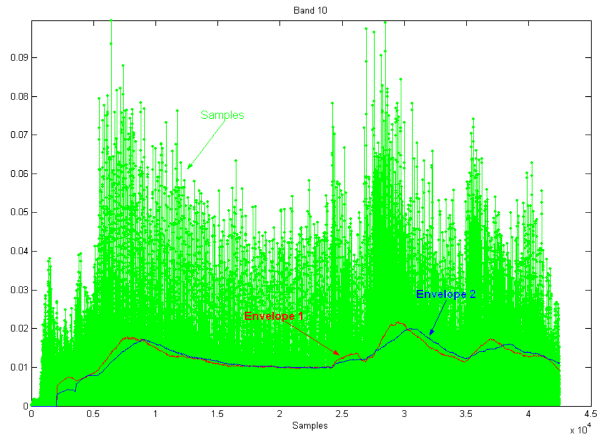


Figure 7 – Envelopes detected with moving Average

This behaviour suggests comparing the two envelopes to detect the transients and their size. We need to compute an index which describes the shape of the transient, in way that is not dependent on the absolute loudness level. If we compute the ratio of the two envelopes:

$$R_i[n] = \frac{acc_i[n]}{acc_i^{long}[n]}$$

we can easily obtain an index with this characteristic (fig. 8).

If Ri[n] > 1, band level is increasing; if Ri[n] < 1 it is decreasing; while if it is close to 1 the level is stationary.

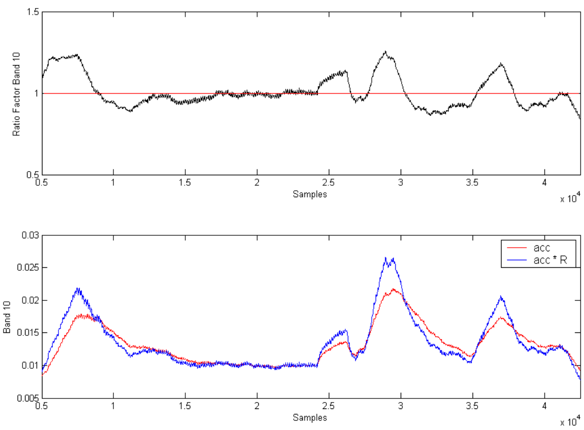


Figure 8 – The R[n] index vs. the two envelopes

The computed index $R_i[n]$ will be used as a multiplier for the signal coming from the band filter output as follows:

$$out_i[n] = c_i[n] \cdot R_i[n]$$

In this way, when a rising edge of the signal occurs, the signal will be amplified, and thus the transient will be enhanced. In the falling edge we will obviously have the inverse behavior. Once all the bands are processed in this way, their output signal will be mixed together again, obtaining a signal with enhanced dynamic transitions.

$$OUT[n] = \sum_{i=0}^{10} out_i[n] \qquad (2)$$

The value of Ri[n] can be very high: for example with uncompressed rhythmical instruments we could reach a index level of 100: this value is too high to be used for enhancing the transients. Thus, it is necessary to limit the values of Ri[n] index. We experimentally found that 0.5 and 2 are quite good limit values. In this way the effect is limited to +/- 6dB on the original signal. Anyway, these limits can be slightly altered according to personal taste and music genres. Another option for setting the amount of effect is setting the time of the slower envelope by making it closer to the short one we can reduce the amount of the effect.

4. APPLICATIONS

With a adequate engineering this method could became the basic principle of a consumer product which could be integrated into common audio systems.
This method is time-domain-based and does not require a very powerful processor for achieving optimal results.

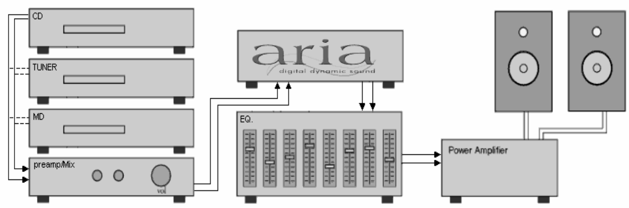


Figure 9 – ARIA based device inserted in a standard HI-FI

Having experience in DSP system development [4][5], We successfully attempted to implement the algorithm on a low-cost DSP platform (Analog Devices 21061 DSK board), proving that it could easily be integrated into today's existing digital audio processors, with little CPU time demand. The commercial system could be conceived as a stand-alone unit, to be inserted in a standard HI-FI chain (Fig.9) as well as in the processing firmware of portable mp3 players. We also attempted to implement the algorithm as a plugin for mp3 player PC application (Winamp), having the same quality as in the DSP base version.

4.1 Further developments

This algorithm could be further optimized, and the quality of its processing improved. The update rate of Ri[n] index on lower bands is quite low so an interpolation between close different coefficients is required and still must be implemented. A "look ahead" technique could be used in order to be better synchronized to fast transients: i.e. the Ri[n] index should be applied to a slightly delayed (few mS) version of the band signal.
For example, the method could store the history of the latest index coefficient, automatically adjusting the amount of the effect according to the music genre (more or less dynamic), designing a fast algorithm based on state-of-the-art feature extraction methods [6][7].
Finally, by giving the user control over each band gain the algorithm could also work as a graphic equalizer.

5. CONCLUSIONS

The goal of this method is to enhance the transients that were heavily modified (and, depending on the music genre, quite deteriorated) in the mastering phase and/or in the audio file compression. The method differs from a typical dynamic expander because it is not dependent on the audio signal loudness, where the expander starts to acts only where some fixed thresholds are crossed. Acting only on short transients it does not alter the equalization carefully made during mastering, described in 2.2.
A preliminary prototype has been implemented and tested on a floating DSP platform, and in a SW based (Winamp plugin) version. The amount of the processing can be easily controlled acting on a single parameter, in a intuitive way.
This, together with the fact that the algorithm itself is not CPU demanding makes it a good starting point for a commercial applications.

ACKNOWLEDGEMENTS

We would like to thank Gabriele Biagiotti for giving us the possibility of working on his idea. We also thanks Sergio Taglioni from StudioLab (Cascina, Italy) for his audio quality evaluations of the ARIA effect.
This paper is presented with the support of EU MUSCLE-NOE (FP6-507752).

REFERENCES

[1] ARIA Italian patent n. 3957PTIT/2, 2003
[2] Bob Katz, *Mastering Audio, The Art and the Science*. Focal Press Book, 2002
[3] Laurent Daudet, "*A review on techniques for the extraction of transients in musical signals*" in Pre-Proc. Of third Int. Simp. On Computer Music Modeling and Retrieval (CMMR 2005), S.Ystad, R. Kronland-Mertinet, T.Voinier, ISTI-CNR, Pisa, Italy, Sep. 2005.
[4] Graziano Bertini, Massimo Magrini, "*A Prototype Lab Box with DSK'C6711/13 for Rapid DSP Algorithm Development"* Texas Instruments, Tech-On-Line web site, Audio and Video/Imaging Series, October 2005.
[5] Graziano Bertini, Massimo Magrini, Diego Gonzales Lorenzo Grassi, "*Uso di tools per "Rapid Prototyping" nel-lo sviluppo di algoritmi DSP complessi"* in Proc. on CD of the DSP Application Day 2005, DICO, Milano Sep. 2005.
[6] Graziano Bertini, Massimo Magrini, "*Spectral Data Management Tools for Additive Synthesis"* in Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-00), pp. 195-200, Verona, Italy, December 7-9, 2000,
[7] Thomas Lidy, Andreas Rauber., "*Evaluation of Feature Extractors and Psycho-acoustic Transformations for Music Genre Classification*" in: Proc. of the Sixth Int. Conference on Music Information Retrieval (ISMIR 2005), pp. 34-41, London, UK, Sep. 11-15, 2005.