Audio Engineering Society

Convention Paper

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Transient Detection for Transform Domain Coders

Venkata Suresh Babu, Ashish kumar Malot, Vijayachandran V M and Vinay M K

Multimedia Technologies division, Emuzed India Pvt. Limited, Bangalore, India

pvsuresh@emuzed.com, malot@emuzed.com, vijay@emuzed.com, vinaymk@emuzed.com

ABSTRACT

State of the art audio encoders are based on transform-domain coding algorithms. Due to time-frequency uncertainty, transform domain coders suffer from “pre-echo” and “diffusion” artifacts during transient portions of the signal. These artifacts occur because of large transform lengths used to achieve higher coding gains. Audio encoders employ various tools such as adaptive transform lengths, TNS etc to efficiently code transient portions of the audio signal. Typically audio signals have time domain transients (e.g. castanets), frequency domain transients (e.g. flute, clarinet) and transients observed in speech signals during consonant to vowel transitions etc. Identification of these transients in an audio signal is vital to achieve perceptual quality at low bit-rates. This paper discusses the various transient classes present in audio signals, apart from describing a transient detector employed for efficient modeling of all classes of transients. The proposed transient detector has been incorporated in MPEG-4 AAC encoder, independent of the psycho-acoustic analysis methodology used. Listening tests as well as OPERA scores indicate substantial improvement in audio quality, over the baseline encoder.

1. INTRODUCTION

With the emergence of multimedia applications in consumer electronics and mobile communication, audio compression technology assumes significance. Current audio coders use time-frequency representation for exploiting the quasi-stationary property of the audio signals for efficient compression. Most of the current audio coders use transform-domain coding approach. Such approaches provide high frequency resolution to achieve better coding gains by employing psycho-acoustic principles effectively. But these representations suffer from pre-echo and frequency diffusion artifacts during transient portions of the signals.

To overcome these problems transform domain audio coders employ transient modeling techniques such as adaptive transform lengths, temporal noise shaping etc. Efficiency of such modeling techniques depends on the effective detection of transients in the audio signals. Hence transient detectors play a major role in improving the quality of an audio encoder. Traditional transient detection solutions are based on variation in time-domain energy function alone [1-3]. These methods fail in detecting *frequency domain transients* with no significant change in time-domain energy. This paper proposes a novel transient detector addressing the above problem. The proposed transient detector has been successfully embedded in the MPEG-4 AAC encoder and is found to improve the quality (by both subjective and objective measures) of the encoded streams over the baseline encoder. Methods employed in the proposed transient detector are computationally cheap and reuse the basic blocks already available in the audio encoders.

In the following section a brief background on the various types of transients are discussed, along with the kind of problems they pose with encoding quality. Section 3 discusses in detail the transient detection methodology proposed. Implementation issues and results indicating the performance of our transient detector are presented in section 4. Section 5 lists our conclusions based on the analysis and experimental results.

2. BACKGROUND

Traditionally transients/attacks are regarded as signal segments, whose time-domain energy function rapidly changes from low to a high value [2]. Transient detection algorithms based on this definition, chose an energy-based criterion to detect transient in the signal. The energy function (*ef*) is defined as,

$$ef(n) = 1/L \sum x^2(k), \;\; k = n\text{-}L/2 \text{ to } n\text{+}L/2$$

The transient character of the signal is determined based on some simple measures like,

$$C(n) = \log(ef(n)) - \log(ef\,(n\text{-}1)) \;\; n = 0 \text{ to } N\text{-}1$$

Where local energy function window length (L) < analysis window length (N)

Figure 1 (castanets) shows the time-domain transient portion of the signal (1.a) along with the corresponding energy function (1.b). The need to detect and treat a transient portion of the input signal can be observed by the pre-echo (smearing effect in time domain) it introduces (1.c) by normal encoding process. Using a "large" window for encoding during a transient portion of the signal results in pre-echo across the analysis window. The energy-based transient detection algorithm performs well in detecting mainly wide-band noise [4] like transients.

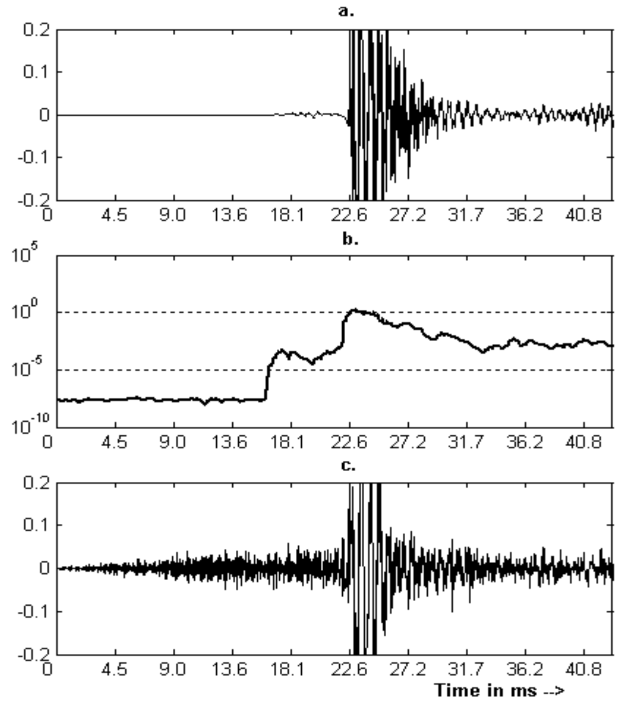

Figure 1. (a) Input audio signal with time domain transient (b) Time-domain energy function (c) Output audio signal with pre-echo artifact

Many stringed and soft percussive musical instrument audio signals have transients, which go undetected by this method (E.g. violin, vibraphone, speech signals). These signals have segments with rapid changes in spectral content. Speech and speech like signals have high activity regions at short fricatives, nasals and stops. Hence it is necessary for these portions of the audio signals to be identified as frequency-domain transients. Figure 2 (vibraphone) shows one such signal portion (2.a), which qualifies to be detected as a transient. But the corresponding energy function (2.b) indicates no steep variation across the (adjacent) frames. The need to detect and treat this transient portion of the audio signal can be observed by the *frequency diffusion,* it introduces (2.c) by normal encoding process.

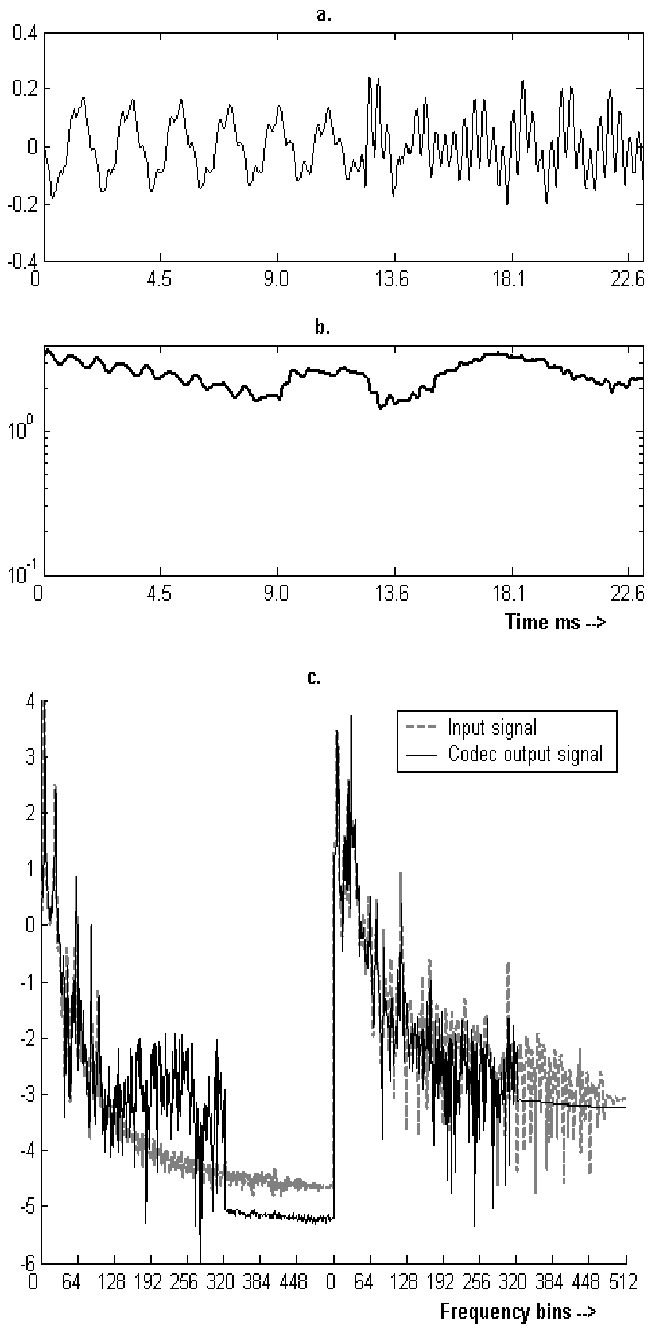

Figure 2. (a) Time domain signal. (b) Energy function in time domain (c) Spectral plots of two consecutive segments of the signal.

3. TRANSIENT DETECTION ALGORITHM

The proposed transient detection algorithm not only detects the time-domain transients but also the frequency domain transients in music signals. Time-domain transients are also reflected in change in spectral content of the underlying signal. Transient detection algorithm is operated in the spectral domain to capture both types of transients. The signal is divided into short-time segments (frames) and for each segment spectrum is obtained. Lower part of spectrum is not used for analysis as the transients are characterized by the higher frequency content. The remaining part of spectrum is divided into uniform frequency bands (N) and for each band energy en(b) is calculated. For each frequency band a transient measure G(b) is estimated by comparing the energy of the current frame-band with the recursive updates of the energy of the previous frame-band updates $f_n(b)$ as given by the following equation,

$$f_n(b) = \alpha * en(b) + (1-\alpha) * f_{n-1}(b)$$

$$G(b) = en(b) / f_n(b)$$

$$F = \sum_{b=0}^{N-1} d(b)\ , \ \ where\ d(b) = \begin{cases} 1 & G(b) > T \\ 0 & otherwise \end{cases}$$

Transient detection decision is made based upon the transient estimates of the frequency bands. If the band energy ratio is greater than the threshold (T) transient in that band is detected. To get a more realistic estimate of the transient presence in the frame the collective effect of all band transients is considered. Determination of the thresholds is based on experimental results and empirical analysis. This transient detection algorithm has been tested across wide range of audio signals and found to be detecting the transients robustly. Figure 3 shows the reduced frequency diffusion effect in the spectral plots with the proposed transient detection algorithm, as compared to those in Figure 2.

Incase of speech signals transition from vowel to fricatives/nasals involves variation in signal activity. These transition portions result in spectral level variations between 3kHz and 9kHz. The frequency-domain based transient detection algorithm can detect these portions as well. To make the detection algorithm more robust against such variations a simple zero crossing measure is used in parallel.

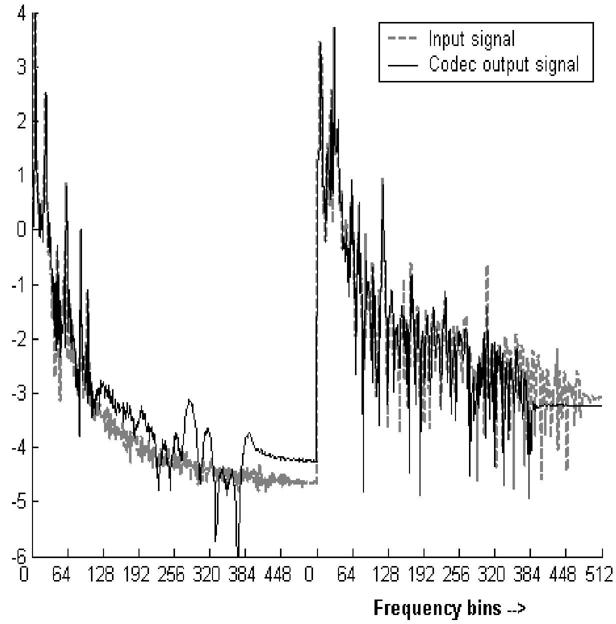

Figure 3. Spectral plots of two consecutive segments of the signal with the proposed transient detection algorithm.

4. IMPLEMENTATION AND RESULTS

The proposed transient detection algorithm is incorporated into MPEG-4 AAC encoder [5] as a stand-alone module. In AAC coding the transient modeling can be performed by adopting various techniques like window switching, temporal noise shaping etc. The standard encoder has a block-switching module, which decides the window size based on the psycho-acoustic model parameters. But the need for stand alone transient detector arises when implementing low complexity encoders without psycho-acoustic models. The module chooses the window size based on the decision of the above transient detection algorithm. The algorithm is tested on the EBU SQAM audio database [6]. Performance of the proposed transient detector is compared against time domain energy based transient detector [2] by incorporating them as stand alone units in MPEG-4 AAC encoder. OPERA ODG scores are used as the performance measure for comparing the two algorithms. The OPERA ODG (Objective Difference Grade) scores give the estimates of the objective quality measurement that relates very much to the perceived audio quality. On a scale of –5 to 0, '0' indicates the transparent quality and '-5' indicates poor quality. From the ODG scores (Table 4.1) it is clear that the proposed frequency based transient detector outperforms the traditional energy based methods. These results further substantiated by the ODG vs. Time plots as in Figures (a) and (b).

Audio signal	ODG for time energy function based method	ODG based on the proposed method
Vibraphone	-1.032	-0.391
Glockenspiel	-0.922	-0.66
Triangle	-0.785	-0.455
Tuba	-0.70	-0.391
Xylophone	-0.67	-0.1985
Trumpet	-0.58	-0.09
Speech Female English	-0.48	-0.085
Violin	-0.45	-0.09

Table 1. Opera ODG scores for EBU SQAM database

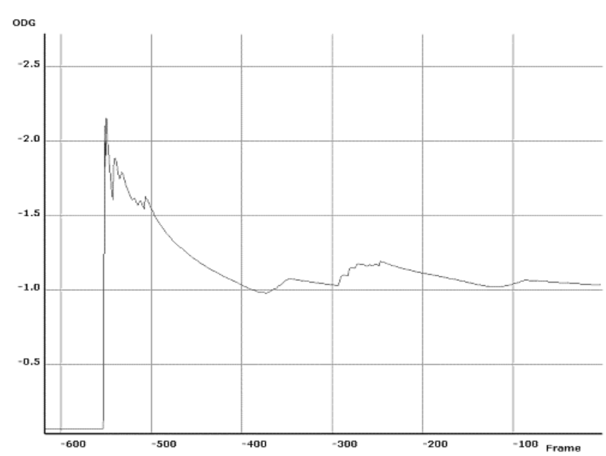

Figure 4. (a) ODG Vs Time plot (vibraphone) with energy based transient detection algorithm

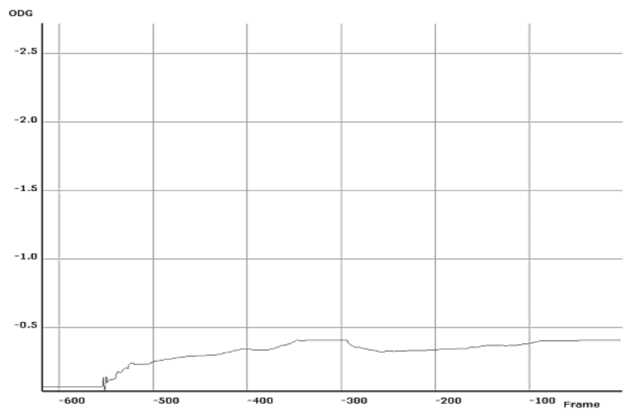

Figure 4. (b) ODG Vs Time plot (vibraphone) with the proposed transient detection algorithm

5. CONCLUSIONS

This paper discusses various types of transients encountered in audio signals as applied to audio coding. The concept of frequency domain transient was introduced. The existing energy based transient detectors fail to detect these frequency domain transients. The paper proposes a frequency domain transient detector algorithm to counter this problem. The proposed transient detector not only takes care of the attack transients but also the frequency domain transients, which cause diffusion or frequency smearing artifacts. This transient detection algorithm has been incorporated into MPEG-4 AAC encoder and tested with EBU SQAM database. The proposed transient detector algorithm is found to improve the objective quality measure over the traditional transient detection methods.

6. ACKNOWLEDGEMENTS

The authors would like to thank Multimedia Division, Emuzed India Pvt. Limited, for the support during the work.

7. REFERENCES

[1] R. Vafin, R. Heusdens, S. van de par and W. Bastiaan Kleijn, “Improved Modeling of Audio Signals by Modifying Transient Locations” ,*IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001

[2] J. Hliewer and A. Mertins, “Audio Subband Coding with Improved Representation of Transient Signal Segments” *Proc IX European signal processing conference (EUSIPCO-98)*, sept. 1998, Rhodes, Greece, pp 1245-1248

[3] P. Masri and A. Bateman, “Improved Modelling of Attack Transients in Music Analysis-Resynthesis”, *ICMC*, 1996

[4] C. Duxbury, J. P. Bello, M. Davies and M. Sandler, “ A Combined Phase and Amplitude Based Approach to Onset Detection for Audio Segmentation”, *Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-03)*, London, UK. April 9-11, 2003

[5] ISO / IEC 14496-3:2001 Information technology - Generic coding of moving pictures and associated audio information: - Part 3: Advanced Audio Coding (AAC).

[6] Sound Quality Assessment Material Recordings for Subjective Tests, Users' handbook for the EBU - SQAM Compact Disc Tech. 3253-E April 1988