# AUDIO ENCODING BASED ON THE EMPIRICAL MODE DECOMPOSITION

*Kais Khaldi[14], Abdel-Ouahab Boudraa[23], Monia Turki[1], Thierry Chonavel[45] and Imen Samaali[1]*

[1]Unité Signaux et Systèmes, ENIT
BP 37, Le Belvedre 1002 Tunis, Tunisia
email: kais.khaldi@gmail.com, m.turki@enit.rnu.tn, imen_samaali@yahoo.fr
[2] IRENav, Ecole Navale / [3] $E^3I^2(EA3876)$, ENSIETA
Groupe ASM, Lanvéoc Poulmic, BP600, 29240 Brest−Armées, France
email: boudra@ecole-navale.fr
[4] Institut Télécom; Télécom Bretagne - LabSTICC UMR / [5] Université européenne de Bretagne
BP 832, 29285 Brest Cedex, France
thierry.chonavel@telecom-bretagne.eu

## ABSTRACT

This paper deals with a new approach for perceptual audio encoding, based on the Empirical Mode Decomposition (EMD). The audio signal is decomposed adaptively into intrinsic oscillatory components by EMD called Intrinsic Mode Functions (IMFs), which can be fully described by their extrema. These extrema are encoded after an appropriate thresholding scheme controlled by a psycho-acoustic model. The decoder recovers the original signal after IMFs reconstruction by means of spline interpolation and their summation. The proposed approach is applied to different audio signals and results are compared to wavelets and to MPEG1-layer3 (MP3) approaches. Relying on exhaustive simulations, the obtained results show that the proposed compression scheme performs better than the MP3 and the wavelet approach in terms of bit rate and audio quality.

## 1. INTRODUCTION

Audio coding at low bit rate and high fidelity is an important task in many applications, such as digital audio broadcasting, multimedia and satellite TV. Achieving low bit rates while ensuring a good audio quality of the decoded signal is an actual challenging problem. Different sub-band coding and transform coding approaches [1] have been proposed for reducing the bit rate. Some of these methods achieve perceptually transparent coding at approximately 96 kb/s. To achieve high perceptual quality, at lower bit rates, approaches presented in [2]-[3] use the masking property of the ear that avoids to encode the non audible signal features. New methods of audio compression based on wavelet have been proposed in [4]-[5] to reduce bit rate requirements. However, for wavelet compression, the basis function is fixed in advance, which makes the compression efficient only over particular classes of non stationary signals.

Recently, a new data driven method, called Empirical Mode Decomposition (EMD), has been introduced by Huang and al. for analyzing nonlinear and non-stationary signals [6]. The EMD is based on the sequential extraction of energy associated with various intrinsic time scales of the signal, called Intrinsic Mode Functions (IMFs), starting from finer temporal scales (high frequency IMFs) to coarser ones (low frequency IMFs). The basis functions of EMD are derived from the signal, in contrast to traditional methods where the basis functions or the filters bandwidth are fixed. The total sum of the IMFs matches the signal very well and therefore ensures completeness [6].

The IMFs are oscillating functions, that can be fully described by their extrema. So, they can be recovered easily from extrema by using spline interpolation. Thus, the EMD seems to be a very interesting decomposition tool to use for a low bit rate sub-band audio coding.

The paper is organized as follows : section II recalls basic ideas of EMD and IMFs. Section III describes our approach for audio compression, while section IV provides simulation results and comparisons with MP3 coder and wavelet based compression.

## 2. EMD BASICS AND ITS INTEREST IN AUDIO CODING

The EMD decomposes a given signal $x(t)$ into a sum of IMFs through an iterative process called *sifting*. Each IMF has a distinct time scale [6]. The decomposition is based on adaptive basis functions derived from $x(t)$. The EMD has similarities with multi-resolution decomposition techniques : each IMF replaces the signal details, at a certain scale or frequency band [7], [8].
By definition, an IMF satisfies two conditions :

1. the number of extrema and the number of zero crossings may differ by no more than one.
2. the average value of the envelope defined by the local maxima, and the envelope defined by the local minima, is zero.

Thus, locally, each IMF contains lower frequency oscillations than the previously extracted one [7] . To be successfully decomposed into IMFs, the signal $x(t)$ must have at least two extrema (one minimum and one maximum). The IMFs are obtained using the following algorithm (sifting process) [6]:

- identify all extrema of $x(t)$.
- interpolate between minima (resp. maxima), ending up with some envelope $e_{min}(t)$ (resp $e_{max}(t)$).
- compute the average $m(t) = (e_{min}(t) + e_{max}(t))/2$.
- extract the detail $d(t) = x(t) - m(t)$.
- iterate on the residual $m(t)$.

The signal $x(t)$ is then described as follows:

$$x(t) = \sum_{j=1}^{C} \mathrm{IMF}_j(t) + r_C(t) \quad (1)$$

where $\mathrm{IMF}_j(t)$ is the IMF of order $j$ and $r_C(t)$ is the residual. The $C$ value is determined automatically using the stopping criterion standard deviation $SD$. Usually, SD is set between 0.2 to 0.3 [6].

As earlier recalled, the IMFs are zero mean and have oscillating shape properties. With a view to compression, these are interesting features. Indeed, most relevant information of the IMF can be represented by its extrema. Roughly, this aims to sample the IMF almost regularly at twice its frequency. Figure 1 shows the plots of an IMF and its approximate obtained by spline interpolation of the extrema. A comparative examination of the actual IMF and its estimate shows the effectiveness of the spline interpolation for the reconstruction of the IMF. Indeed, we see that the error corresponding to the difference between the true and the reconstructed IMF is negligible.

## 3. IMFS ENCODING

The proposed coding scheme is shown in figure 2. The first step consists in a segmentation of the audio signal into frames of length 512 samples [3]. Then, each frame is decomposed by EMD into IMFs and a residual. Since the residual represents a trend, and the audio signal IMFs is zero mean, it is clear that the residual $r_C(t)$ can be neglected and only the IMFs are considered for encoding.

In a second step, the number of extrema for each IMF are reduced by using an appropriate masking threshold fixed so that the energy of IMF's estimating error remains below the masking curve corresponding to the IMF. The latter is computed using the psychoacoustic model used in MPEG1 [3]. In fact, when, the error's energy level is under the masking curve, the threshold needs to be modified. This reduction of the number of extrema controlled by the masking curve gives significant compression gain while preserving the listening quality.

Since the EMD can be seen as a type of wavelet decomposition [8], the threshold parameter is very soon given by a standard wavelet coefficients thresholding procedure [10]. The following expression gives for each IMF the initial value of the thresholded ($\tau_{j,0}$ or $\tau_j$) [9]:

$$\tau_j = \begin{cases} 0.05 \max |\mathrm{IMF}_j(t)|, & \text{if } \tilde{\sigma}_j = 0 \\ \tilde{\sigma}_j, & \text{else} \end{cases} \quad (2)$$

where the noise level ($\tilde{\sigma}_j$) is given by [10], [11]:

$$\tilde{\sigma}_j = \mathrm{Median}\left\{ |\mathrm{IMF}_j(t) - \mathrm{Median}\left\{\mathrm{IMF}_j(t)\right\}| \right\}. \quad (3)$$

Although there exist different non linear thresholding functions [12], in the present work, hard thresholding is used:

$$e_j = \begin{cases} \mathrm{E}_j, & \text{if } |\mathrm{E}_j| > \tau_j \\ 0, & \text{if } |\mathrm{E}_j| \leq \tau_j, \end{cases} \quad (4)$$

where $e_j$ et $E_j$ represent respectively the thresholded and the initial extremal values. To confirm this first choice of the threshold, the estimated IMF is reconstructed from non zero thresholded extrema by using spline interpolation. If the error's energy is under the masking curve of the IMF, we iterate the thresholding procedure by reducing the threshold value, as following [13].

$$\tau_{j,i} = \frac{\tau_{j,i-1}}{2}, \quad (5)$$

where $\tau_{j,i}$ is the threshold parameter of the $\mathrm{IMF}_j$ at the iteration number $i$ ($i \geq 1$). Clearly, the thresholding procedure is an iterative process: aiming at estimating an IMF from a reduced number of extrema, while ensuring the inaudibility of the estimation error. Figure 2 illustrates such aspect for a frame of the audio signal.

In this work, each extremum is coded over 8 bits and a classical scalar quantization is used for the purpose. In fact, 8 bits are enough to guarantee that the quantization error remains below the masking curve. However, better performance can be achieved by using lossless compression such as Huffman or Lempel-Ziv encoding techniques to store data. These techniques account for probability of occurrence of encoded data to reduce the number of bits allocated to. Although Lempel-Ziv is not optimum, the decoder needs not to know the encoding dictionary [16].
The different steps of the compression procedure are summarized as follows:

- Divide the original signal into frames.
- Fix the stopping criterion of the sifting process, and extract the $j^{\text{th}}$ IMF, $j \in \{1,...,C\}$, and the residual $r_C(t)$ for each IMF.
- For IMF $j$, determine all the extrema, and calculate the threshold $\tau_j$ using relations (2) and (5).
- Threshold the extrema of the IMFs as in relation (4).
- Quantize and encode non-zero thresholded extrema.
- Decoder : reconstruct the compressed frame by summoning interpolated IMFs and rebuilt the signal.

## 4. RESULTS

Three test audio signals (guitar, violin and sing) sampled at 44.1 kHz are used to illustrate the effectiveness of the proposed compression scheme. Each sample of the audio signal is coded over 9 bits. Figure 3 shows the considered signals. The results of the proposed codec is compared to those of the MP3 and the wavelet compression technique. The performance evaluation is based on objective criteria such as the Peak Signal to Noise Ratio (PSNR) and Compression Ratio (CR) [14], and subjective ones such as Subjective Difference Grade (SDG) and instantaneous Perceptual Similarity Measure (PSMt) [15].
Each signal is divided into frames of size 512 samples [3]. Using the EMD, each frame is decomposed into IMFs and residual. For illustration, figure 4 shows that the EMD decomposes a frame of the song signal into 5 IMFs and a residual. According to this decomposition, we can see that the number of extrema decreases from one IMF to the next.

Figure 5 shows the number of initial and thresholded extrema versus the number of IMFs for this frame. By applying the thresholding controlled by the psycho-acoustic model, the number of the thresholded extrema of each IMF decreases compared to the number of initial extrema. In fact, this reduction of the number of extrema to be coded

besides that 8 bits are allowed to each extremum ensures a significant gain in terms of bit rate.

For deeper investigation, we apply the proposed codec, the MP3 codec and the wavelet compression technique to the three considered signals. The obtained values of PSNR, CR, SDG and PSMt are summarized in Table 1. For wavelet compression, Daubechies wavelet of order 8 is used. This choice is motivated by its good behavior for audio encoding, compared to other wavelets [5]. A careful examination of the results reported in Table 1, shows that the proposed coder performs remarkably better than wavelet and MP3 compressions. Indeed, it offers higher values of CR with similar or slightly better audio quality. In particular for violin signal an improvement is achieved in terms of bite rate and listening quality as the CR, ODG and PSMt are higher than those offered by MP3 and Wavelet compression.

## 5. CONCLUSION

In this paper, a new approach for low bit rate perceptual audio coding is proposed. It is based on the efficient Empirical Mode Decomposition and psycho-acoustic model. The obtained results for three different audio signals are very promising. Indeed, they have proved that the present coding scheme offers lower bit rate and comparable or better listening quality than the MP3 coder and the wavelet compression technique. A large class of audio signals is necessary to confirm the obtained results. This work constitutes a first step in audio coding based on EMD. As future work, we plan to study bit allocation using psychoacoustic model and coding technique for more reduced bit rate.

## REFERENCES


[1] J.D. Johnston, "Transfom coding of audiosignals using perceptual criteria," *IEEE. Select Areas Commun.*, vol. 6, pp. 314-323, 1988.

[2] K. Brandenburg and G. Stoll, "ISO-MPEG-1 audio: A generic standard for coding of high-quality digital audio," *J. Audio Eng. Soc.*, vol. 42, no. 10, pp. 780-792, 1994.

[3] P. Noll, "MPEG digital audio coding," *IEEE Sig. Process. Magazine,* vol. 14, no. 5 pp 59-81, 1997.

[4] P. Srinivasan and L. H. Jamieson, "High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling," *IEEE Trans. Sig. Process.*, vol. 46, no. 4, 1998.

[5] P. R. Deshmukh, "Multiwavelet decomposition for audio coding," *IE(I) Journal-ET* vol. 87, pp. 38-41, 2006.

[6] N.E. Huang et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Royal Society A*, vol. 454, no. 1971, pp. 903-995, 1998.

[7] N.E. huang, "Empirical mode decomposition for analyzing acoustical signals", *US Patent*, App. 10/073,957, 2002

[8] P. Flandrin, G. Rilling and P. Gonçalvès, "Empirical mode decomposition as a filter bank," *IEEE Sig. Proc. Lett.*, vol. 11, no. 2, pp. 112-114, 2004.

[9] M. Misiti, Y. Misiti, G. Oppenheim, J. M. Poggi, "Matlab Wavelet ToolBox," *Math Works Inc.*, 1996.

[10] A.O. Boudraa and J.C. Cexus, "Denoising via empirical mode decomposition," *Proc. IEEE ISCCSP*, Marrakech, Morocco, 4 pages, 2006.

[11] K. Khaldi, A.O. Boudraa, A. Bouchiki and M. Turki-Hadj Alouane, "Speech Enhancement via EMD", *EURASIP Journal on Advances in Signal Processing*, Special issue on "The Empirical Mode Decomposition and the Hilbert-Huang Transform", vol. 2008, Article ID 873204, 8 pages, 2008.

[12] S. Mallat, *Une exploration des signaux en ondelettes*, Ellipses, Ecole Polytechnique, Palaiseau, 2000.

[13] A. Mertins, *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*, 1999.

[14] E. B. Fgee, W. J. Phillips and W. Robertson, "Comparing Audio Compression using Wavelets with other audio Compression schemes," *Proc. IEEE Canadian Conf. Electrical and Comput. Eng.*, Canada, pp.698-701, 1999.

[15] R. Huber and B. Kollmeier, "PEMO-QA New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," *IEEE Trans. Audio, Speech and Language Process.*, vol. 14, no 6, 2006.

[16] T. Welch,"A technique for high-performance data compression." *Computer*. Vol. 17, pp. 8-19, June 1984.


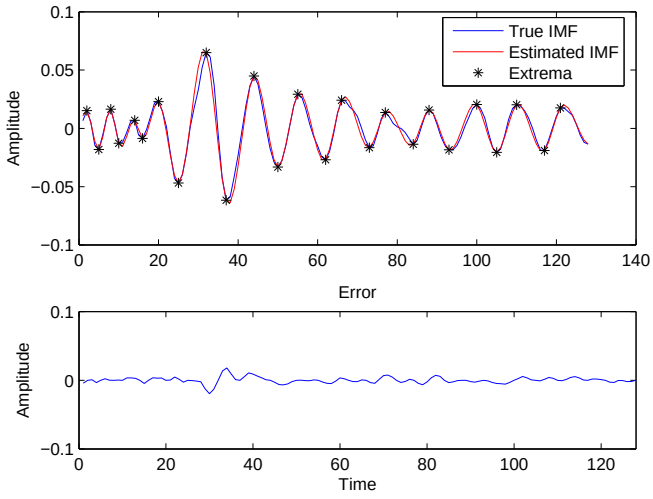


Figure 1: Original IMF and estimated version by spline interpolation.

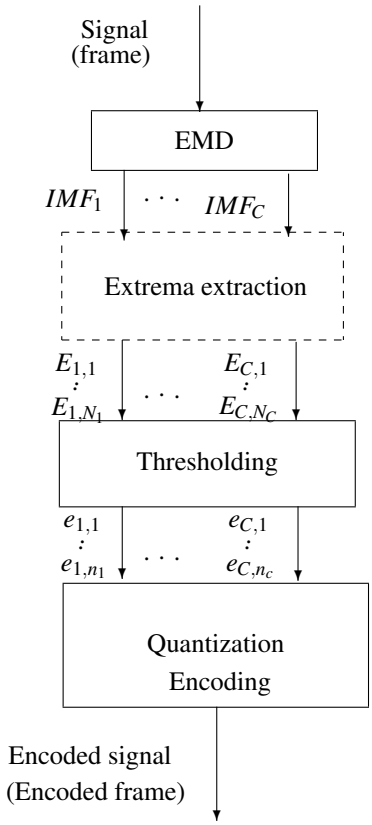


Figure 2: Encoding scheme.

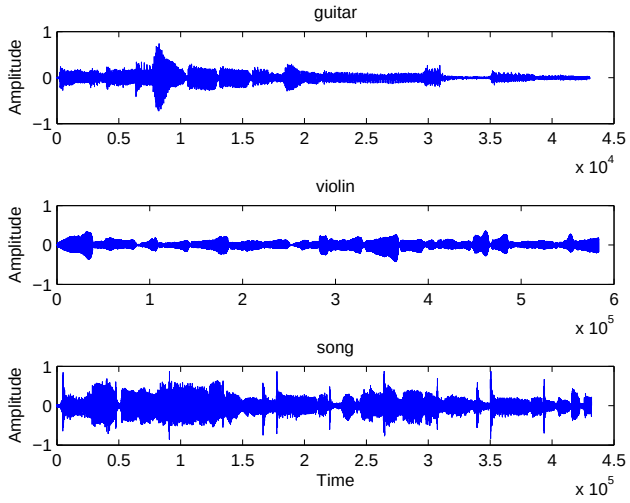


Figure 3: Original audio signals (guitar, violin and song).

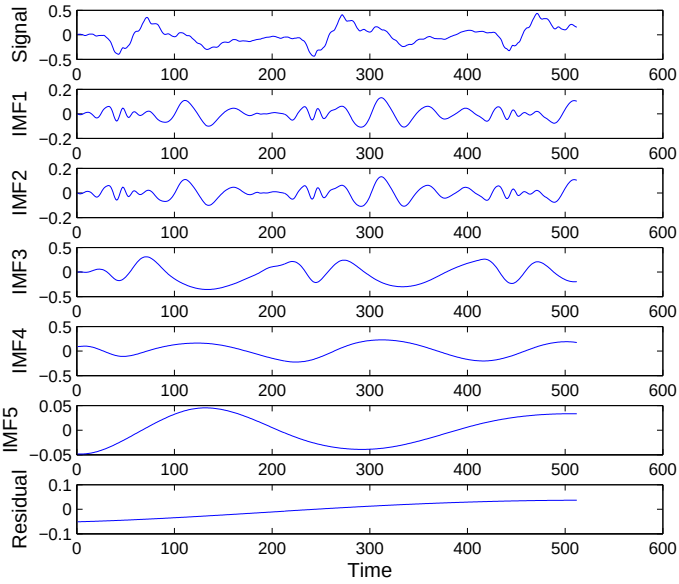


Figure 4: EMD of a frame (song).

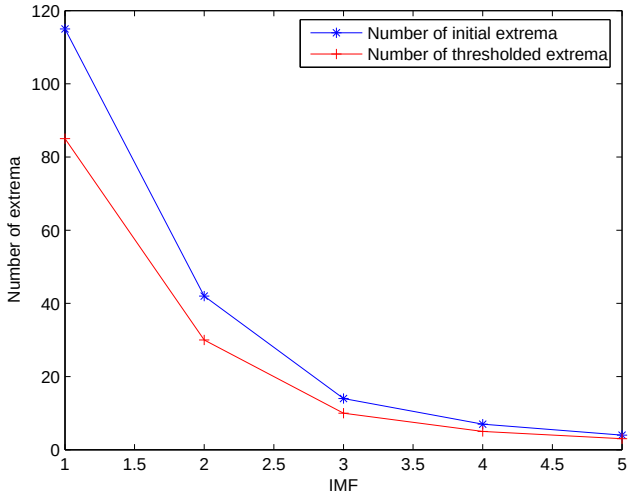


Figure 5: Variations of average number of initial and thresholded extrema values versus the number of IMFs in the frame.

Table 1: Compression results of audio signals (guitar, violin and song) by the proposed approach, MP3 and the wavelet.

| | Signal | guitar | violin | song |
|---|---|---|---|---|
| Proposed approach (EMD) | Cr | 10.12:1 | 10.50:1 | 11:1 |
| | PSNR | 50.81 | 47.16 | 57.63 |
| | SDG | -0.82 | -0.91 | -0.71 |
| | PSMt | 0.91 | 0.89 | 0.95 |
| Wavelet approach | Cr | 9.42:1 | 9.83:1 | 10.11:1 |
| | PSNR | 34.56 | 31.19 | 27.48 |
| | SDG | -1.51 | -1.76 | -1.94 |
| | PSMt | 0.85 | 0.83 | 0.81 |
| MP3 | Cr | 7.37:1 | 7.84:1 | 6.92:1 |
| | PSNR | 51.23 | 46.65 | 60.12 |
| | SDG | -0.79 | -1.05 | -0.67 |
| | PSMt | 0.92 | 0.86 | 0.96 |