AUDIO ENCODING USING HUANG AND HILBERT TRANSFORMS

K. Khaldi [1, 2, 4], *A.O. Boudraa* [2], *B. Torrésani* [3], *Th. Chonavel* [4] *and M. Turki* [1]

[1]Unité Signaux et Systèmes, ENIT, BP 37, Le Belvedre 1002 Tunis, Tunisia.
[2]IRENav (EA 3634), Ecole Navale, BCRM Brest, CC 600, 29240 Brest Cedex 9, France.
[3]Université de Provence, LATP, CMI, 39 rue F. Joliot-Curie, 13453 Marseille Cedex 13, France.
[4]Institut Télécom; Télécom Bretagne, LabSTICC UMR, BP 832, 29285 Brest Cedex, France.
(Kais.Khaldi,thierry.chonavel)@telecom-bretagne.eu, boudra@ecole-navale.fr
Bruno.Torresani@cmi.univ-mrs.fr, m.turki@enit.rnu.tn

ABSTRACT

In this paper an audio coding scheme based on the Empirical Mode Decomposition (EMD) in association with the Hilbert transform is presented. The audio signal is decomposed adaptively into intrinsic oscillatory components by EMD called Intrinsic Mode Functions (IMFs) and the associated instantaneous amplitudes and the instantaneous phases calculated. The basic principle of the proposed approach consists in encoding the instantaneous amplitudes and the instantaneous phases. The decoder recovers the original signal after IMFs reconstruction by demodulation, and their summation. The compression method is applied to different audio signals, and results compared to MP3 and to wavelet approaches.

1. INTRODUCTION

Audio signal compression of hight quality, and at low bit rate has become very important in many applications, such as digital audio broadcasting, multimedia and satellite TV, that request a lower bit rates and high fidelity. Different coding methods has been proposed for reducing the bit rate [1]-[2]. Furthermore, new methods of audio compression based on wavelet have been proposed in to reduce bit rate requirements [3]-[4]. However, a limit of the wavelet approach, is that the basis functions are fixed, and thus do not necessarily match all real signals. In this work we investigates the interest of the EMD or Huang transform for audio encoding. The EMD has been introduced by Huang et al. [5] for analyzing data from non-stationary and nonlinear processes. The major advantage of the EMD is that the basis functions are derived from the signal itself. Hence, the analysis is adaptive in contrast to the traditional methods where the basis functions are fixed. The basic idea of the proposed method is to encode the instantaneous amplitude (IA) and the instantaneous phase (IP) for each IMF, exploiting the smoothness of these instantaneous quantities. This method is applied to audio signals, and the results are compared to the wavelet and MP3 approaches.

2. HUANG TRANSFORM: EMD

The EMD breaks down any signal $x(t)$ into a series of IMFs through an iterative process called *sifting*; each one with a distinct time scale [5]. The decomposition is based on the local time scale of $x(t)$, and yields adaptive basis functions. The EMD can be seen as a type of wavelet decomposition whose subbands are built up as needful to separate the different components of $x(t)$. Each IMF replaces the signals detail, at a certain scale or frequency band. The EMD picks out the highest frequency oscillation that remains in $x(t)$. By definition, an IMF satisfies two conditions :

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

Thus, locally, each IMF contains lower frequency oscillations than the just extracted one. The EMD does not use a predetermined filter or a wavelet function, and is a fully data-driven method [5]. 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) [5]:

- 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 $m(t)$.

Signal $d(t)$ is a true IMF, if it satisfies conditions (1) and (2).

3. ANALYTIC SIGNAL

With the Hilbert transform, $\mathcal{H}[.]$, the analytic signal $z(t)$ corresponding to $s(t)$ is given by :

$$z(t) = s(t) + i\mathcal{H}[s(t)] = a(t)e^{i\theta(t)} \quad (1)$$

where the given time series $s(t)$ is the real part of (1), and the imaginary part is the Hilbert transform of $s(t)$,

$$\mathcal{H}[s(t)] = \frac{1}{\pi}\mathrm{PV}\int_{+\infty}^{-\infty}\frac{s(\tau)}{t-\tau}d\tau \tag{2}$$

PV is the Cauchy principal value of the integral. An analytic signal represents a rotation in the complex plane with the radius of rotation $a(t)$ and the IP $\theta(t)$, where

$a(t) = \sqrt{[s(t)]^2 + \mathcal{H}[s(t)]^2}$ and $\theta(t) = \tan^{-1}\left(\frac{\mathcal{H}[s(t)]}{s(t)}\right)$

The IA $a(t)$ and IP $\theta(t)$ of IMFs of audio signal are slowly varying, which is not true for general audio signals. The combination of the EMD applied to $s(t)$ to generate IMFs, and the Hilbert transform of each IMF is called the Hilbert-Huang Transform (HHT).

4. PROPOSED APPROACH

Compared to our recently published works [6],[7], where essentially extrema are encoded, in the present work IA and IP which are valuable pieces of information are exploited for coding.

4.1. Segmentation

In the proposed method, the audio signal is first segmented adaptively into frames where it remains quasi stationary within each frame. This segmentation is based on the Local Entropic Criterion (CEL) which is a non parametric detector. The CEL at instant n for a signal $x(n)$ is given by [8]:

$$\mathrm{CEL}_x(n) = \frac{\mathrm{E}_{xc}(n) - [\mathrm{E}_{xl}(n) + \mathrm{E}_{xr}(n)]}{|\mathrm{E}_{xc}(n)|} \tag{3}$$

where $\mathrm{E}_{xc}(n)$, $\mathrm{E}_{xl}(n)$ and $\mathrm{E}_{xr}(n)$ denotes the entropies of the principal window and the left and right sub-windows respectively.
$\mathrm{E}_{xc}(n) = \mathrm{E}_{x[n-\frac{N}{2},n+\frac{N}{2}-1]}$,
$\mathrm{E}_{xl}(n) = \mathrm{E}_{x[n-\frac{N}{2},n-1]}$,
$\mathrm{E}_{xr}(n) = \mathrm{E}_{x[n,n+\frac{N}{2}-1]}$.

Shannon entropy of a signal $x(n)$ in the interval $[0, N-1]$, $\mathrm{E}_{x[0,N-1]}$, is defined by :

$$\mathrm{E}_{x[0,N-1]} = -\sum_{k=0}^{N-1}|X(k)|^2\log|X(k)|^2 \tag{4}$$

where $X(k)$ is the discrete Fourier transform of $x(n)$. So the CEL has a value in the range of -1 to 1. A transient in the signal is characterized by a CEL > 0. An example of CEL variations for an audio frame guitar is shown in figure 1.

4.2. HHT

After adaptive segmentation, each audio frame $x(t)$ is decomposed into sum of IMFs by the EMD, as follows:

$$x(t) = \sum_{j=1}^{L}\mathrm{IMF}_j(t) + r_L(t) \tag{5}$$

where $\mathrm{IMF}_j(t)$ is the j^{th} IMF and $r_L(t)$ is the residual. The L value is determined automatically using standard deviation SD as stopping criterion which usually is set between 0.2 and 0.3 [5]. An example of decomposition of an audio frame is illustrated in figure 2. For each IMF, IP $\theta(n)$ and the IA $a(n)$, are determined using Hilbert transform. Figure 3 shows the IA and IP of an IMF.

4.3. Encoding

For class of audio signals studied, it is found that IA of IMFs are correlated. An example of such correlations is shown in figure 4. So, AR model is used to efficiently exploit this temporally correlated information.

$$\hat{a}(n) = \sum_{k=1}^{p}c(k)a(n-k) + \epsilon(n) \tag{6}$$

where $[1, c(2), ..., c(p)]$ are the coefficients of the model and $\epsilon(n)$ is stationary zero mean input sequence that is independent of past outputs. Analysis of variation of IMFs IP show that for coding the classical scalar quantization can be used. Thus, only extrema of IP are encoded. Figure 5 shows the extrema (red circles) of an IMF that are coded. This information corresponds to encoding zero crossings of the imaginary parts of IMFs. Finally, the encoded coefficients of the IA and extrema of the IP is improved, 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 does not require the encoding dictionary [9].

4.4. Decoding

The decoder requires only the encoded extrema of IP and calculates the remaining phase values by linear interpolation. IA is also generated by linear prediction. Finally, the estimated IMF is calculated as follows:

$$\hat{\mathrm{IMF}}(n) = |\hat{a}(n)|\cos(\hat{\theta}(n)) \tag{7}$$

The audio frame is constructed by IMFs summation and the decoded audio signal is obtained by frames concatenation.

5. RESULTS

The method is tested on different audio signals, sampled at 44.1 Khz. The results are compared to the MP3 et wavelet

approaches. As criteria to evaluate the performance of the method, Signal to Noise Ratio (SNR) and Compression Ratio (CR), Subjective Difference Grade (SDG) and instantaneous Perceptual Similarity Measure (PSMt) are used [10]. Due to its good behavior for audio encoding, compared to other wavelets, the Daubechies wavelet of order 8 is used [4]. Table 1 shows the variation of TC and SDG against the number of AR level. So, it is clear that order 9 represents a good compromise between the TC and listening quality (SDG).

Table 1. Variations of the TC and the SDG over the AR order.

	guitar		violin		sing	
order	TC	SDG	TC	SDG	TC	SDG
5	13.40:1	-2.17	13.35:1	-2.87	16.43:1	-2.32
7	11.39:1	-1.08	11.72:1	-1.91	13.20:1	-1.12
9	10.15:1	-0.85	9.96:1	-1.09	11.30:1	-0.75
11	8.94:1	-0.83	8.70:1	-1.05	9.48:1	-0.73
13	8.14:1	-0.71	7.74:1	-1.01	8.31:1	-0.67
15	7.51:1	-0.63	7.01:1	-0.92	7.39:1	-0.51

Table 2, shows that the improvement in TC provided by the proposed method varies from 9.96:1 to 11.3:1 than the TC achieved by wavelets and MP3. Even for a sing signal, we still can observe the effectiveness of the proposed method in compression. A careful examination of the results reported in Table 2, shows that the proposed approach performs remarkably better than wavelet and MP3 methods. Furthermore, when listening the decoded signal, the proposed method produces lower noise compared to the wavelet method and MP3. This result is shown in table 2, when we see the acquired SDG values depending to TC is better than the other methods. The obtained results show the interest to encode both IA and IP.

Table 2. Compression results of audio signals (guitar, violin and sing) by the proposed approach, MP3 and the wavelet.

	Signal	guitar	violin	sing
EMD	Cr	10.15:1	9.96:1	11.3:1
	SNR	20.27	20.41	22.86
	SDG	-0.85	-1.09	-0.75
	PSMt	0.89	0.84	0.91
Wavelet	Cr	9.42:1	9.83:1	10.11:1
	SNR	20.17	19.65	23.43
	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
	SNR	21.84	19.72	23.69
	SDG	-0.79	-1.05	-0.67
	PSMt	0.92	0.86	0.96

6. CONCLUSION

In this paper, a new coding method combining Huang and Hilbert transforms is presented. The estimated IP and IA of the extracted IMFs are used for audio signals coding. Obtained results for different audio signals show that the proposed method, performs better than the wavelet and MP3 approaches, and confirm our previous findings [6],[7]. These results also show the interest of the EMD as basis for signals coding. To confirm the obtained results and the effectiveness of the EMD-compression approach, the scheme must be evaluated with a large class of audio signals and in different experimental conditions, such as sampling rates, sample sizes.

7. REFERENCES

[1] J.D. Johnston, "Transform coding of audio signals using perceptual criteria," *IEEE. Select Areas Commun.*, vol. 6, pp. 314–323, 1988.

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

[3] 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.

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

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

[6] K. Khaldi, A.O. Boudraa, M. Turki, T. Chonavel and I. Samaali, "Audio encoding based on the empirical mode decomposition," in *EUSIPCO*, Glasgow, 2009.

[7] K. Khaldi, A.O. Boudraa, M. Turki and T. Chonavel, "Codage audio perceptuel à bas débit par décomposition en modes empiriques," in *GRETSI*, Dijon, France, 2009.

[8] G. Gonon, S. Montrésor and M. Baudry, "Segmentation multibande adaptée basée sur le critre entropique local pour le codage audio," in *GRETSI*, Toulouse, 2001.

[9] T. Welch, "A technique for high-performance data compression," *Computer*, vol. 17, pp. 8–19, 1984.

[10] R. Huber and B. Kollmeier, "PEMO-Q a 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.

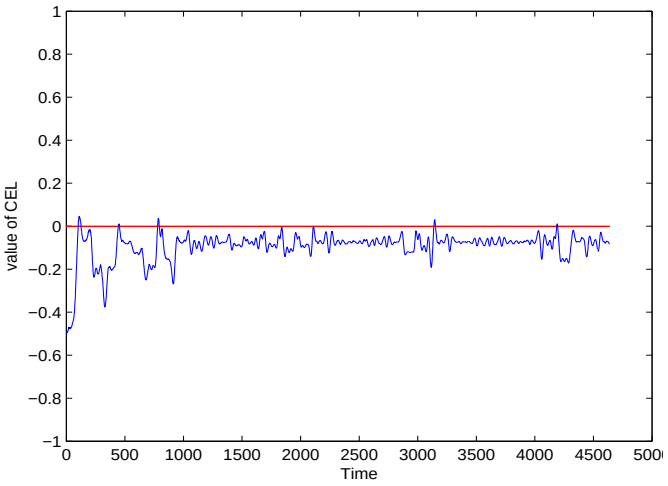

Fig. 1. CEL variation of the audio frame guitar.

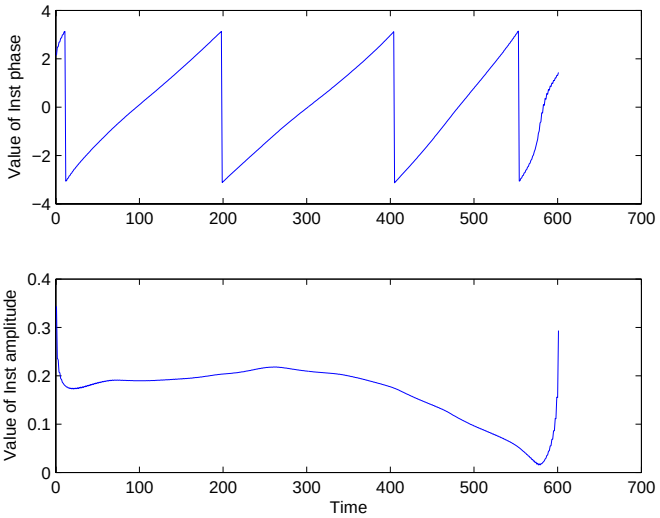

Fig. 3. Instantaneous phase and instantaneous amplitude of IMF3.

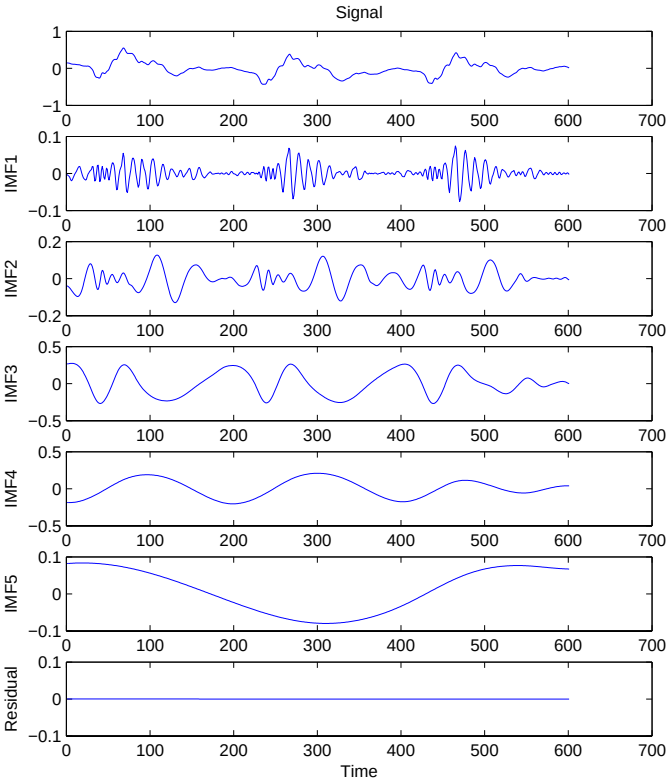

Fig. 2. Decomposition of an audio frame by EMD.

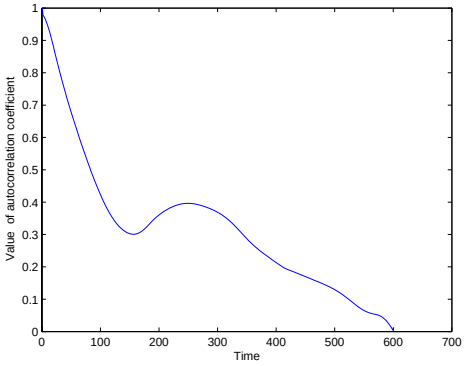

Fig. 4. Autocorrelation function of instantaneous amplitude of IMF3.

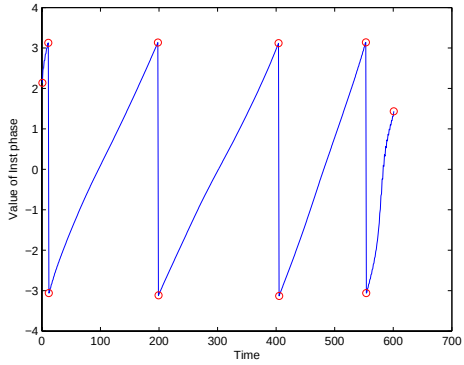

Fig. 5. Instantaneous phase and their extrema of the IMF3.