Global Journal of Computer Science and Technology
Graphics & Vision
Volume 13 Issue 3 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Data Adaptive Multi-Band Approach for Robust Audio Watermarking using Adaptive Threshold

By K.M. Ibrahim Khalilullah, A K M Wasif Abedin, Mahfuza Khatun, Mohammad Tareq & Md. A.R. Pramanik

Dhaka International University, Dhaka, Bangladesh

Abstract - This paper presents a data adaptive digital audio watermarking process with the use of Empirical Mode Decomposition (EMD) and Hilbert Transform (HT). The host audio signal is decomposed into several Intrinsic Mode Functions (IMFs). A binary image or a set of -1 and 1, which is obtained by mapping from standard normal distributed pseudo random numbers, is embedded as secret information into the significant coefficients of the highest energetic IMF that are greater than a specified adaptive threshold. Thus watermarked IMF is less sensitive to common signal processing attack. As a result this method increases more robustness. The experimental results show that the proposed method has good imperceptibility and robustness under common signal processing attacks such as additive noise (in time domain and in frequency domain), low pass filtering, re-sampling, re-quantization, MP3 compression, and sound processing effects such as, delay, a natural sounding reverberation (Schroeder's Reverberator), flanging effect, equalization effect, etc. A Bit Error calculation equation is provided to measure the accuracy of the extracted watermark.

Keywords : audio watermarking, empirical mode decomposition, hilbert transform, wavelet transform, adaptive threshold, binary image.

GJCST-F Classification: I.4.5

DATA ADAPTIVE MULTI-BAND APPROACH FOR ROBUST AUDIO WATERMARKING USING ADAPTIVE THRESHOLD

Strictly as per the compliance and regulations of:

Data Adaptive Multi-Band Approach for Robust Audio Watermarking using Adaptive Threshold

K.M. Ibrahim Khalilullah [α], A K M Wasif Abedin [σ], Mahfuza Khatun [ρ], Mohammad Tareq [Ѡ] & Md. A.R. Pramanik [¥]

Abstract - This paper presents a data adaptive digital audio watermarking process with the use of Empirical Mode Decomposition (EMD) and Hilbert Transform (HT). The host audio signal is decomposed into several Intrinsic Mode Functions (IMFs). A binary image or a set of -1 and 1, which is obtained by mapping from standard normal distributed pseudo random numbers, is embedded as secret information into the significant coefficients of the highest energetic IMF that are greater than a specified adaptive threshold. Thus watermarked IMF is less sensitive to common signal processing attack. As a result this method increases more robustness. The experimental results show that the proposed method has good imperceptibility and robustness under common signal processing attacks such as additive noise (in time domain and in frequency domain), low pass filtering, re-sampling, re-quantization, MP3 compression, and sound processing effects such as, delay, a natural sounding reverberation (Schroeder's Reverberator), flanging effect, equalization effect, etc. A Bit Error calculation equation is provided to measure the accuracy of the extracted watermark.

Keywords : audio watermarking, empirical mode decomposition, hilbert transform, wavelet transform, adaptive threshold, binary image.

I. Introduction

In any given application, practitioners had already recognized two different technologies for multimedia data protection: encryption and digital watermarking [1, 2]. However, a digital watermark is complementary approach than cryptographic processes. The availability of digital multimedia contents has been grown rapidly in the recent years. Today, digital media documents can be distributed via the World Wide Web to a large number of people without much effort and money. Unlike traditional analog copying, with which the quality of the duplicated content is degraded, digital tools can easily produce large amount of perfect copies of digital documents in a short period.

Author α ¥ : Lecturer, Dept. of Computer Science & Engineering, Dhaka International University, Dhaka, Bangladesh.
E-mails : ibrahim_5403@yahoo.com, pramanikanis@gmail.com
Author σ : B.Sc Engineering from Dhaka International University major in Computer Science & Engineering. E-mail : abedinwasif@gmail.com
Author ρ Ѡ : Lecturer, Dept. of Electrical, Electronics & Telecommunication Engineering, Dhaka International University, Dhaka, Bangladesh. E-mails : mahfuza.khatun@yahoo.com, tareq.mhd@gmail.com

This ease of digital multimedia distribution over the Internet, together with the possibility of unlimited duplication of this data, threatens the intellectual property (IP) rights more than ever. Therefore, there is a strong need for security services in order to keep ownership for the document owner and reliable to the customer. Watermarking plays an important role for that purposes. In a digital watermarking process, typically, a pattern or sequence of bits is embedded into a digital image, an audio or video file such that the embedded pattern can be detected or extracted later to make an assertion about the object. Some proposed or actual watermarking applications are [3]: broadcast monitoring, owner identification, proof of ownership, transaction tracking, content authentication and tampering detection, copy control, and device control, fingerprinting, information carrier.

Embedding secret information into audio sequences is a more tedious task than image watermarking, due to dynamic preeminence of the human auditory system (HAS) over human visual system (HVS). The main confront in digital audio watermarking and steganography is that if the perceptual transparency parameter is predetermined, the design of a watermark system cannot obtain high robustness and a high watermark data rate at the same time. We have addressed this challenging situation with the use of EMD specially developed for analyzing non-linear and non-stationary signals [4].

In the wavelet transform domain [5, 6], the watermark is embedded in high frequency band or in low frequency band; therefore watermark can't be embedded in the whole transform domain. In EMD-based method, we can add the watermark into the whole transform domain. The EMD method provides perfect time-frequency localization properties due to its instantaneous frequency than discrete wavelet transform (DWT) domain and leads to implicit audio-visual masking [7]. This decomposition method is adaptive, and highly efficient for perceptual transparency (inaudibility) and robustness. That is HAS is perfectly met too. In the previous EMD-based method [8], the watermark is added into the whole coefficients of the highest energetic IMF, therefore the watermarked IMF is sensitive to common signal processing attacks than this proposed method. This paper presents a method, which

Global Journal of Computer Science and Technology (F) Volume XIII Issue III Version I Year 2013

Global Journal of Computer Science and Technology (F) Volume XIII Issue III Version I Year 2013

can embed the watermark into the filtered coefficients of the highest energetic IMF using adaptive threshold. Since adaptive threshold is used to add watermark into the significant IMF, our method is much more resistant to common signal processing manipulation when compared with other methods.

II. Proposed Audio Watermarking Technique

The Empirical Mode Decomposition is a new method for analyzing nonlinear and non-stationary data. This method is adaptive, and, therefore, highly efficient. Since this decomposition method is based on the local characteristic time scale of the data, it is applicable to nonlinear and non-stationary processes. Using this method any complicated data set can be decomposed into a finite and often small number of 'Intrinsic Mode Functions (IMF)'. An Intrinsic mode Function (IMF) is a function that satisfies two conditions [9]: (i) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. With the Hilbert transform, the 'Intrinsic Mode Functions' yield instantaneous frequencies as functions of time that provide sharp identifications of imbedded structures.

We have extended the previous idea by using EMD and embedding watermark into the filtered coefficients of the highest energetic IMF for increasing robustness against different signal processing attacks. In the proposed method, at first we have divided the host signal into a number of frames. Then each frame is decomposed into a small number of intrinsic mode functions (IMFs) that acknowledge well-behaved Hilbert transforms. This method is applicable to nonlinear and non-stationary process because of the local characteristic time scale of the data. However, after calculating energy of the every IMF, we have picked an IMF contained highest energy. The following Fig. 1 shows the energy distribution of the IMFs for the first frame of an audio signal:

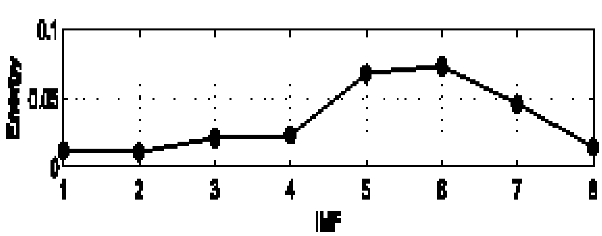


Figure 1 : Energy distribution for IMFs of the first frame

From the energy distribution of IMFs, we can select the 6th IMF because of the highest energetic value. A scaled version of the pseudo random watermark sequence $w[n]$, n is the size of the watermark which is equal to the number of frame, added to the filtered coefficients of the selected IMFs. Only one bit of $w[n]$ is added to the only one frame. The i[th] bit of $w[n]$ is added to the selected IMF (e.g. 6th IMF) of the i[th] frame by using the following additive embedding formula (1):

$$d'_m(l) = d^i_m(l) + \alpha w(i) \tag{1}$$

Where, i=1 for the first frame, i=2 for the second frame and so on. l runs over all coefficients of the IMF > T (Adaptive threshold), $d^i_m(l)$ is the filtered coefficient of m[th] IMF of i[th] frame, $w(i)$ is the i[th] bit of the watermark, and $d'_m(l)$ is the watermarked coefficient of the selected IMF. The watermarked frame is generated by simply summing all IMF of that frame. The watermarked signal $x'[t]$ is generated after embedding watermarks to all frames in this way. The scaling factor α is chosen to be as high as possible while remaining inaudible. Here α is taken as 0.02. The threshold value (T) is calculated by this equation (2):

$$T = (Max\_Val + Min\_Val)/2 \tag{2}$$

Where Max_Val is the maximum coefficient value and Min_Val is the minimum coefficient value of the selected IMF.

Since, EMD is empirical, there is no reason the decomposition will result in same (or even similar) IMFs when the signal is modified (or even the watermark added). If it does not result in the same IMFs, the entire scheme fails. Therefore, in the detection process, the original signal $x[t]$ and watermarked signal $x'[t]$ are transformed into Hilbert transform domain by excluding the use EMD. Then the watermark is extracted simply by using the following subtraction equation (3):

$$w'_i[t] = H'_i[x'[t]] - H_i[x[t]]/\alpha \tag{3}$$

Where, $H'_i[x'[t]]$ is the watermarked audio signal in Hilbert transform domain, $H_i[x[t]]$ is the original audio signal of i[th] frame in Hilbert transform domain, and $w'_i[t]$ is the extracted watermark of the i[th] frame. The resulted watermark bit is calculated by averaging over the number of embedded watermark to the frame. In this calculation, two different techniques are applied to two different watermarks.

For the sequence of 1 and -1, the watermark bit is a +1 If the average is greater than zero and if the average is less than zero the corresponding watermark bit is a –1. If we use binary number as a watermark , then the watermark bit is a 1 if the average is greater than zero; if the average is less than equal zero the corresponding watermark bit is a zero. Finally we get the resulted watermark $w''[n]$ after calculating the above operation for all frames.

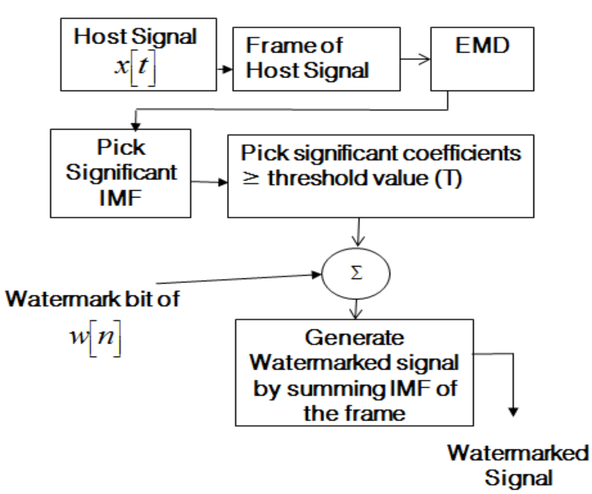
The similarity between the original watermark and resulted watermark is calculated to detect the existence of watermark. The similarity should be increased or decreased according to the amount of the inserted watermark, as well as to the degree of the attack. To calculate the detector response value (S_{dr}), dr means detector response, the proposed method utilizes the vector projection as defined in the following equation (4):

$$S_{dr} = w[n] \cdot w''[n] / \sqrt{w''[n] \cdot w''[n]} \qquad (4)$$

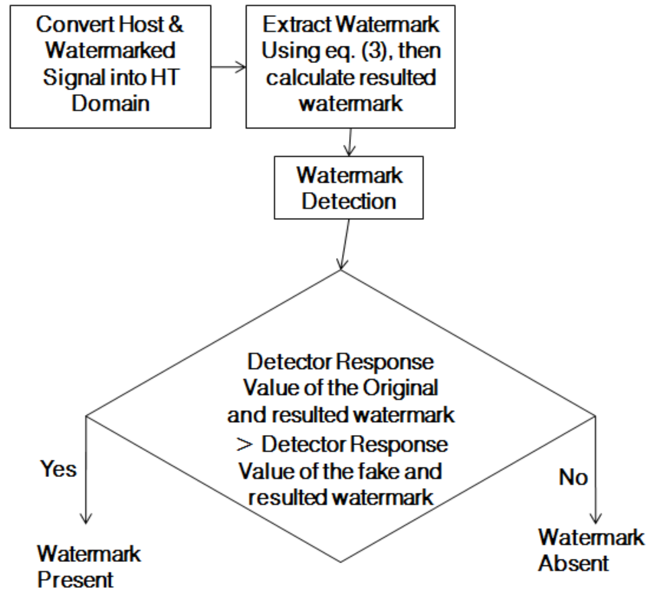
Where, $w[n]$ is the original watermark or fake watermark, and $w''[n]$ is the resulted watermark. We also calculated the normalized correlation (ψ) to measure the performance of the embedding algorithm using the following equation (5):

$$\psi = \left[\sum_j w(j) w''(j) \Big/ \sum_j w(j) w(j) \right] \qquad (5)$$

When the detector response value of the original watermark and extracted watermark is greater than other detector response values of the fake watermarks and resulted watermark, then we can conclude that the watermark is present. Fig. 2 shows a block diagram of the proposed method for a single frame.



(a)

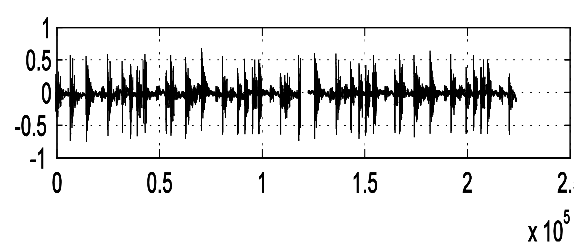


(b)

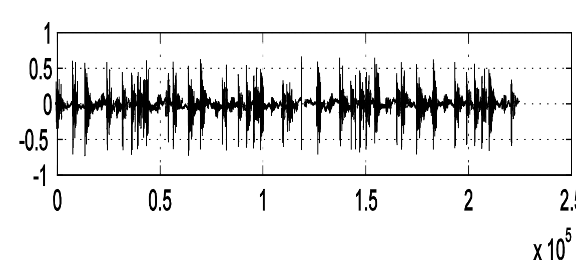
Figure 2 : The proposed method: (a) Watermark embedding and (b) Watermark detection

III. Experimental Results and Discussion

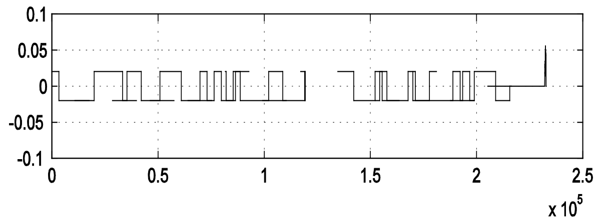
To test the algorithm we have used eight audio music clips (44 kHz, 705 kpbs), which has been collected from Fruity Loops software. An original audio signal (drum music), watermarked audio signal, and the extracted watermark signal are shown in Fig. 3. After converting the original and watermarked audio signal into HT domain, the extracted watermark is calculated using equation (3).



(a)



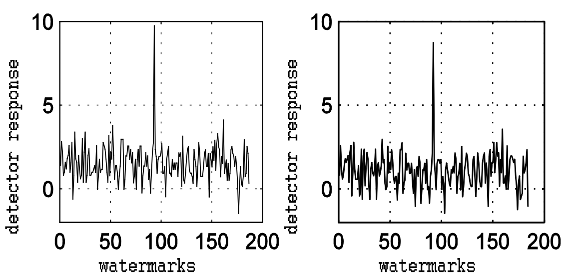
(b)



(c)

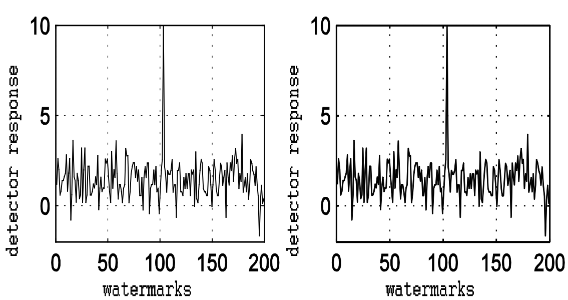
Figure 3 : (a) Original audio signal, (b) Watermarked audio signal, (c) Extracted Watermark

The detector response values (S_{dr}) on the y-axis and randomly generated watermark on the x-axis against three attacks as example for the piano music and drum music are shown in the following figures (from Fig. 4 to Fig. 6). The original audio signal is watermarked with a seed of 100. Therefore, in each figure, the detector response values with the original watermark are located at 100 on the x-axis. Therefore, at this position the detector response values are greater than any other values.



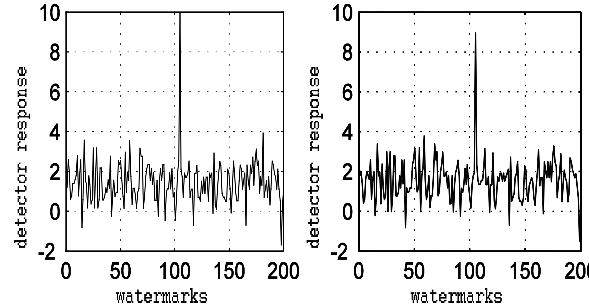
(a) (b)

Figure 4 : Detector responses for Gaussian noise attack. (a) Piano music with = 0.96, SNR= -7.26 db and (b) Drum music with =0.9, SNR= -15.104db



(a) (b)

Figure 5 : Detector responses for Flanging effect (a) Piano music with =1, SNR= 6.02db and (b) Drum music with =1, SNR= 6.05db



(a) (b)

Figure 6 : Detector responses for MP3 compression (a) Piano music with ψ =1 and (b) Drum music with ψ =0.9, 56 kbps

The performance against Gaussian noise addition is calculated to illustrate the robustness of the proposed algorithm as:

$$R_p = V_W - V_F \tag{6}$$

Where, V_W is the similarity measurement value between original watermark and resulted watermark extracted from the attacked watermarked signal and V_F is the highest similarity measurement value between false watermark and resulted watermark extracted from the attacked watermarked signal. The changes of the robustness performance for drum music are shown in Fig. 7:

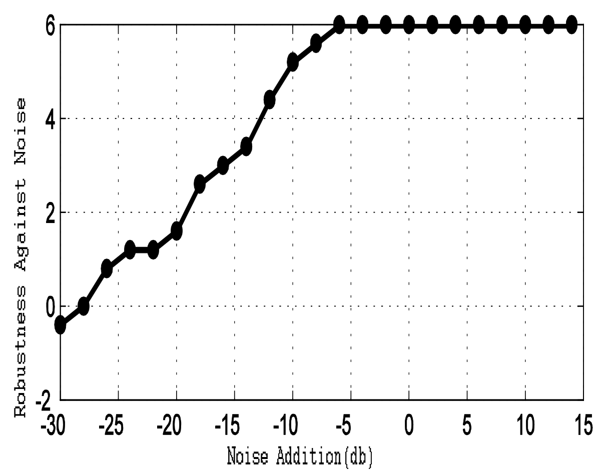


Figure 7 : Robustness performance values against different noise levels (db)

If $R_p > 0$, then watermark is detected correctly. The highest value of R_p represents maximum performance. The proposed method works for -26db also. We have also calculated Bit Error (β) to examine extracted watermark by the following equation (7):

$$\beta = \frac{\sum_j \left| w(j) - w''(j) \right|}{2L} \qquad (7)$$

Where, L is the length of the watermark i.e. the number of bit of the watermark. The minimum value of β is 0 and maximum value of β is 1. Experimental results of the eight music clips are shown in Table-I.

Table 1 : Experimental results of different attacks for various types of audio signals

Music Clip → / Attacks ↓		Drum	C_HC	Dream-zofluxury	NewStuff	ISuck-AtTitles	404lament	H441-GOA	Average β
Gaussian noise	ψ =0.96 SNR=-7.26 β =0.020	ψ =0. 90 SNR=-15.10 β =0.050	ψ =0.980 SNR=-20.10 β =0.010	ψ =0.94 SNR=- 4.23 β =0.030	ψ =0.96 SNR=- 5.075 β =0.020	ψ =0.94 SNR=-7.36 β =0.030	ψ =0.94 SNR=- 23.60 β =0.030	ψ =0.94 SNR=-17.25 β =0.030	0.0275
Add FFT Noise	ψ =1 SNR=30.5 β =0	ψ =1 SNR= 28.5 β =0	ψ =1 SNR=30.2 β =0	ψ =1 SNR=35. 5 β =0	ψ =1 SNR=40. 5 β =0	ψ =1 SNR=64.21 β =0	ψ =1 SNR= 48.14 β =0	ψ =1 SNR=54.46 β =0	0.0
Down Sampling	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	ψ =1 22 kHz β =0	0.0
Requan Tization	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	ψ =1 352 kbps β =0	0.0
Low pass filter	ψ =1 SNR=28.68 β =0	ψ =1 SNR= 16.68 β =0	ψ =1 SNR=17.62 β =0	ψ =1 SNR=31.88 β =0	ψ =1 SNR=20.81 β =0	ψ =1 SNR=22.19 β =0	ψ =1 SNR=24.83 β =0	ψ =1 SNR=30.39 β =0	0.0
Delay Effect	ψ =1 SNR=6.04 β =0	ψ =1 SNR= 6.104 β =0	ψ =1 SNR=6.0309 β =0	ψ =0.86 SNR=6.116 β =0.070	ψ =0.94 SNR=6.094 β =0.030	ψ =0.92 SNR=6.16 β =0.040	ψ =1 SNR=6.12 β =0	ψ =1 SNR=6.12 β =0	0.0175
Flanging Effect	ψ =1 SNR= 6.02 β =0	ψ =1 SNR= 6.05 β =0	ψ =1 SNR=6.01 67 β =0	ψ =0.98 SNR=6.185 β =0.01	ψ =0.96 SNR=6.051 β =0.020	ψ =0.98 SNR=6.096 β =0.010	ψ =1 SNR=6.061 β =0	ψ =1 SNR=6.05 β =0	0.005
AllPass Reverberation	ψ =1 SNR=5.77 β =0	ψ =1 SNR=3.88 β =0	ψ =0.96 SNR=4.0 β =0.020	ψ =0.90 SNR=5.495 β =0.050	ψ =1 SNR=4.343 β =0	ψ =0.94 SNR=3.905 β =0.030	ψ =1 SNR=6.76 β =0	ψ =1 SNR=2.80 β =0	0.0125

The average bit errors of the eight music clips for different Gaussian noise addition are sown in Fig. 8:

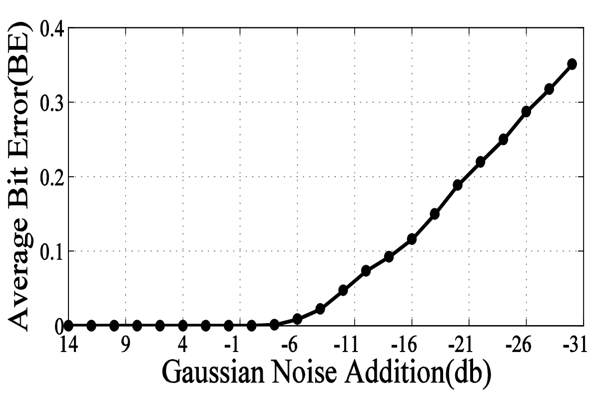


Figure 8 : Average bit-error against different level of Gaussian noise (db)

IV. Conclusion

This paper proposed an audio watermarking schema using the principle advantages of Empirical Mode Decomposition. A multiband approach is presented here for digital audio watermarking. The data adaptive time varying filtering is implemented for sub band decomposition of the host audio signal using EMD. Traditional frequency domain filtering techniques (i.e. Fourier and wavelet) depend on the priori basis functions and hence the signal is fitted with those bases. In the proposed method, the decomposition is fully data dependent and without the use of any basis function. The EMD based decomposition is a complete decomposition i.e. the original signal can easily be obtained by simply summing up the individual IMFs and hence the method performs better for audio watermarking. The efficiency of the proposed algorithm has been shown by a series of experiment. In this method we can embed a binary logo image as a watermark also.

References Références Referencias

1. M. Eskicioglu & E. J. Delp (April 2001). Overview of Multimedia Content Protection in Consumer Electronics Devices. *Signal processing: Image Communication,* 16(7), pp: 681-699.
2. A. M. Eskicioglu & E.J. Delp (April 2003). Security of Digital Entertainment Content From Creation to Consumption. *Signal Processing : Image Communication, Special Issue on Image Security,* 118 (4), pp. 237-262.
3. I. J. Cox, M. L. Miller, and J. A. Bloom (2002). Digital Watermarking*. Morgan Kaufmann Publishers*.
4. Huang, N. E. et. al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London A,* Vol. 454, pp. 903-995.
5. Xianghong Tang, Yamei Niu, Hengli Yue, Zhongke Yin (2005). A Digital Audio Watermark Embedding Algorithm. *International Journal of Information Technology, China,* Vol. 11 No.12.
6. Martinez-Noriega, R. et. al. (July 8-11, 2007). Audio Watermarking Based on Wavelet Transform and Quantization Index Modulation. *The 22nd International Technical Conference on Circuits/-Systems, Computers and Communications,* ITC-CSCC 2007, pp. 133-134, Busan, Korea.
7. LingFei Liang, ZiLiang Ping (2008). Information Hiding Based on Empirical Mode Decomposition. *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application,* paciia, vol. 1, pp. 561-565.
8. A. N. K. Zaman, K. M. I. Khalilullah, M. W. Islam and M. K. I. Molla (May, 2010). A robust digital audio watermarking algorithm using empirical mode decomposition. *Proc. of 23rd Canadian Conf. on Electrical and Computer Engineering, (CCECE'10).*
9. Foo S, Yeo T & Huang D. (2001). An adaptive audio watermarking system. *In: Proc. IEEE Region 10 International Conference on Electrical and Electronic Technology*, Phuket Island-Langkawi Island, Singapore, p 509–513.