International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 05, May-2015

Audio Watermarking using Doubly Iterative Empirical Mode Decomposition

Veena Junjarwad
Department of Telecommunication Engineering
Siddaganga Institute of Technology,
Tumakuru-572103, Karnataka, India

Chandrashekar H M
Department of Telecommunication Engineering
Siddaganga Institute of Technology,
Tumakuru-572103, Karnataka, India

***Abstract*—This paper presents an adaptive audio watermarking algorithm based on Doubly Iterative Empirical Mode Decomposition. Initially audio signal is divided into number of frames, each of same length. Then each frame is decomposed adaptively into zero mean intrinsic oscillatory components called as Intrinsic Mode Functions (IMFs) by Doubly Iterative Empirical Mode Decomposition. To achieve good performance against various attacks, watermark bits along with the Synchronization Code (SC) bits are embedded into extrema of the last IMF obtained from each frame, using Quantization Index Modulation (QIM) by which good rate-distortion-robustness performance can be achieved. Simulation results shows that the audio watermarking scheme is robust against attacks like additive noise, resampling, filtering, cropping, requantization, echo-addition and MP3 compression.**

***Keywords*—Doubly Iterative Empirical Mode Decomposition; Intrinsic Mode Function; Synchronization Code; Audio Watermarking; Quantization Index Modulation.**

I. INTRODUCTION

Copyright protection of digital media is done by embedding a watermark in original audio signal, called as host audio signal. According to International Federation of the Phonographic Industry (IFPI), the main requirements of audio watermarking are imperceptibility, robustness and data payload. Imperceptibility refers to inaudibility of watermark within the host audio signal. The quality of the audio should not be degraded after the addition of watermark. Imperceptibility is evaluated using both subjective listening and objective measures. According to IFPI recommendation, watermarked audio signal should maintain Signal to Noise Ratio (SNR) of 20dB. Robustness corresponds to ability to extract watermark bits from the watermarked audio signal subjected to various attacks. Data payload refers to amount of data that can be embedded into host audio signal per unit time, without degrading its quality and watermarking algorithm should offer data payload of more than 20 bps. The applications of watermarking are copyright control, identifying the owner, tracking transaction, copy control, content authentication and broadcast monitoring. Different watermark techniques have been proposed and are referred in [1]-[5]. The spread spectrum audio watermarking scheme referred in [5] is robust against various attacks but the limitation is lower transmission bit rate. The bit rate can be improved using watermarking schemes in wavelet domain. In [3], watermark bits are embedded into low frequency co-efficient in Discrete Wavelet Domain (DWT) and time-frequency localization characteristics are exploited. In [4], watermarking is done in DWT domain based on Singular Value Decomposition (SVD). A limitation of performing watermarking in wavelet domain is fixed basis functions which may not necessarily match all real signals. To overcome the drawback, a new signal decomposition method which expresses the audio signal as an expansion of basis functions is introduced. These basis functions are signal dependent and are estimated using iterative procedure called as sifting. As priori choice of basis functions is not involved, the technique is advantageous. It works by breaking audio frame down into a number of zero mean signals, termed Intrinsic Mode Functions (IMFs). Each IMF represents an embedded characteristic oscillation on a separated time scale. Any time-varying audio signal x(t) can be expanded by EMD as

$$x(t) = \sum_{j=1}^{n} \mathrm{IMF}_j(t) + r_n(t) \qquad (1)$$

where n represents number of IMFs and $r_n(t)$ is final residue. The IMFs are zero mean functions and are orthogonal to each other. The number of extrema and zero crossing must be either equal or differ at most by one, in an IMF [6].

In standard version of EMD, local extrema points are used as the interpolation points and natural cubic splines are used for interpolation operation during sifting procedure. The performance of EMD can be enhanced by the improved criteria for interpolation point selection. Instead of local extrema of signal, the extrema points of subsignal which is having higher instantaneous frequency are used. As extrema points are not known in advance, interpolation point selection criterion is exploited in doubly iterative sifting schemes, which leads to improved decomposition performance. Thus a novel EMD variant termed as Doubly Iterative Empirical Mode Decomposition is considered [7]. The decomposition is completed with finite number of IMFs and frequency goes on decreasing from one IMF to the next. Higher order IMFs are of low frequency, signal dominated functions and are more vulnerable to attacks. Thus last IMF is considered to embed watermark bits. Quantization Index Modulation (QIM) is used to embed the watermark bits because of its blind nature and good robustness [8]. Parameters of QIM are chosen such that inaudibility constraint is maintained. Experimental results show that the scheme is robust against attacks like addition of white Gaussian noise, resampling, requantization, filtering, cropping, echo addition and MP3compression.

II. PROPOSED WATERMARKING ALGORITHM

The host audio signal is segmented into number of frames each consists of 128 samples. Doubly Iterative Empirical Mode Decomposition is applied on each frame and corresponding IMFs are extracted as shown in figure 1.

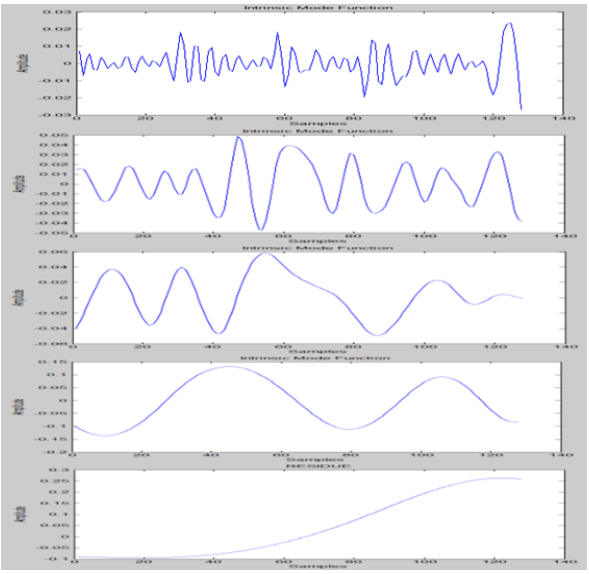


Figure 1: Decomposition of an audio frame by Doubly Iterative Empirical Mode Decomposition

Combination of Synchronization Code and watermark bits is as shown in figure 2.

Synchronization code 1111100110101110	Watermark bits	Synchronization code 1111100110101110

Figure 2: Data structure m_i

The combination is then embedded into extrema of last IMFs of consecutive frames, a bit (either 0 or 1) per extrema. The number of bits embedded in extrema of last IMF of one frame to the following is not constant, as number of IMFs and number of extrema depend on amount of data of each frame. As the number of extrema per last IMF is small compared to length of watermark image to be embedded, all the bits cannot be embedded in last IMF of one frame. Thus last IMFs of consecutive frames are considered. The length of binary sequence to be embedded is $2N_1+N_2$ where N_1 is length of Barker code i.e. equal to 16 and N_2 is length of watermark bits. This binary sequence is length $2N_1+N_2$ is embedded in host audio signal P times. The value of P depends on the length of host audio signal. Then, the superposition of all the IMFs along with residue results in audio frame, which is inverse transformation of Doubly Iterative Empirical Mode Decomposition. Finally, all the frames are concatenated as shown in figure 3.

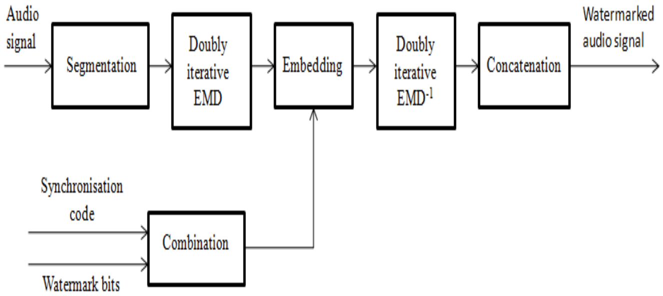


Figure 3: Watermark embedding

For watermark extraction, the watermarked audio signal is divided into number of frames each containing 128 samples. To each frame Doubly Iterative Empirical Mode Decomposition is applied to obtain IMFs. The number of IMFs obtained for original audio frame and watermarked audio frame remains same, as the sifting of the watermarked signal extracts same number of mode functions as before watermarking, as shown in figure 4.

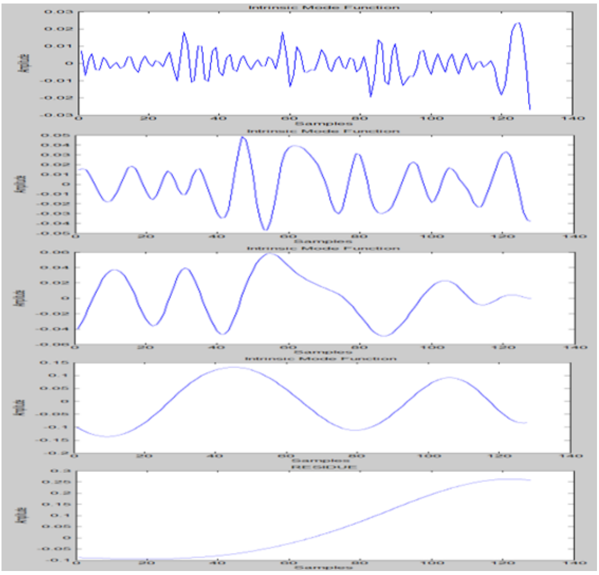


Figure 4: Decomposition of watermarked audio frame by Doubly Iterative Empirical Mode Decomposition

If the number of IMFs obtained before and after watermarking is same, it guarantees that last IMF always contains the watermark information. Binary watermark bits are extracted from the extrema of consecutive last IMFs by searching for Synchronization Codes as shown in figure 5. As the host signal is not required to extract watermark, the watermarking scheme is blind.

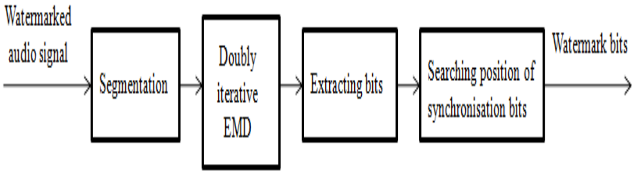


Figure 5: Watermark extraction

A. Synchronization Code

Synchronization Code is used to locate the embedding position of the watermark bits in host audio signal. Let U={1111100110101110} be original Synchronization code and V be unknown sequence of length same as U. If the number of positions in which U and V differ is $\leq \tau$, a predefined threshold, then V is considered as Synchronization Code.

B. Watermark Embedding

A binary sequence whose i^{th} bit is denote by $m_i \in \{0,1\}$, is formed by combining Synchronization Code with the watermark image before embedding. Watermark embedding is detailed as follows:

Step 1: Audio signal is segmented into frames, each of 128 samples.

Step 2: Each frame is decomposed into IMFs by Doubly Iterative Empirical Mode Decomposition.

Step 3: The binary sequence $\{m_i\}$ is embedded P times into the extrema of last IMF by Quantization Index Modulation (QIM) using rule [8]:

$$e_i^* = \begin{cases} \left\lfloor \frac{e_i}{S} \right\rfloor \cdot S + \frac{3S}{4} & \text{if } m_i = 1 \\ \left\lfloor \frac{e_i}{S} \right\rfloor \cdot S + \frac{S}{4} & \text{if } m_i = 0 \end{cases} \qquad (2)$$

where e_i and e_i^* are the extrema of last IMF of audio signal before and after embedding watermark and $\lfloor\ \rfloor$ represents floor function. S denotes embedding strength which is chosen appropriately for the maintenance of inaudibility.

Step 4: Audio frame is reconstructed using modified last IMF and all the frames are concatenated to obtain watermarked signal.

C. Watermark Extraction

The steps for watermark extraction are detailed as follows:

Step 1: Watermarked audio signal is segmented into frames.

Step 2: Each frame is decomposed into IMFs by Doubly Iterative Empirical Mode Decomposition.

Step 3: Extrema $\{e_i^*\}$ of last IMF is extracted.

Step 4: Binary bits $\{m_i^*\}$ are extracted from the extrema using following rule [8]:

$$m_i^* = \begin{cases} 1 & \text{if } e_i^* - \left\lfloor \frac{e_i^*}{S} \right\rfloor \cdot S \geq \frac{S}{2} \\ 0 & \text{if } e_i^* - \left\lfloor \frac{e_i^*}{S} \right\rfloor \cdot S < \frac{S}{2} \end{cases} \qquad (3)$$

Step 5: Consider sliding window of size $L=N_1=16$. The start index of the extracted binary data, Y, is set to INDEX=1.

Step 6: Similarity between extracted binary segment $V = Y\,(\mathrm{INDEX} : \mathrm{L})$ and U is evaluated. If the number of bits in which they differ is $\leq \tau$, then Y (INDEX : L) is considered as Synchronization Code and go to Step 8.

Step 7: INDEX value is increased by 1. Slide the window to next L=16 samples and go to Step 6.

Step 8: Similarity between second extracted segment $\dot{V} = Y\,(\mathrm{INDEX} + \mathrm{N1} + \mathrm{N2} : \mathrm{INDEX} + 2\mathrm{N1} + \mathrm{N2})$ and U bit by bit.

Step 9: INDEX is assigned with new value, $\mathrm{INDEX} \leftarrow \mathrm{INDEX} + \mathrm{N1} + \mathrm{N2}$, repeat Step 7.

Step 10: P watermarks are extracted and bit by bit comparison is made, for correction. Finally desired watermark is extracted.

Watermark embedding and extraction processes are as shown in figure 6.

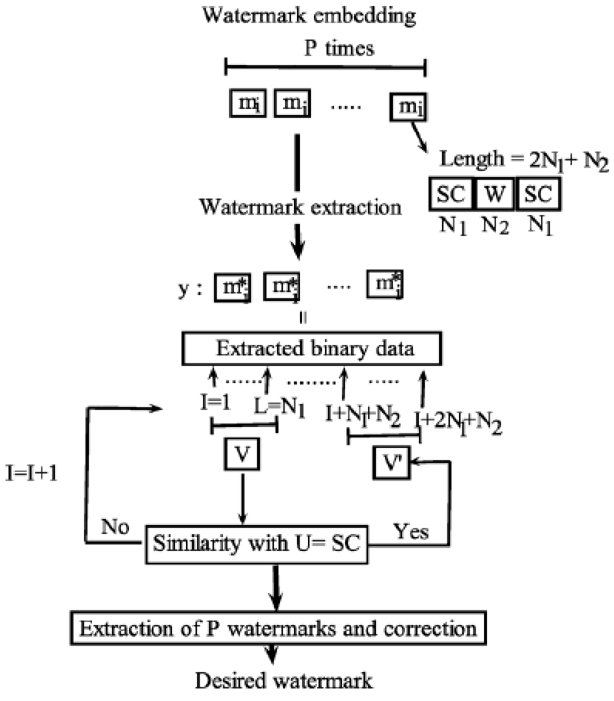


Figure 6: Embedding and extraction of watermark bits

III. PERFORMANCE ANALYSIS

The performance is evaluated in terms of data payload, Signal to Noise Ratio (SNR) between the host audio signal and watermarked signal, probability of error, Bit Error Rate (BER), and Normalized cross-Correlation (NC). According to IFPI standard, SNR of watermarked audio should be more than 20 dB. The SNR is calculated using following formula

$$\mathrm{SNR}(X, \tilde{X}) = 10 \log_{10} \frac{\sum_{i=1}^{L} X^2(i)}{\sum_{i=1}^{L} [X(i) - \tilde{X}(i)]^2} \qquad (4)$$

Watermark detection accuracy is evaluated using BER and NC defined by

$$\mathrm{BER}(W, \tilde{W}) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} W(i,j) \oplus \tilde{W}(i,j)}{M \times N} \qquad (5)$$

where $\oplus$ represents EX-OR operation and M×N is watermark image size. W represents original watermark image and $\tilde{W}$ represents extracted image. Normalized cross-Correlation is used to evaluate the similarity between the original watermark and extracted one is calculated by

$$\mathrm{NC}(\mathrm{W},\widetilde{\mathrm{W}}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} \mathrm{W}(i,j)\widetilde{\mathrm{W}}(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} \mathrm{W}^2(i,j)}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} \widetilde{\mathrm{W}}^2(i,j)}} \qquad (6)$$

Large value of NC near to 1 represents presence of watermark and lack of watermark is identified when value of NC is low. While searching for Synchronization codes two types of errors occur namely, False Positive Error (FPE) and False Negative Error (FNE) and associated probabilities are given by

$$\mathrm{P_{FPE}} = \frac{1}{2^p}\sum_{k=[0.8p]}^{p}\binom{p}{k} \qquad (7)$$

$$\mathrm{P_{FNE}} = \frac{1}{2^p}\sum_{k=0}^{[0.8p]-1}\binom{p}{k}(\mathrm{BER})^k(1-\mathrm{BER})^{p-k} \qquad (8)$$

where p is length of SC and τ is threshold [4]. The probability that SC is detected in false position is called probability of False Positive Error. The probability that a watermarked signal is declared as unwatermarked is called probability of False Negative Error. Data payload or information embedding rate of audio watermarking scheme is the number of bits embedded into one second audio section and calculated using

$$\mathrm{DP} = \frac{\mathrm{M}}{\mathrm{L}}\ (\mathrm{bps}) \qquad (9)$$

where length of audio signal is L seconds and watermark data is of size M bits.

IV. EXPERIMENTAL RESULTS

Simulations are performed on different audio signals sampled at 44.1 kHz with the length of about 10 seconds in WAVE format, namely FOLK.wav, JAZZ.wav, CLASSICAL.wav and LATIN.wav. A binary logo image of size M×N = 36×36 = 1296 bits (here M=N) is embedded watermark and is represented as W and is shown in figure 7. The 2D binaty image is converted in to 1D sequence of 1296 bits. 16 bit Barker sequence 1111100110101110 is used as Synchronization Code (SC) and is considered before and after watermark bits as shown in the figure 2. Each audio signal is segmented into number of frames each with 128 samples and threshold value τ is set to 4. The paramerter S is set to 0.98 to achieve imperceptibility. Figure 8 shows portion of audio signal FOLK.wav and its watermarked version which can not be visually distinguished from original audio.

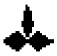
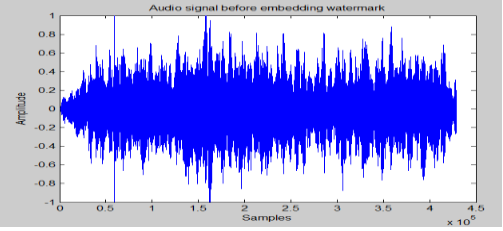


Figure 7: Binary watermark image



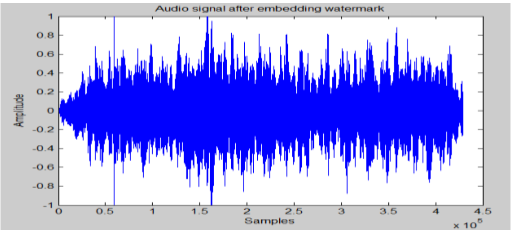


Figure 8: A portion of audio signal FOLK.wav and its watermarked version

Perceptual quality assessment is done using subjective listening tests involving 15 persons and objective evaluation tests by measuring Signal-to-Noise Ratio of the watermarked audio signal. Participants were asked to grade the dissimilarities between original and watermarked audio files. 5-grade impairment scale [11] is used and is shown in Table I. Grade 5 corresponds to excellent quality of audio signal and Grade 1 corresponds to bad quality. According to IFPI standard, SNR of above 20 dB denotes good quality of audio signal. Table II show calculated value of SNR of different test audio signals which are above 20 dB confirming to IFPI standard and Average Mean Opinion Score (MOS) of audio signals.

Table I: ITU-R quality and impairment scales

Five-grade scale	
Quality	Impairment
5 Excellent	Imperceptible
4 Good	Perceptible, but not annoying
3 Fair	Slightly annoying
2 Poor	Annoying
1 Bad	Very annoying

Table II: SNR and Average MOS between original and watermarked audio signal

Audio file	SNR (dB)	Average MOS
FOLK.wav	32.2806	4.51
JAZZ.wav	22.0357	4.47
CLASSICAL.wav	25.1445	4.43
LATIN.wav	22.8816	4.22

A. Robustness test

The watermarked audio signal is subjected to various attacks such as

Noise: White Gaussian Noise is added to watermarked signal till the Signal to Noise Ratio of resulting signal is 20 dB.

Cropping: Audio segments, each consisting of 512 samples are removed from 15 random positions and are added with noise. The segments which are contaminated with WGN are then replaced in watermarked signal.

Resampling: The watermarked signal is originally sampled at 44.1 kHz. It is re-sampled at frequencies 22.05 kHz and 11.025 kHz and then restored back at frequency 44.1 kHz.

Filtering: Low-pass Butterworth filter of order 2 with cut-off frequency 11.025 kHz is used.

Requantization: Re-quantization down to 8 bits/sample is done and then back to 16 bits/sample.
Compression (64 kbps and 32kbps): The watermarked signal is compressed and then decompressed using MPEG layer 3.
Echo-addition: An echo signal with a decay of 40% and a delay of 100 ms and is added to the watermarked audio signal.

Table III shows extracted watermarks and corresponding Bit Error Rate (BER) and Normalized cross-Correlation (NC) or different attacks on FOLK.wav file. The extracted and original watermarks are visually similar. Due to the insertion of watermark bits into extrema of last IMF, error is not detected even if WGN is added as low frequency sub-band has high robustness against the addition of noise. Table IV shows NC and BER results for JAZZ.wav, CLASSICAL.wav and LATIN.wav. Normalized cross correlation (NC) values are above 0.9778 and Bit Error Rate (BER) values are below 4%. Even the perceptual characteristics of individual audio file are different from another, Doubly Iterative Empirical Mode Decomposition adapts to all the audio signals and hence robustness is achieved. Table V shows comparison of different watermarking algorithms in terms of data payload and robustness to MP3compression. Proposed watermarking algorithm achieves highest payload.

Table III: BER and NC of extracted watermark for FOLK.wav audio signal

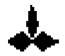
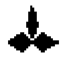
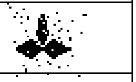
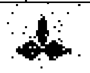
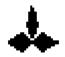
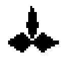
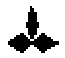
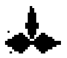
Attack type	**BER %**	**NC**	**Extracted watermark**
No attack	0	1	
AWGN	0	1	
Cropping	0	1	
Resampling (22.05 kHz)	0	1	
Resampling (11.025 kHz)	4	0.9778	
Filtering	3	0.9843	
Requantization	0	1	
MP3 Compression (64 kbps)	0	1	
MP3 Compression (32 kbps)	0	1	
Echo addition	0	0.999	

Figure 9 shows variation of probability of False Positive Error (P_{FPE}) with respect to p, which represents length of Synchronization Code. P_{FPE} tends to 0 when $p \geq 16$, confirming chosen length of SC. Figure 10 shows variation of P_{FNE} with respect to length of embedding bits. For embedding bit length ≥ 25, P_{FNE} tends to 0. Since we are using watermark image of size 1296 bits, probability of FNE is very low.

Table V: Comparison of audio watermarking methods with respect to data payload and robustness to MP3 compression

Reference	Payload (b/s)	Robustness to MP3 compression
Proposed algorithm	130	32
Khaldi & Boudraa [1]	46.9-50.3	32
Bhat et al. [4]	45.9	32
Lie & Chang [9]	43	80
Cvejic & Seppanen [10]	27.1	32

Table IV: BER and NC of extracted watermark for different audio signals

Audio signal	Attack type	BER%	NC
JAZZ.wav	No attack	0	1
	AWGN	0	1
	Cropping	0	1
	Resampling (22.05 kHz)	0	1
	Resampling (11.025kHz)	1	0.9958
	Filtering	3	0.9843
	Requantization	0	1
	MP3 Compression (64 kbps)	0	0.9980
	MP3 Compression (32 kbps)	2	0.9889
	Echo addition	3	0.9797
CLASSICAL.wav	No attack	0	1
	AWGN	0	1
	Cropping	0	1
	Resampling (22.05 kHz)	0	1
	Resampling (11.025kHz)	0	0.9972
	Filtering	0	1
	Requantization	3	0.9829
	MP3 Compression (64 kbps)	0	1
	MP3 Compression (32 kbps)	0	0.9995
	Echo addition	1	0.9940
LATIN.wav	No attack	0	1
	AWGN	0	1
	Cropping	0	1
	Resampling (22.05 kHz)	0	1
	Resampling (11.025kHz)	1	0.9950
	Filtering	0	1
	Requantization	0	1
	MP3 Compression (64 kbps)	0	1
	MP3 Compression (32 kbps)	0	1
	Echo addition	1	0.9935

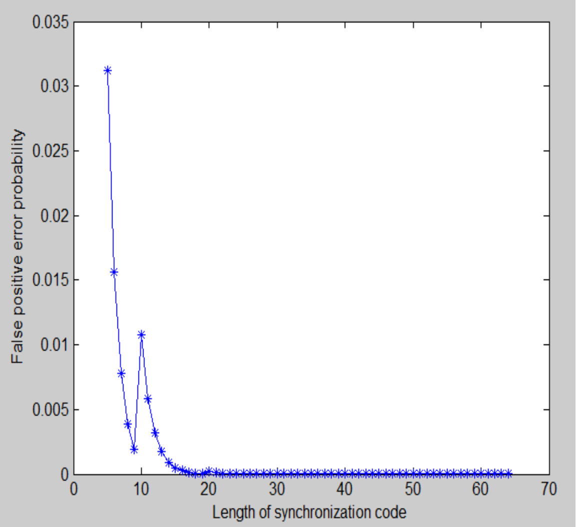


Figure 9: P_{FPE} versus SC length

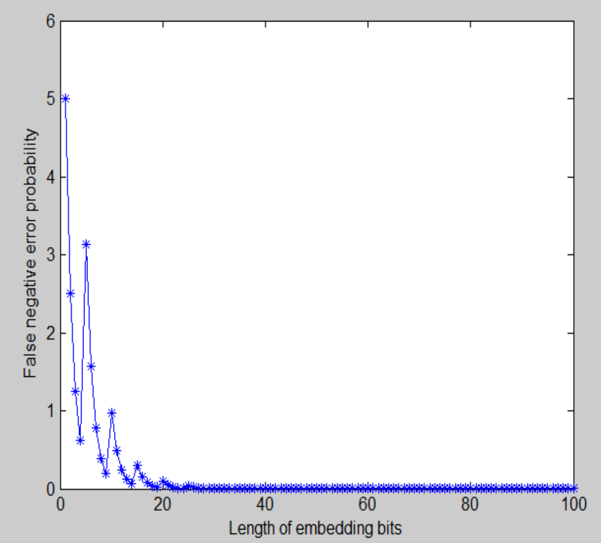


Figure 10: P_{FNE} versus length of embedding bits

V. CONCLUSION

In this paper, an adaptive and efficient watermarking scheme based on Doubly Iterative Empirical Mode Decomposition is proposed. As watermark bits are embedded in extrema points of last IMF, good performance against different kind of attacks is achieved. Binary data bits to be embedded include Synchronization Code and watermark bits, and are embedded into extrema based on QIM. Experimental results demonstrate that original and watermarked audio signals are indistinguishable. Watermark extraction algorithm is efficient and blind since host audio signal is not required during extraction. Proposed watermarking scheme is robust against various attacks like addition of WGN, cropping, resampling, filtering, requantization, MP3 compression and echo addition. This watermarking scheme has higher information embedding rate compared to other schemes and involves easy calculations. The probability of False Positive Error and False Negative Error is very low for chosen length of Synchronization code and watermark bits. The embedded data is the information related to the owner of audio file, the proposed algorithm is useful in applications like copyright protection and tracking the owner. The embedding strength S is kept constant during experiments. Parameter S should be chosen adaptively depending on magnitude values of host audio signal, to further improve the performance. Future work includes designing watermarking method for adaptive embedding problem.

REFERENCES

[1] Kais Khaladi and Abdel-Ouahab Boudraa, “Audio watermarking via EMD,” IEEE Transactions on audio, speech, and language processing, vol. 21, No. 3, March 2013.

[2] M. D. Swanson, B. Zhu, and A. H. Tewfik, “Robust audio watermarking using perceptual masking,” *Signal Process.*, vol. 66, no. 3, pp. 337–355, 1998.

[3] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, “Efficiently self-synchronized audio watermarking for assured audio data transmission,” *IEEE Trans. Broadcasting*, vol. 51, no. 1, pp. 69–76, Mar. 2005.

[4] V. Bhat, K. I. Sengupta, and A. Das, “An adaptive audio watermarking based on the singular value decomposition in the wavelet domain,” *Digital Signal Process.*, vol. 2010, no. 20, pp. 1547–1558, 2010.

[5] D. Kiroveski and S. Malvar, “Robust spread-spectrum audio watermarking,” in *Proc. ICASSP*, 2001, pp. 1345–1348.

[6] N. E. Huang *et al.*, “The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc.*, vol. 454, no. 1971, pp. 903–995, 1998.

[7] Yannis Kopsinis and SteveMcLaughlin, “Improved EMD Using Doubly-Iterative Sifting and High Order Spline Interpolation”, EURASIP Journal on Advances in Signal Processing, volume 2008

[8] B. Chen and G. W. Wornell, “Quantization index modulation methods for digital watermarking and information embedding of multimedia,” *J. VLSI Signal Process. Syst.*, vol. 27, pp. 7–33, 2001.

[9] W.-N. Lie and L.-C. Chang, “Robust and high-quality time-domain audio watermarking based on low frequency amplitude modification,” *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 46–59, Feb. 2006.

[10] N. Cvejic and T. Seppanen, “Spread spectrum audio watermarking using frequency hopping and attack characterization,” *Signal Process.*, vol. 84, no. 1, pp. 207–213, 2004.

[11] Recommendation ITU-R BT.500-11