



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume3, Issue1)

Available online at: www.ijariit.com

Audio Water Marking Using Empirical Mode of Decomposition

Prof. Karimella Vikram

Computer Science & Engineering Dept.
TKR Engineering College, Hyderabad.

Dr. K. Madhusudana

Professor in Computer Science & Engg. Dept
TKR Engineering College, Hyderabad.

Abstract: In this paper a new adaptive audio watermarking algorithm based on Empirical Mode Decomposition (EMD) is introduced. The audio signal is divided into frames and each one is decomposed adaptively, by EMD, into intrinsic oscillatory components called Intrinsic Mode Functions (IMFs). The watermark and the synchronization codes are embedded into the extreme of the last IMF, a low frequency mode stable under different attacks and preserving audio perceptual quality of the host signal. The data embedding rate of the proposed algorithm is 46.9–50.3 b/s. Re-lying on exhaustive simulations, we show the robustness of the hidden watermark for additive noise, MP3 compression, re-quantization, filtering, cropping and resampling. The comparison analysis shows that our method has better performance than watermarking schemes reported recently.

Keywords: Empirical mode decomposition, intrinsic mode function, audio watermarking, quantization index modulation, synchronization code.

I. INTRODUCTION

Digital audio watermarking has received a great deal of attention in the literature to provide efficient solutions for copyright protection of digital media by embedding a watermark in the original audio signal [1]–[5]. Main requirements of digital audio watermarking are imperceptibility, robustness and data capacity. More precisely, the water-mark must be inaudible within the host audio data to maintain audio quality and robust to signal distortions applied to the host data. Finally, the watermark must be easy to extract to prove new watermarking schemes is a very challenging problem [5]. Different watermarking techniques of varying complexities have been proposed [2]–[5]. In [5] a robust watermarking scheme to different attacks is proposed but with a limited transmission bit rate. To improve the bit rate, watermarked schemes performed in the wavelets domain have been proposed [3], [4]. A limit of wavelet approach is that the basis functions are fixed, and thus they do not necessarily match all real signals. To overcome this limitation, recently, a new signal decomposition method referred to as Empirical Mode De-composition (EMD) has been introduced for analyzing non-stationary signals derived or not from linear systems in totally adaptive way [6]. A major advantage of EMD relies on no a priori choice of filters or basis functions. Compared to classical kernel based approaches, EMD is fully data-driven method that recursively breaks down any signal into a reduced number of zero-mean with symmetric envelopes AM-FM components called Intrinsic Mode Functions (IMFs). The decomposition starts from finer scales to coarser ones. Any signal $x(t)$ is expanded by EMD as follows:

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

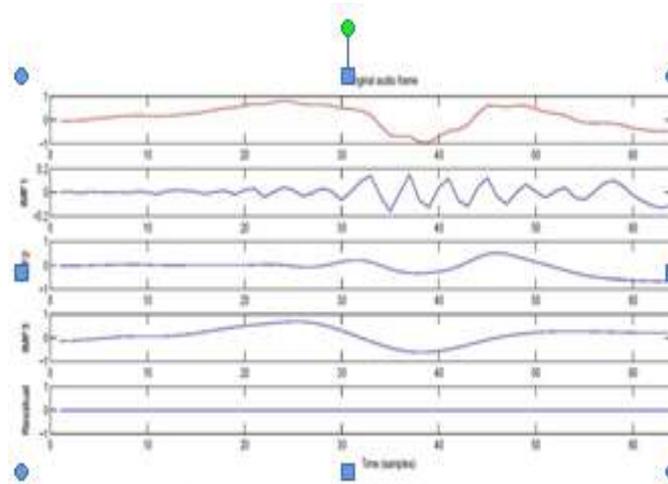


Fig. 1. Decomposition of an audio frame by EMD.



Fig. 2. Data structure (ym).

Where C is the number of IMFs and $r_{res}(t)$ denotes the final residual. The IMFs are nearly orthogonal to each other, and all have nearly zero means. The number of extrema is decreased when going from one mode to the next, and the whole decomposition is guaranteed to be completed with a finite number of modes. The IMFs are fully described by their local extrema and thus can be recovered using these extrema [7], [8]. Low frequency components such as higher order IMFs are signal dominated [9] and thus their alteration can lead to degradation of the signal. As result, these modes can be considered to be good locations for watermark placement. Some preliminary results have appeared recently in [10], [11] showing the interest of EMD for audio watermarking. In [10], the EMD is combined with Pulse Code Modulation (PCM) and the watermark is inserted in the final residual of the subbands in the transform domain. This method supposes that mean value of PCM audio signal may no longer be zero. As stated by the authors, the method is not robust to attacks such as band-pass filtering and cropping, and no comparison to watermarking schemes reported recently in literature is presented. Another strategy is presented in [11] where the EMD is associated with Hilbert transform and the watermark is embedded into the IMF containing highest energy. However, why the IMF carrying the highest amount of energy is the best candidate mode to hide the watermark has not been addressed. Further, in practice an IMF with highest energy can be a high frequency mode and thus it is not robust to attacks.

Watermarks inserted into lower order IMFs (high frequency) are most vulnerable to attacks. It has been argued that for watermarking robustness, the watermark bits are usually embedded in the perceptually components, mostly, the low frequency components of the host signal [12]. Compared to [10], [11], to simultaneously have better resistance against attacks and imperceptibility, we embed the watermark in the extrema of the last IMF. Further, unlike the schemes introduced in [10], [11], the proposed watermarking is only based on EMD and without domain transform. We choose in our method a watermarking technique in the category of Quantization Index Modulation (QIM) due to its good robustness and blind nature [13]. Parameters of QIM are chosen to guarantee that the embedded watermark in the last IMF is inaudible. The watermark is associated with a synchronization code to facilitate its location. An advantage to use the time domain approach, based on EMD, is the low cost in searching synchronization codes. Audio signal is first segmented into frames where each one is decomposed adaptively into IMFs. Bits are inserted into the extrema of the last IMF such that the watermarked signal inaudibility is guaranteed. Experimental results demonstrate that the hidden data are robust against

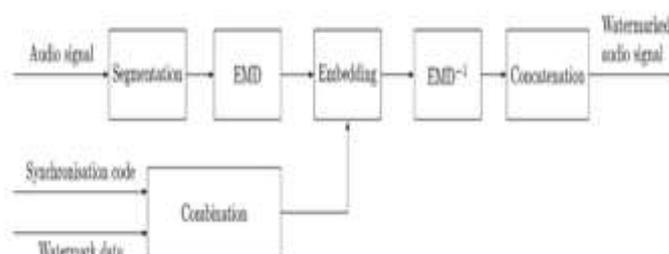


Fig. 3. Watermark embedding

attacks such as additive noise, MP3 compression, requantization, crop-ping and filtering. Our method has high data payload and performance against MP3 compression compared to audio watermarking approaches reported recently in the literature.

II. PROPOSED WATERMARKING ALGORITHM

The idea of the proposed watermarking method is to hide into the original audio signal a watermark together with a Synchronized Code (SC) in the time domain. The input signal is first segmented into frames and EMD is conducted on every frame to extract the associated IMFs (Fig. 1). Then a binary data sequence consisted of SCs and informative watermark bits (Fig. 2) is embedded in the extrema of a set of consecutive last-IMFs. A bit (0 or 1) is inserted per extrema.

Since the number of IMFs and then their number of extrema depend on the amount of data of each frame, the number of bits to be embedded varies from last-IMF of one frame to the following. Watermark and SCs are not all embedded in extrema of last IMF of only one frame. In general the number of extrema per last-IMF (one frame) is very small compared to length of the binary sequence to be embedded. This also depends on the length of the frame. If we design by N_1 and N_2 the numbers of bits of SC and watermark respectively, the length of binary sequence to be embedded is equal to $2N_1 + N_2$. Thus, these $2N_1 + N_2$ bits are spread out on several last-IMFs (extrema) of the consecutive frames. Further, this sequence of $2N_1 + N_2$ bits is embedded P times. Finally, inverse transformation (EMD⁻¹) is applied to the modify extrema to recover the watermarked audio signal by superposition of the IMFs of each frame followed by the concatenation of the frames (Fig. 3). For data extraction, the watermarked audio signal is split into frames and EMD applied to each frame (Fig. 4). Binary data sequences are extracted from each last IMF by searching for SCs (Fig. 5). We show in Fig. 6 the last IMF before and after watermarking. This figure shows that there is little difference in terms of amplitudes between the two modes. EMD being fully data adaptive, thus it is important to guarantee that the number of IMFs will be same before and after embedding the watermark (Figs. 1, 4). In fact, if the numbers of IMFs are different, there is no guarantee that the last IMF always contains the watermark information to be extracted. To overcome this problem, the sifting of the watermarked signal is forced to extract the same number of IMFs as before watermarking. The proposed watermarking scheme is blind, that is, the host signal is not required for watermark extraction. Overview of the proposed method is detailed as follows:

A. Synchronization Code:

To locate the embedding position of the hidden watermark bits in the host signal a SC is used. This code is unaffected by cropping and shifting attacks [4].

Let U be the original SC and V be an unknown sequence of the same length. Sequence V is considered as a SC if only the number of different bits between U and V , when compared bit by bit, is less or equal than to a predefined threshold τ [3].

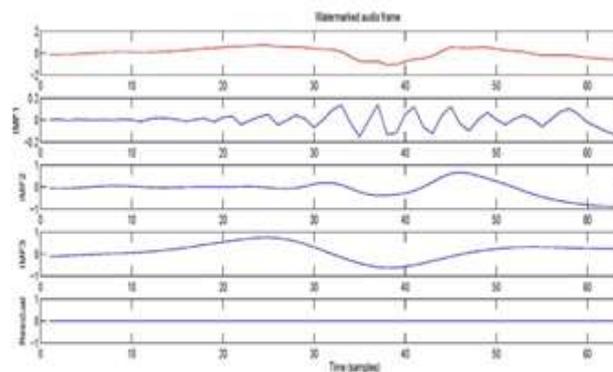


Fig. 4. Decomposition of the watermarked audio frame by EMD.

B. Watermark Extraction

For watermark extraction, host signal is splitted into frames and EMD is performed on each one as in embedding. We extract binary data using rule given by (3). We then search for SCs in the extracted data. This procedure is repeated by shifting the selected segment (window) one sample at time until a SC is found. With the position of SC determined, we can then extract the hidden information bits, which follows the SC. Let m denote the binary data to be extracted and U denote the original SC. To locate the embedded watermark we search the SCs in the sequence m bit by bit. The extraction is performed without using the original audio signal. Basic steps involved in the watermarking extraction, shown in Fig. 5, are given as follows:

- Step 1: Split the watermarked signal into frames.
- Step 2: Decompose each frame into IMFs.
- Step 3: Extract the extrema _____ of _____ .

desired watermark.

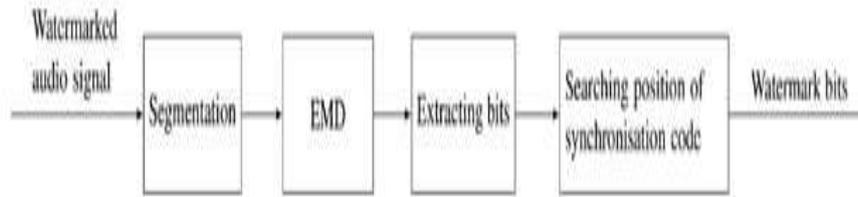


Fig. 5. Watermark extraction.

Fig. 7. Watermarking embedding and extraction processes are summarized.

III. PERFORMANCE ANALYSIS

We evaluate the performance of our method in terms of data payload, error probability of SC, Signal to Noise Ratio (SNR) between original and the watermarked audio signals, Bit Error Rate (BER) and Normalized cross-Correlation (NCC). According to International Federation of the Photographic Industry (IFPI) recommendations, a watermark audio signal should maintain more than 20 dB SNR. To evaluate the water-mark detection accuracy after attacks, we used the BER and the NCC de-fined as follows [4]:

$$BER(W, \hat{W}) = \frac{\sum_{i=1}^M \sum_{j=1}^N W(i, j) \oplus \hat{W}(i, j)}{M \times N} \quad (4)$$

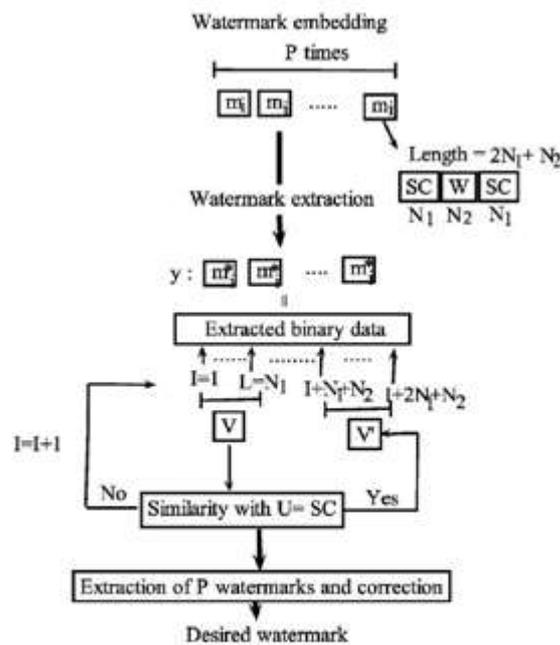


Fig. 7. Embedding and extraction processes.



Fig. 8. Binary watermark.

The original watermark and the extracted one we use the NCC measure defined as follows:

$$NC(W, \hat{W}) = \frac{\sum_{i=1}^M \sum_{j=1}^N W(i, j) \hat{W}(i, j)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N W^2(i, j)} \sqrt{\sum_{i=1}^M \sum_{j=1}^N \hat{W}^2(i, j)}} \quad (5)$$

A large NC indicates the presence of watermark while a low value suggests the lack of watermark. Two types of errors may occur while searching the SCs: the False Positive Error (FPE) and the False Negative Error (FNE). These errors are very harmful because they impair the credibility of the watermarking system. The associated probabilities of these errors are given by [3], [4]:

$$P_{FPE} = \frac{1}{2^p} \sum_{k=p-\tau}^p C_p^k$$

$$P_{FNE} = \frac{1}{2^p} \sum_{k=p+\tau}^p C_p^k (BER)^k (1 - BER)^{p-k} \quad (6)$$

where p is the SC length and τ is the threshold. P_{FPE} is the probability that a SC is detected in false location while P_{FNE} is the probability that a watermarked signal is declared as unwatermarked by the decoder. We also use as performance measure the payload which quantifies the amount of information to be hidden. More precisely, the data payload refers to the number of bits that are embedded into that audio signal within a unit of time and is measured in unit of bits per second (b/s).

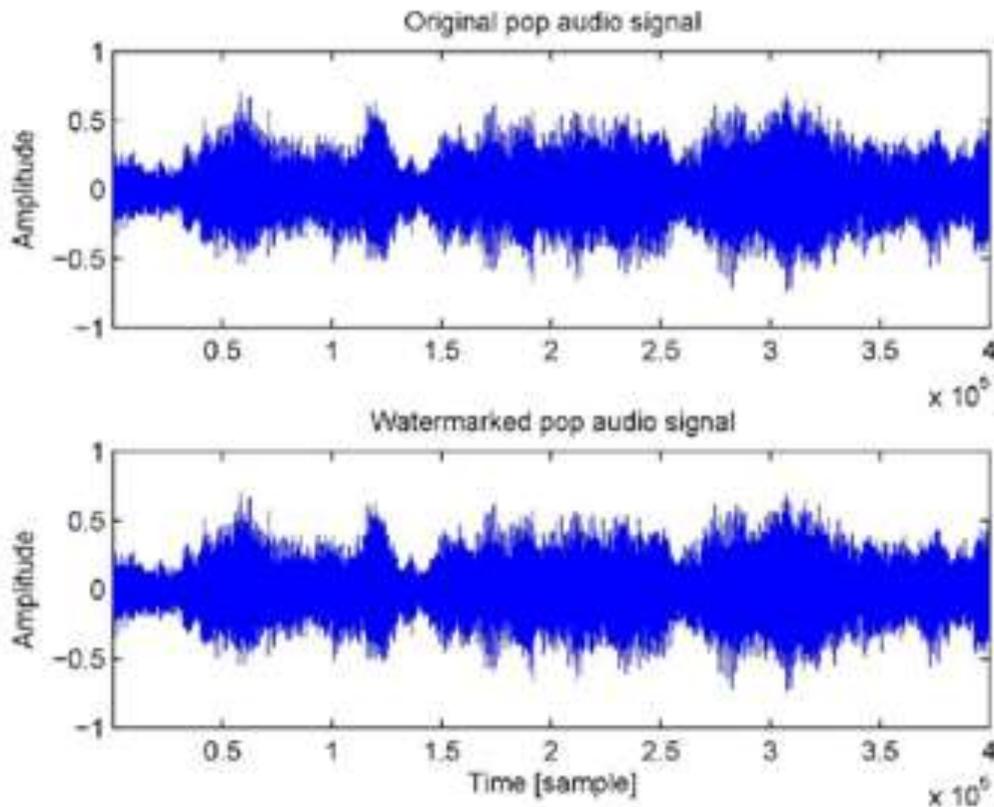


Fig. 9. A portion of the pop audio signal and its watermarked version.

TABLE I
SNR AND ODG BETWEEN ORIGINAL AND WATERMARKED AUDIO

Audio file	SNR (dB)	ODG
classic	25.67	-0.5
jazz	26.38	-0.4
pop	24.12	-0.6
rock	25.49	-0.5

RESULTS

To show the effectiveness of our scheme, simulations are performed on audio signals including pop, jazz, rock and classic sampled at 44.1 kHz. The embedded watermark, W, is a binary logo image of size $256 \times 256 \times 1$ bits (Fig. 8). We convert this 2D binary image into 1D sequence in order to embed it into the audio signal. The SC used is a 16 bit Barker sequence 1111100110101110. Each audio signal is divided into frames of size 64 samples and the threshold τ is set to 4. The S value is fixed to 0.98. These parameters have been chosen to have a good compromise between imperceptibility of the watermarked signal, payload and robustness. Fig. 9 shows a portion of the pop signal and its watermarked version. This figure shows that the watermarked signal is visually indistinguishable from the original one.

Perceptual quality assessment can be performed using subjective lie-telling tests by human acoustic perception or using objective evaluation tests by measuring the SNR and Objective Difference Grade (ODG). In this work we use the second approach. ODG and SNR values of the four watermarked signals are reported in Table I. The SNR values are above 20 dB showing the good choice of S value and confirming to IFPI standard. All ODG values of the watermarked audio signals are between -1 and 0 which demonstrates their good quality.

A. Robustness Test

To assess the robustness of our approach, different attacks are per-formed:

Noise: White Gaussian Noise (WGN) is added to the watermarked signal until the resulting signal has an SNR of 20 dB.

Filtering: Filter the watermarked audio signal using Wiener filter.

Cropping: Segments of 512 samples are removed from the watermarked signal at thirteen positions and subsequently replaced by segments of the watermarked signal contaminated with WGN.

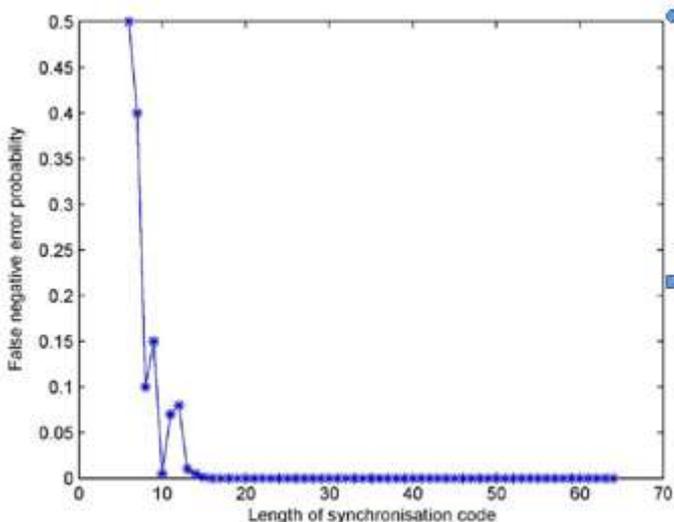


Fig. 10. False negative error probability versus synchronization code length.

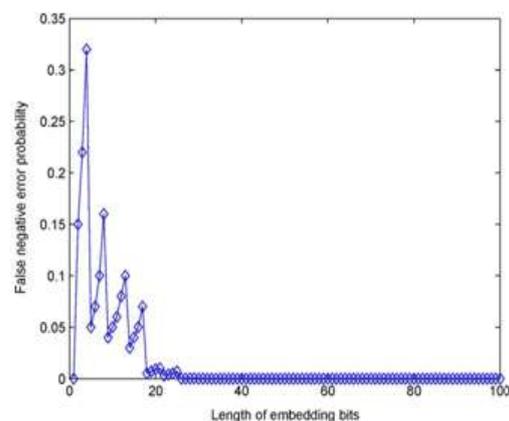


Fig. 11. False negative error probability versus the length of embedding bus.

Resampling: The watermarked signal, originally sampled at 44.1 kHz, is re-sampled at 22.05 kHz and restored back by

sampling again at 44.1 kHz.

Compression: (64 kb/s and 32 kb/s) Using MP3, the watermarked signal is compressed and then decompressed.

Requantization: The watermarked signal is re-quantized down to 8 bits/sample and then back to 16 bits/sample.

Table II shows the extracted watermarks with the associated BER and NC values for different attacks on pop audio signal. NC values are all above 0.9482 and most BER values are all below 3%. The extracted watermark is visually similar to the original watermark. These results shows the robustness of watermarking method for pop audio signal. Even in the case of WGN attack with SNR of 20 dB, our approach does not detects any error. This is mainly due to the insertion of the water-mark into IMF_c extrema. In fact low frequency subband has high ro-bustness against noise addition [3], [4]. Table III reports similar results for classic, jazz and rock audio files. NC values are all above 0.9964 and BER values are all below 3%, demonstrating the good performance robustness of our method on these audio files. This is robustness is due to the fact that even the perceptual characteristics of individual audio files vary, the EMD decomposition adapts to each one. Table IV shows comparison results in terms of payload and robustness to MP3 com-pression attack of our method to nine recent watermarking schemes

TABLE II
BER AND NC OF EXTRACTED WATERMARK FOR POP AUDIO SIGNAL BY PROPOSED APPROACH

Attack type	BER %	NC	Extracted watermark
No attack	0	1	
AWGN (20 dB)	0	1	
Filtering	6	0.9482	
Cropping	0	1	
Resampling	3	0.9783	
MP3 (64 Kb/s)	0	0.9996	
MP3 (32 Kb/s)	1	0.9876	
Requantization	0	1	

TABLE III
BER AND NC OF EXTRACTED WATERMARK FOR DIFFERENT AUDIO SIGNALS (CLASSIC, JAZZ, ROCK) BY OUR APPROACH

Audio signal	Attack type	BER %	NC
classic	No attack	0	1
	AWGN	0	1
	Filtering	0	1
	Cropping	0	1
	Resampling	2	0.9986
	MP3 (64 kb/s)	0	1
	MP3 (32 kb/s)	0	1
jazz	No attack	0	1
	AWGN	0	1
	Filtering	3	0.9964
	Cropping	0	1
	Resampling	2	0.9983
	MP3 (64 kb/s)	0	1
	MP3 (32 kb/s)	1	0.9973
rock	No attack	0	1
	AWGN	0	1
	Filtering	0	1
	Cropping	0	1
	Resampling	1	0.9989
	MP3 (64 kb/s)	0	1
	MP3 (32 kb/s)	0	1
Requantization	0	1	

[4].

[14]–[20]. Due to diversity of these embedding approaches, the comparison is sorted by attempted data payload. It can be seen that our method achieves the highest payload for the three audio files. Also, for these signals our scheme has a good performance against MP3 (32 kb/s) compression, where the maximum of BER against this last is of 1%.

Fig. 11 shows that the P_{err} is dependent on the length of watermark bits. So, we note that for the embedding bits length ≥ 25 , the P_{err} tends to 0. Since the wa-termark size in bits used is o f 1632 (>25), the obtained P_{err} is very low.

TABLE IV
COMPARISON OF AUDIO WATERMARKING METHODS, SORTED BY ATTEMPTED PAYLOAD

Reference	payload (b/s)	Robustness to MP3 (kb/s)
Proposed algorithm	46.9-50.3	32
Bhat et al. [4]	45.9	32
Lie & Chang [14]	43	80
Cvejic & Seppanen [18]	27.1	32
Yeo & Kim [15]	10	96
Tachibana et al. [17]	8.5	96
Li et al. [19]	4.2	32
Mansour & Tewfik [20]	2.3	56
Xiang et al. [21]	2	64
Kirovski & Malvar [16]	0.5-1	32

CONCLUSION

In this paper a new adaptive watermarking scheme based on the EMD is proposed. Watermark is embedded in very low frequency mode (last IMF), thus achieving good performance against various attacks. Watermark is associated with synchronization codes and thus the synchronized watermark has the ability to resist shifting and cropping. Data bits of the synchronized watermark are embedded in the extrema of the last IMF of the audio signal based on QIM. Extensive simulations over different audio signals indicate that the proposed watermarking scheme has greater robustness against common attacks than nine recently proposed algorithms. This scheme has higher payload and better performance against MP3 compression compared to these earlier audio watermarking methods. In all audio test signals, the watermark introduced no audible distortion. Experiments demonstrate that the water-marked audio signals are indistinguishable from original ones. These performances take advantage of the self-adaptive decomposition of the audio signal provided by the EMD. The proposed scheme achieves very low false positive and false negative error probability rates. Our water-marking method involves easy calculations and does not use the original audio signal. In the conducted experiments the embedding strength S is kept constant for all audio files. To further improve the performance of the method, the S parameter should be adapted to the type and magnitudes of the original audio signal. Our future works include the design of a solution method for adaptive embedding problem. Also as future research we plan to include the characteristics of the human auditory and psychoacoustic model in our watermarking scheme for much more improvement of the performance of the watermarking method. Finally, it should be interesting to investigate if the proposed method supports various sampling rates with the same payload and robustness and also if in real applications the method can handle D/A-A/D conversion problems.

[4], [14]–[20]. Due to diversity of these embedding approaches, the comparison is sorted by attempted data payload. It can be seen that our method achieves the highest payload for the three audio files. Also, for these signals our scheme has a good performance against MP3 (32 kb/s) compression, where the maximum of BER against this last is of 1%.

REFERENCES

- [1] I. J. Cox and M. L. Miller, "The first 50 years of electronic water-marking," *J. Appl. Signal Process.*, vol. 2, pp. 126–132, 2002.
- [2] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust audio water-marking 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-synchro-nized 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 water-marking," 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] K. Khaldi, A. O. Boudraa, M. Turki, T. Chonavel, and I. Samaali, "Audio encoding based on the EMD," in *Proc. EUSIPCO*, 2009, pp. 924–928.