International Journal of Advanced Science, Engineering and Technology.
ISSN 2319-5924
Vol 3, Issue3, 2014, pp38-43
http://www.bipublication.com

Non-Blind Audio Watermarking Scheme Based on Empirical Mode Decomposition

Jeebananda Panda[1], Karan Raj Gera[2], and Asok Bhattacharyya[3]

[1,3] Deptt of Electronics and Communication Engineering
Delhi Technological University. Delhi, India
[2] Deptt of Electronics and Communication Engineering
S.V National Institute of Technology. Surat, India
[1]jpanda@dce.ac.in, [2]u10ec017@eced.svnit.ac.in, [3]asok_bhattacharyya@yahoo.com
[*]Corresponding author: Email: jpanda@dce.ac.in, Tel: +91 9868417813;

[Received-03/06/2014, Published-21/07/2014]

ABSTRACT:

In this paper, an audio watermarking scheme based on Empirical Mode Decomposition (EMD) is proposed. The audio signal is separated into segments, which, by EMD, are decomposed adaptively to produce intrinsic mode functions (IMFs) that can be fully described by their extrema. These points of the last IMF are then encoded with the watermarking and synchronization bits through Quantization Index Modulation (QIM) embedding technique. Though, watermarking is done on the last IMF, the extraction process is non-blind and can be carried out on the lower order IMFs. The percentage similarity between the extracted watermark and the original watermark decreases as we extract from increasing order of IMFs. This gives an added advantage of extraction of watermark from multiple IMFs in case of masking of a particular IMF. Through rigorous simulations, the algorithm is proven to be robust against various attacks like additive noise, MP3 compression, high pass filtering, low pass filtering, and cropping.

Keywords: Empirical Mode Decomposition, Intrinsic Mode Function, Non-blind Extraction, Quantization Index Modulation, Digital Audio, Binary Watermark.

[I] INTRODUCTION

With the proliferation of the development of Internet and wireless networks, multimedia security and Digital Rights Management (DRM) have received a great deal of attention. Conventional cryptographic techniques do provide security against unauthorized access of the media but once the media has been decrypted, there is no way to track back its reproduction. Thus, Digital watermarking has been viewed as a plausible compliment to the cryptographic methods, for security against unauthorized duplication and redistribution of these multimedia data [1].

The digital audio watermark must be imperceptible, robust against attacks and should have enough data capacity. Imperceptibility and robustness are the most important of them all, but they are also incompatible. A good scheme must compromise on each other, for which, many watermarking schemes have been proposed. As, the watermark embedding process requires the lowest frequency possible in the audio signal, different methods to separate the frequencies have been considered. In [2], the watermarking bits are embedded into the peaks of the frames on which Discrete Fourier Transform (DFT) has been performed. The use of Fourier transformation, though being the most valuable tool in spectral data analysis, is refrained, as it is limited only to linear systems and stationary data series. Wavelet transform has also been considered for the same. In [3], Discrete Wavelet Transform (DWT) is used to decompose the host audio signal into several multi-resolution sub-bands. The highest resolution sub-band is chosen to embed the watermarking bits. This scheme proves to be promising

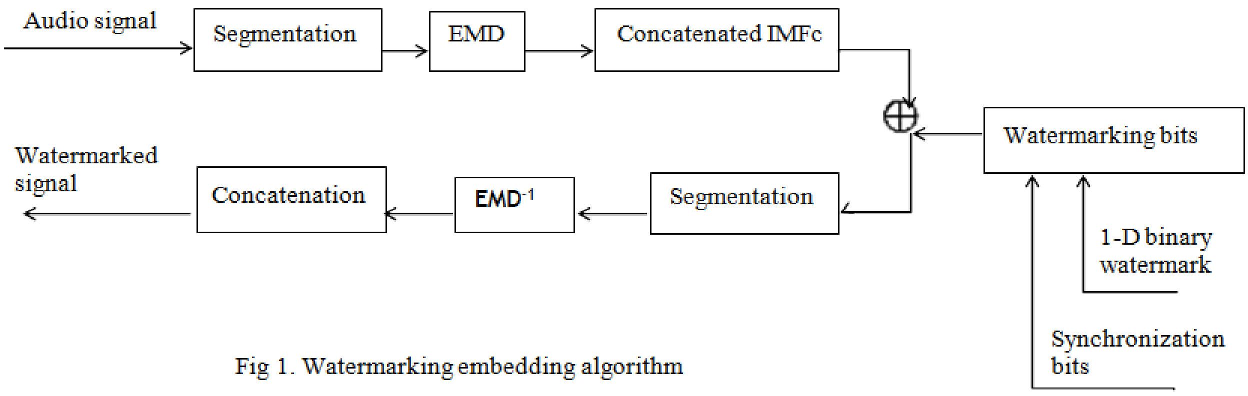


Fig 1. Watermarking embedding algorithm

considering the limitations of the fourier transform, but yet poses a limitation: the basic functions of wavelet transformation are fixed and do not necessarily match the shape of the data series considered in every instant in time. To overcome this barrier, Empirical Mode Decomposition (EMD) has been introduced that observes non-stationary signal that may or may not be derived from linear systems [4]-[6]. The basic concept of the EMD method is to expand any oscillatory signal into a set of functions, called the Intrinsic Mode Functions (IMF), defined by the signal itself. This process is called the sifting process. According to EMD, any signal can be expanded as:

$$x(t) = \sum_{j=1}^{C} IMF_j(t) + r_c(t)$$

Where C is the number of IMFs and $r_c(t)$ denotes the final residual. The IMFs bear some specific properties; the first being that the number of zero crossings and extrema points are either equal or differ by unity, and the second being that the envelopes defined by the local maxima and minima, respectively, are locally symmetric around the envelope mean. Low frequency components, such as higher order IMFs are signal dominated [7] which makes them good locations for watermark placement as their alteration can cause substantial degradation of the signal. On the other hand, watermarks embedded into lower order IMFs are most vulnerable to attacks such as additive noise, MP3 compression, filtering, cropping etc making the watermark extraction difficult.

The usage of a binary watermark is based on the fact that there are only two values to resort to with respect to a decided threshold value. This threshold value can be adaptive as in [11]. In our work, adaptive threshold is used which changes with respect to the IMF order.

In [8], the EMD process is carried out on the frames of the host audio signal. The embedding and the extraction process are performed on the higher order IMFs with a blind watermarking scheme. Whereas, during the EMD process, as the frequency selection is dependent on the standard deviation, there is a relation established between the magnitude and frequency of a signal. Hence, during the extraction process, the watermark spikes are shifted to the lower order IMFs, with their strength decreasing with the increase in the order of the IMF. Thus, in this paper, the extraction process is performed on the lower order IMFs in a non-blind fashion to yield satisfactory results in case of attacks.

[II] THE WATERMARK EMBEDDING ALGORITHM

The basic concept in embedding, as stated in [8], is that the watermark image is first converted to the binary form so that it can be represented by a matrix $M \in \{0,1\}$. It is then converted into a 1-D array $M' \in \{0,1\}$, which is combined with the synchronization bits to form the watermark array $A \in \{0,1\}$. The synchronization code is a stream of bits $\in \{0,1\}$ chosen, to differentiate between two consecutive message bits. The basic steps to embed the watermark, as shown in fig. 1, are explained as follows:

1. Divide the original audio signal (samples) into segments by considering specific number of samples for each segment. In this case, a total of 88200 samples of the audio signal have been divided into 200 segments of 441 samples each.

2. Perform the Empirical Mode decomposition separately on all the segments separately.

3. Embed P times the binary array A into the extrema points of the last IMF (IMFc) of all the segments concatenated together, by QIM [9]:

$$e_i^* = \begin{cases} \lfloor e_i / S \rfloor . S + \mathrm{sgn}(3S/4) \rightarrow m_i = 1 \\ \lfloor e_i / S \rfloor . S + \mathrm{sgn}(S/4) \rightarrow m_i = 0 \end{cases}$$

e_i denotes the extrema of the host audio signal and e_i^*, that of the watermarked signal. sgn function denotes '+' if e_i is a maxima, else it represents a '-' operation. $\lfloor\ \rfloor$ is the floor function used when e_i is a maxima, else the ceil function, is used. S is an optimum embedding strength [10].

4. Separate the concatenated watermarked IMFc into its respective segments.

5. Reconstruct segments by EMD^{-1} using the modified last IMFs of each segment and concatenate them to retrieve the watermarked signal.

[III] PROPOSED NON BLIND WATERMARK EXTRACTION ALGORITHM

The proposed extraction process is based on non-blind method, and is carried out on the lower order IMFs. As stated earlier, the dependency on standard deviation causes the magnitude of the watermarked spikes to move up to the higher frequencies from where it can be successfully extracted. As the deviation reduces as the IMF order increases, the strength of the watermark also

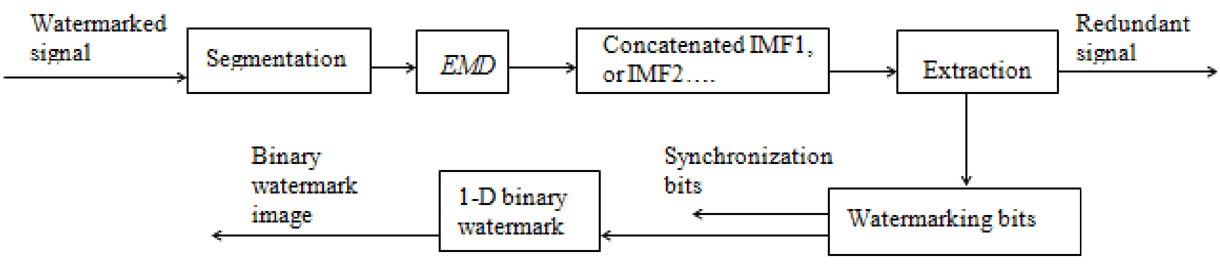


Fig 2. Watermark extraction algorithm

reduces. The basic steps involved in the extraction process, as shown in fig. 2, are explained as follows:

1. Divide the watermarked signal into segments as discussed previously.
2. Perform EMD on each segment to obtain the IMFs.
3. Concatenate the IMF order from which the watermark is to be extracted, of all the segments, from which extrema points $\{e_i^*\}$ is to be extracted.
4. Collect the extrema points of the corresponding IMF order of all the segments from the original audio signal as $\{e_i\}$.
5. Obtain $\{m_i^*\}$ from $\{e_i^*\}$ by the following equation,

$$m_i^* = \begin{cases} 1 & if\, e_i^* - e_i \geq \operatorname{sgn}\frac{S}{2^{j+1}} \\ 0 & if\, e_i^* - e_i < \operatorname{sgn}\frac{S}{2^{j+1}} \end{cases}$$

S is again the embedding strength and j denotes the IMF order with the sgn function having its usual meaning as stated earlier. The value of $j = 2$ for IMF 1 as an exception.

6. Obtain different sets of the image bits recovered from the extracted bits $\{m_i^*\}$ after deleting the synchronization codes from it.
7. Calculate the percentage similarity of each set with the original embedded 1-D array $M' \in \{0,1\}$.
8. Use the set with the highest percentage to recreate the 2-D watermark image.

[IV] EXPERIMENTAL ANALYSIS

We have formulated our scheme over a 2 sec wav audio signal sampled at 44.1 khz which gives us 88200 samples to embed the image. As discussed previously, the audio signal has been divided into 200 segments of 441 samples each. The binary image to be watermarked is taken as of $M \times N = 10 \times 20 = 200$ bits (fig. 4). This 2-D image is converted into a 1-D array in order to embed it into the audio signal. The synchronization code is chosen to be [0 1 1 1 1 1 1 0] based on the concept of byte-stuffing in synchronous transmissions in serial medium.

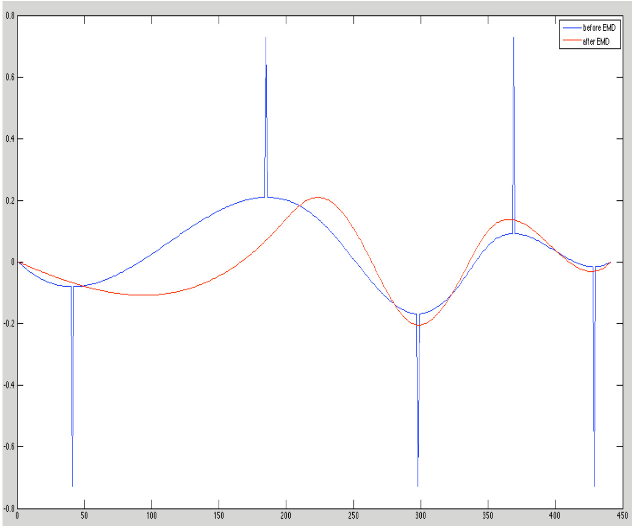
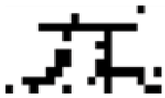


Fig. 3 Last IMF of the watermarked signal before and after EMD

According to [8], if we attempt to extract the watermark from the last IMFs of all the segments, there would, as it is be a need to consult the original host audio signal due to the shift in the extrema after watermarking as shown in fig. 3. Also, the changes in magnitude of the extrema were marginal which led us to believe that the change in magnitude due to embedding has shifted to the higher frequencies.
The embedded bits can be observed across IMF 1, IMF 2, IMF 3 clearly as shown in figs. 5, 6 and 7. However, the strength of these bits could not be observed adequately across IMF 4.

[V] RESULTS

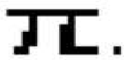
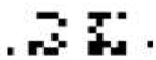
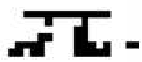
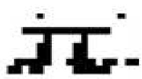
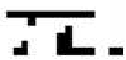
Keeping a balance of both, the imperceptibility as well as the robustness, the value of the embedding strength has been taken as 0.97, which provides us with 98.18%, match of the extracted watermark with the embedded one. The rest of the statistics have been shown in table no. 1.

IMF order from which the watermark has been extracted	Percentage similarity of the extracted watermark	Extracted watermark
IMF 1	98.18	
IMF 2	86.64	
IMF 3	72.37	

[Table-1]

To add to the robustness of the scheme, different attacks have been performed:

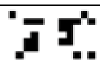
1) **Noise** – White Gaussian Noise (WGN) is added to the signal to make the Signal to Noise Ratio (SNR) as 20 dB.
2) **Filtering** – The signal is filtered through the low pass as well as a high pass filter.
3) **Cropping** – Segments of 441 samples are removes at about 13 locations and are replaced with those segments contaminated with WGN.
4) **Compression** – The watermarked signal compressed and eventually decompressed using MP3 compression at 320 kbps.

Attack type	**Percentage similarity**	**Extracted watermark**
No attack	98.18	
AWGN (20 dB)	64.93	
Cropping	81.79	
MP3 (320 kbps)	79.85	
Low pass filter	84.59	

[Table-2]

Table no. 2 shows the results of the attacks on the watermarked signal. All the attacks deter the binary watermark by less than 50%. Even with adding WGN, detection over 50% is possible. However, the detection, incase of attacks is only possible from IMF 1; IMF 2 and IMF 3 fail to respond to attacks other than the filtering attack.

Detection through multiple IMFs gives us the advantage over filtering attack. If a particular frequency corresponding to an IMF is masked, the watermarked can be successfully extracted from the other frequencies bearing the magnitude change as shown in table no. 3.

IMF order	Percentage similarity	Extracted watermark
IMF 2	74.23	
IMF 3	66.05	

[Table-3]

Moreover, an observation about the embedding strength was made. The embedding strength was expected to give linear results as compared to the percentage similarity of the extracted watermark, which it did for when no attack is performed on the watermarked signal. But, in case of an attack, the embedding strength produces a Gaussian curve with respect to the similarity of the extracted watermark.

[VI] CONCLUSION

In this paper, a new scheme of non-blind extraction of the watermark has been presented. This scheme is found to be as robust as compared to blind watermark extraction technique presented in [8] with an added property of the adaptive nature of the threshold value while detecting the watermark. Further, the extraction process has got an advantage over the filtering attack. It allows the extraction of the watermark from a different IMF if a filter masks the frequency of a particular IMF. As the watermark bits are associated with the synchronization codes, they also tend to resist a change by cropping attacks, which has been demonstrated in the simulation result of this scheme. Simulation results prove that the extracted watermark has a similarity percentage over 50% in all cases of attacks. In this scheme, the embedding strength S is kept constant at all times as it has given the optimum result in terms of robustness. To improve the inaudibility feature of watermarked audio by this algorithm, the value of embedding strength S can be made adaptive. This would allow better SNR and robustness against attacks. This can be achieved through application of power spectrum compliant technique in our future research work.

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] Dhar, Pranab Kumar, Mohammad Ibrahim Khan, and Jong-Myon Kim. "A new audio watermarking system using discrete fourier transform for copyright protection." *IJCSNS* 10.6 (2010): 35-40.

[3] Al-Haj, A., L. Bata, and A. Mohammad. "Audio watermarking

[4] using wavelets." *Networked Digital Technologies, 2009. NDT'09. First International Conference on*. IEEE, 2009.

[5] Khaldi, Kais, et al. "Audio encoding using Huang and Hilbert transforms." *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*. IEEE, 2010.

[6] Schlurmann, T. "The empirical mode decomposition and the Hilbert Spectra to analyse embedded characteristic oscillations of extreme waves." *Actes de Colloques-IFREMER* 32 (2001).

[7] 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.

[8] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, “A secure, robust watermark for multimedia,” LNCS, vol. 1174, pp. 185–206, 1996.

[9] Kais Khaldi and Abdel-Ouahab Boudraa, “Audio Watermarking Via EMD,” IEEE Trans. Audio Speech Language Process., vol. 21, no. 3, pp. 675–680, Mar. 2013.

[10] 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.

[11] Vuckovic, Vesna. "EMBEDDING STRENGTH CRITERIA FOR AWGN WATERMARK, ROBUST AGAINST EXPECTED DISTORTION." *Computing and Informatics* 29 (2010):357-387.

[11] N. K. Zaman, K. M. I. Khalilullah, Md. W. Islam, and Md. K. I.Molla, “A robust digital audio watermarking algorithm using empirical mode decomposition,” in Proc. IEEE CCECE, 2010, pp. 1-4.