Digital Audio Watermarking using EMD for Voice Message Encryption with Added Security

Kunal Gawale[1], Harshali Chaudhari[2], Vasundhara Kandesar[3], Smita Sakharwade (Makade)[4]

Student, Smt. Indira Gandhi College of Engineering, Koparkhairne, Navi Mumbai, India
Assistant Professor, Smt. Indira Gandhi College of Engineering, Koparkhairne, Navi Mumbai, India

Abstract— *Several accurate watermarking methods for image watermarking have being suggested and implemented to secure various forms of digital data, images and videos however, very few algorithms are proposed for audio watermarking. This is also because human audio system has dynamic range which is wider in comparison with human vision system. In this paper, a new audio watermarking algorithm for voice message encryption based on Empirical Mode Decomposition (EMD) is introduced. The audio signal is divided into frames and each frame is then decomposed adaptively, by EMD, into intrinsic oscillatory components called Intrinsic Mode Functions (IMFs). The watermark, which is the secret message that is to be sent, along with the synchronization codes are embedded into the extrema of the last IMF, a low frequency mode stable under different attacks and preserving the perceptual quality of the host signal. Based on exhaustive simulations, we show the robustness of the hidden watermark for audio compression, false decryption, re-quantization, resampling. The comparison analysis shows that our method has better performance than other steganography schemes recently reported.*

Keywords— Empirical mode decomposition, intrinsic mode function, audio watermarking, voice message encryption, quantization index modulation, synchronization code.

I. INTRODUCTION

A. What is a Watermark?

A watermark is defined as a distinguishing mark impressed on paper during manufacture; visible when paper is held up to the light (e.g. $ Bill). Physical objects can be watermarked using special dyes and inks or during paper manufacturing. A digital watermark is a kind of marker covertly embedded in a noise tolerant signal such as an audio, video or image data. It is typically used to identify ownership of the copyright of such signal. Our paper focuses on using the watermarking technique to embed a secret message within the transmitted audio signal (voice message).

B. Digital Audio Watermarking

Digital audio watermarking has been proved successful to provide efficient solutions for copyright protection of digital media by means of embedding a watermark in the original audio signal. Main requirements of digital audio watermarking are imperceptibility, robustness and data capacity. More precisely, the watermark must not be audible within the host audio data or the secret message hidden should not affect the audio signal in any way to maintain the quality of audio and must be robust to signal distortions applied to the host data. Finally, the watermark must be easy to extract in order to prove ownership. To achieve these requirements, seeking new watermarking schemes is a very challenging problem. Various watermarking techniques of varying complexities have been proposed. In a robust watermarking scheme solutions to different attacks are proposed but with a limited transmission bit rate. To improve the watermarking bit rate, watermarked schemes performed in the wavelets domain have been proposed. A limitation of wavelet approach is that the basis functions are fixed, and thus they do not necessarily match all real signals.

C. Audio Watermarking Techniques

There are several major audio watermarking techniques. They are:

- Discrete Cosine Transformation (DCT)
- Discrete Wavelet Transformation (DWT)
- Quantization Index Modulation (QIM)
- Empirical Mode Decomposition (EMD)

1) Discrete Cosine Transformation (DCT) :

DCT like a Fourier Transform, it represents data in terms of frequency space rather than an amplitude space. DCT based watermarking techniques are robust compared to spatial domain techniques. Such algorithms are robust against simple image processing operations like low pass filtering, brightness and contrast adjustment, blurring etc. However, they are difficult to implement and are computationally more expensive.

2) Discrete Wavelet Transformation (DWT) :

The Discrete Wavelet Transform (DWT) is currently used in a wide variety of signal processing applications, such as in audio and video compression, removal of noise in audio, and the simulation of wireless antenna

distribution. Wavelets have their energy concentrated in time and are well suited for the analysis of transient, time-varying signals. Since most of the real life signals encountered are time varying in nature, the Wavelet Transform suits many applications very well. The major drawback is lack of shift invariance, which means that small shifts in the input signal can cause major variations in the distribution of energy between DWT coefficients at different scales.

3) *Quantization Index Modulation (QIM) :*
The quantization index modulation method is resilient to noise, signal strength variations. Also it does not require linear amplifiers in the transmitter. This method enables greater efficiency than many other modes. However, it requires more complicated demodulator. Compared to this method some other modes have higher data spectral efficiency. The sidebands extend to infinity either side.

4) *Empirical Mode Decomposition (EMD) :*
In EMD, the essential functions are fastened, and therefore they are doing not essentially match all real signals. To overcome this limitation, recently, a new signal decomposition method referred to as Empirical Mode Decomposition (EMD) has been introduced for analysing non-stationary signals derived or not from linear systems in totally adaptive way. 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). This method is adequate for both non-`linear and non-stationary signals. EMD gives sharper spectrum of audio than any other method.

II. PROPOSED SYSTEM

A. *Overview of the Proposed Scheme*

The sample audio signal is first segmented to form smaller segments of the original signal. These segments undergo EMD to obtain even smaller components called Intrinsic Mode Functions (IMFs). The bits of data that is to be sent followed by the synchronization code is embedded into the IMFs. Encryption key is used during embedding, which adds a level of security to the watermarked data. During extraction, the same EMD algorithm is applied inversely. The IMFs are combined along with the residue signal to form the watermarked audio signal which contains the hidden message.

1) *Synchronization Code (SC) :*
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. 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 than or equal than to a predefined threshold.

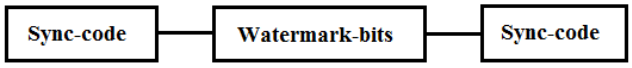

Fig.1: Data Structure

2) *Watermark Embedding :*
Before embedding, SCs are combined with watermark bits to form a binary sequence denoted by $\{m_i\}$, i^{th} bit of watermark. 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 $(2N_1+ N_2)$ sequence of bits is embedded times. Basics of our watermark embedding are detailed as follows:

- Step 1: Split original audio signal into frames.
- Step 2: Decompose each frame into IMFs.
- Step 3: Embed P times the binary sequence $\{m_i\}$ into extrema of the last IMF_c by QIM.
- Step 4: Reconstruct the (EMD^{-1}) frame using modified IMF_c and concatenate the watermarked frames to retrieve the watermarked signal (secret text message to be passed).
  $\{e_i^*\} = [e_i/S] \cdot S + sgn(3S/4)$ if $m_i = 1$
  $\{e_i^*\} = [e_i/S] \cdot S + sgn(S/4)$ if $m_i = 0$
  where e_i and e_i^* are the extrema of the host audio signal and the watermarked signal respectively.

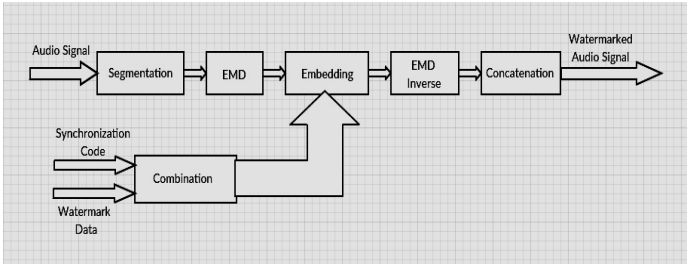

Fig.2: Watermark Embedding

3) *Watermark Extraction :*
For watermark extraction, host signal is split into frames and EMD is performed on each one as in embedding. We extract binary data using the below given rule . 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 $y = \{m_i^*\}$ 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 $\{m_i^*\}$ bit by bit. The extraction is performed without using the original audio

signal. Basic steps involved in the watermarking extraction are given as follows:

- Step 1: Split the watermarked signal into frames.
- Step 2: Decompose each frame into IMFs.
- Step 3: Extract the extrema $\{e_i^*\}$ of IMF_c.
- Step 4: Extract m_i^* from e_i^* using the following rule $m_i^* = 1$ *if* $e_i^* - [\, e_i^*/S] \cdot S \geq sgn\ (S/2)$
  $m_i^* = 0$ *if* $e_i^* - [\, e_i^*/S] \cdot S < sgn\ (S/2)$
- Step 5: Set the start index of the extracted data, y , *to* $I = 1$ and $L = N_1$ select samples (sliding window size).
- Step 6: Evaluate the similarity between the extracted segment $V = y(I : L)$ and bit by bit. If the similarity value is $\geq r$, then V is taken as the SC and go to Step 8. Otherwise proceed to the next step.
- Step 7: Increase I by 1 and slide the window to the next $L = N_1$ samples and repeat Step 6.
- Step 8: Evaluate similarity between the second extracted segment, $V' = y(I + N_1 + N_2 : I + 2N_1 + N_2)$ and U bit by bit.
- Step 9: $I \leftarrow I + N_1 + N_2$, of the new I value is equal to sequence length of bits, go to Step 10 else repeat Step 7.
- Step 10: Extract the P watermarks and make comparison bit by bit between these marks, for correction, and finally extract the desired watermark.

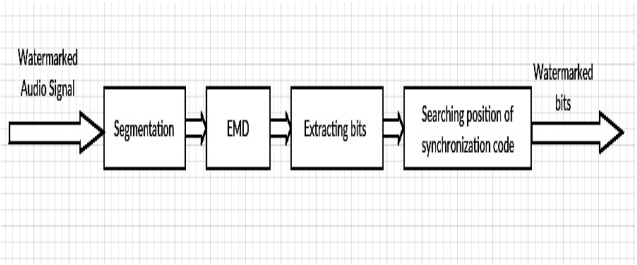

Fig. 3: Watermark Extraction

4) *Encryption :*
Encryption is the most effective way to achieve data security. In encryption, a key is a piece of information that determines the functional output of a cryptographic algorithm. In our system, the DES algorithm is used for encrypting and decrypting data. DES was classified as a predominant symmetric-key algorithm for the encryption of electronic data. The key is entered during the embedding process which needs to be verified before the secret message is extracted. The same key is used for encryption and decryption, thus making it a symmetric key algorithm. This enhances the robustness of the watermarking technique. The key based technique for encryption ensures that the hidden data is not retrieved by someone who's not intended to access the file. A valid key should always be entered at the decryption side.

5) Decryption :
Decryption is the process of taking encoded or the encrypted data and using a valid key, converting it back to understandable form or in this case, to extract the watermark. Decryption is done by entering the valid key during watermark extraction. This is done symmetrically i.e. the same key used for encryption and decryption. Because both parties have the same key, the decryption essentially is performed by reversing some part of the encryption process. In case if an invalid key is entered during decryption then the user won't be able to extract the embedded watermark.

III. WORKFLOW

The figure explains the flow of the system and the steps of shifting process used for watermark embedding. For the watermark i.e. the text message to be embedded in an audio, the audio signal needs to be segmented into multiple IMFs. Each IMF can then be embedded with the secret data along with the synchronization bits. The concatenation of all the IMFs gives the resultant watermarked audio signal. The extremas of the audio signal are first found so as to form the mean envelope. The envelope is then processed using EMD to find the resultant IMFs. The signal that is left after extracting the IMFs is called the residue signal.

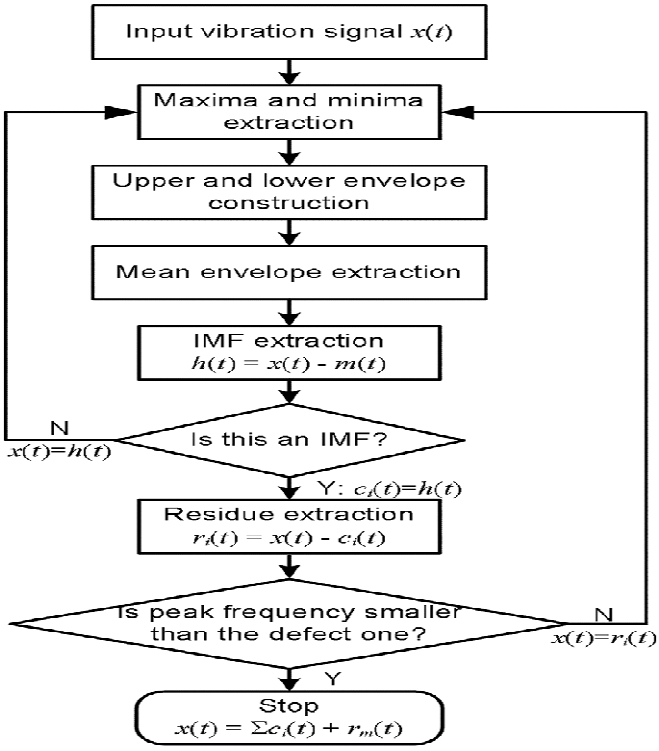

Fig. 4:Finding IMF and Watermark Embedding

IV. IMPLEMENTATION

A. *Shifting Process :*

The Shifting process is used for watermark embedding. There are two conditions that need to be satisfied by the IMF function, viz.

- In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one.
- At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero**.**

The implementation of the shifting process can explained using the following steps-

- Step 1: Take a complicated data set *x(t)* (the audio signal that will hold the text message data) and identify all the upper extrema of *x(t).*
- Step 2: Interpolate the local maxima to form the upper envelope *u(x).*
- Step 3: Identify all the lower extrema of *x(t).*
- Step 4: Interpolate the local minima to form the lower envelope *l(x).*
- Step 5: Calculate the mean envelope,
  $m(t) = [u(x) + l(x)]/2$
- Step 6: Extract the mean from the signal,
  $h(t) = x(t) - m(t)$
- Step 7: Check whether *h(t)* satisfies the IMF condition.
  YES: *h(t)* is an IMF, stop shifting.
  NO: let *x(t) = h(t),* keep shifting.

There seems to be negligible or no difference between the original audio and the watermarked audio, that contains the hidden text message. obtained by this method. The extraction is carried out in the similar fashion by applying EMD inversely and concatenating the IMFs and the residue signal.

B. *Experimental Results :*

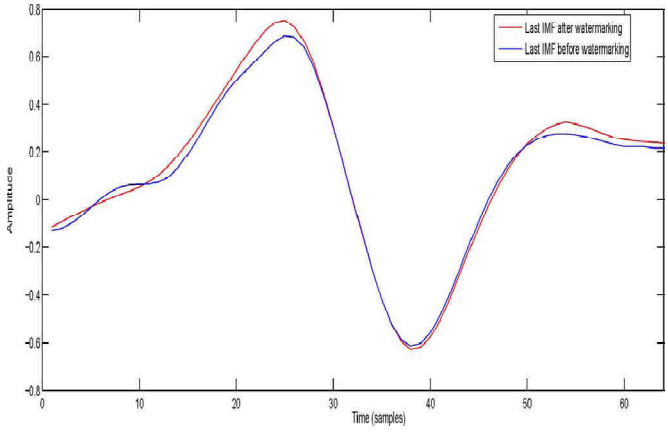

Fig. 5: Comparison of Watermarked and Original Signal

The above graph shows the difference between the original signal and the signal after watermarking. The difference is not much and the distortion is not audible to the human ear.

V. CONCLUSION

The secret text message that is embedded in the form of a watermark is in very low frequency, it is also associated with synchronization codes and thus the synchronized watermark has the ability to resist shifting and cropping. Extensive simulations over different audio signals indicate that the proposed watermarking scheme has comparatively greater robustness against common attacks. In all audio test signals, the text message introduced no audible distortion to the transmitted voice message. Experiments demonstrate that the watermarked audio signals are indistinguishable from original ones. Our watermarking method involves easy calculation and uses the same algorithm in inverse for extracting the text data from the voice messages. In the conducted experiments the embedding strength is kept constant for all audio files.

VI. FUTURE SCOPE

This method can be effectively used in sending hidden text messages over audio signals. With the increasing use of social media applications that allow us to send voice messages, our technique can be collaborated with such applications so as to send secret text messages within the voice message in real time. This technique can also be used for message passing in military where prime focus is given to maintain the confidentiality and authenticity of transmitted data. Secret text messages can be hidden in the form of a watermark within the audio signals to be transmitted. A key can also be specified during the encoding process. A text message can masked with a random audio signal thus, fooling the intruder. The proposed technique also includes DES algorithm, a symmetric key matching algorithm, to maintain optimum security. Intruder trying to access the hidden text has to know the key used during encryption. No further action is possible in case a invalid key is entered.

REFERENCES

[1] Kais Khaldi and Abdel-Ouahab Boudraa, Senior member, “Audio Watermarking via EMD” IEEE transactions on audio, speech and language processing, VOL.21, NO.3 MARCH 2013.

[2] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques,” in *Proc. ACL-02 Conf. Empirical Methods Natural Lang. Process.*, 2002, pp. 79–86.

[3] P. D. Turney, “Thumbs up or thumbs down?: Semantic orientation applied To unsupervised classification of

reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 417–424.

[4] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms Through gloss classification," in *Proc. 14th ACM Int. Conf. Inf. Knowl. Manage.* 2005, pp. 617–624.

[5] S. H. Choi, Y.-S. Jeong, and M. K. Jeong, "A hybrid recommendation Method with reduced data for large-scale application," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 5, pp. 557–566, Sep. 2010.

[6] T. Mullen and N. Collier, "Sentiment analysis using support vector machines With diverse information sources," in *Proc. EMNLP*, 2004, pp. 412–418.

[7] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc.10th ACMSIGKDD Int. Conf.Knowl. Discov.DataMining*, 2004, pp. 168–177.

[8] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proc. 8th Conf. Eur. Chap. Assoc. Comput. Linguist.*, Morristown, NJ:Assoc. Comput. Linguist., 1997, pp. 174–181.

[9] A. Esuli and F. Sebastiani, "SENTIWORDNET: A publicly available Lexical resource for opinion mining," in *Proc. 5th Conf. Lang. Res. Eval.*,2006, pp. 417–422.