Design and Simulation of Audio Watermarking using Empirical Mode Decomposition for Copyright Protection

Pragathi T L 1 Meghashree A C 2
1 M. Tech Student 2 Assistant Professor
1,2 Department of DECS
1,2 VTU PG Centre, Mysore-570019

Abstract—Digital watermarking is drawing attention as a new method for protecting multimedia content from illegal copying. In this paper, we present a alternative audio watermarking procedure to provide copyright protection to the digital audio by modifying audio samples directly by making use of algorithm called Empirical Mode Decomposition (EMD). The original audio signal is sampled at 44.1 kHz to split it into non-overlapping frames. These frames are further decomposed by EMD adaptively into intrinsic oscillatory mono-component signals known as Intrinsic Mode Functions (IMFs). The digital watermark along with synchronization codes are embedded into the last IMF extrema. This is low frequency mode, stable under different attacks and preserving the audio quality of host signal. Based on the simulations, robustness of hidden watermark is demonstrated for different attacks such as noise, filtering, compression, cropping, resampling and re-quantization. It is seen that the data embedding rate or payload of the algorithm is 46.9-50.3 bits per second.

Keywords: Audio watermarking, Quantization Index Modulation, Synchronization code, Empirical Mode Decomposition, Intrinsic Mode Functions

I. INTRODUCTION

The recent growth in computer networks has allowed multimedia data to be easily distributed over the internet. The widespread and easy access to multimedia contents and possibility of duplication has motivated the need for digital rights. To provide copyright protection for digital data, two techniques were developed. They are encryption and watermarking. Encryption protects the content during transmission from sender to receiver. But once the data is received and decrypted, the content is not protected any longer. So, the watermarking technique replaces the encryption technique by embedding into the original data in such a way that it is always present in the host data. The watermark is a secret code or image or pattern of bits that is embedded into the original data which acts to verify both the content of the data and the owner of the file. Digital audio files are particularly the most abused for illegal copying because they can be downloaded and can be copied very easily. Audio watermarks are the special signs or patterns that are embedded into audio file to identify the ownership of copyright of that audio. Compared to image and video watermarking, audio watermarking is bit difficult to design because Human Auditory System [HAS] is more sensitive compared to Human Visual System [HVS]. Thus good audio watermarking schemes are very difficult to design [1]. The audio watermarking techniques are divided into two types: Time domain techniques and frequency transform domain techniques. The time domain techniques include methods where the embedding is performed without any transformation. Frequency domain transformation techniques include embedding after applying transformations like Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) to the host signal.

The watermark should be inaudible within the audio signal to maintain audio quality and robust to the signal distortions applied to host audio signal. To achieve these requirements seeking new watermark schemes is very challenging [2]. Therefore different techniques of various complexities have been proposed [2]-[5]. In [2], a robust audio watermarking scheme for different attacks is proposed but transmission bit rate was limited. To increase the bit rate, watermark schemes in the wavelet domain have been proposed [4], [5]. In the wavelet transform domain, the watermark is embedded either in the high or in low frequency band; therefore watermark can’t be embedded in the whole transform domain and the basis functions are fixed and thus they do not necessarily match the real signals. To overcome this we present a method called Empirical Mode Decomposition (EMD) which can embed the watermark to the last IMF that contains less vulnerable to attacks. Thus this method is much more resistant to common signal processing manipulation when compared with other methods [3] [6] [7] [8]: The EMD method provides perfect time-frequency localization properties than discrete wavelet transform (DWT) domain masking. This decomposition method is highly adaptive and very efficient for perceptual quality (inaudibility) and robustness. The watermark can be embedded into the last IMF which is not vulnerable to attacks; therefore, the robustness is great than other methods and the embedded information can be restored fully.

II. PROPERTIES OF WATERMARK

The following are some of the properties that need to be satisfied for efficient application of watermark technology. The important properties are robustness, imperceptibility, data capacity and security.

A. Robustness:

Even after attacks like cropping, filtering, compression, resampling and re-quantization, watermark should be readable and extractable.

B. Imperceptibility:

The watermark that is embedded should not degrade the perceptual quality of host audio signal.

C. Data Capacity:

The amount of information that can be embedded into a signal is also an important issue. A user has to be able to change the amount embedded to suit different applications. The host signal should be able to carry more information without degrading the quality of host signal.
D. Security:
The property of security is of great importance to watermarking schemes. It implies that the watermark should be impossible to be detected.

III. EMPirical MODE Decomposition

Empirical Mode Decomposition (EMD) is a new method of analyzing nonlinear non-stationary signals. It was proposed by Norden E. Huang in 1996 [9]. EMD is an entirely data driven algorithm and it does not depend on any predefined basis functions. EMD breaks down non-stationary, multi-component signal into its mono-components. Such mono-components are called Intrinsic Mode Functions (IMFs). Each IMF is a signal that must meet the following criteria [9]:

1) The number of extrema and zeros are equal or their difference is not greater than 1.
2) The signal has “zero mean” the mean value of the envelope determined by maxima and the envelope determined by minima is equal 0 at every point.

These conditions show the essence of the EMD decomposition: non-stationary signal is decomposed into symmetrical oscillations around zero amplitude, which are easy for further analysis. The decomposition starts from finer scales to coarser ones. Any signal \( x(t) \) is expanded by EMD as shown below:

\[
x(t) = \sum_{j=1}^{C} IMF_j(t) + r_e(t)
\]

Where \( C \) is the number of IMFs and \( r_e(t) \) denotes the final residual. The IMFs are nearly orthogonal and all have nearly zero means. The number of extrema is decreased when going from one mode to next, and the whole decomposition is guaranteed to be completed with a finite number of modes. Low frequency components such as higher order IMFs are signal dominated and thus their alteration may lead to degradation of signal. Hence these are considered to be good locations for watermark placement.

A. EMD Algorithm:

EMD method decomposes a complex signal into a number of intrinsic mode functions (IMFs). Decomposition consists of following steps:

1) Step 1: Identify all the local extrema, and then connect all the local maxima by an interpolation method to get upper envelope \( e_{\text{max}} \). Repeat the procedure for the local minima to produce the lower envelope \( e_{\text{min}} \).
2) Step 2: Find the mean value of the envelopes,

\[
m(t) = \frac{e_{\text{max}} + e_{\text{min}}}{2}
\]

3) Step 3: Determine the difference between the signal \( x(t) \) and the mean of upper and lower envelope \( m(t) \) to obtain the first component. If it satisfies the IMF condition, then it would be the first IMF. Otherwise, it is treated as the original signal and sifting process (Step 1-3) is repeated.
4) Step 4: Separate IMF from the original signal \( x(t) \) to obtain the residue and consider it as the new data and repeat the above described process until the residue is monotonic. The sum of all IMF components and the residue gives the original signal.

IV. SYNCHronization CODE

Synchronization is the key issue during watermark extraction process especially when the host audio is manipulated by desynchronization attacks. Any shift in the bits positions makes extracting schemes unable to succeed. The main goal of the synchronization schemes is to find where the new shifted positions are. In the first step of a watermark extraction procedure, the detecting process tries to align itself with the watermarked block. If it fails, it is impossible to extract the digital watermark bits from the host audio, causing a false detection. In practice, time or frequency scaling attacks can lead to desynchronization. Therefore, in order to have a robust audio watermarking, it is required to employ a synchronization technique that can resist such attacks. A very common technique uses synchronization codes to insert into the audio signal.

A synchronization code is used to locate the embedding position of the hidden watermark bits in the host signal. When choosing a synchronization code, the following three issues need to be considered,

- How the code bits are inserted into the audio signal.
- The number of code bits.
- Distribution of 0s and 1s over the code length.

The synchronization code used here is a 16 bit Barker sequence 111100110101110. This code is unaffected by any cropping and shifting attacks during signal transmission. Let U be the original SC and V be an unknown sequence of data of the same length. Sequence V is considered as a SC, then compared bit by bit to the length of U which is less or equal than to a predefined threshold \( \tau \) [4] which is set to 4.

V. PROPOSED METHOD

The proposed watermarking algorithm performs an idea of hiding a watermark bit together with a synchronization code (SC) to the host audio signal in the time domain manner. The digitalized watermarked bit together with SC (Fig. 2) is embedded into the last extrema of the IMFs. A bit 0 or 1 is inserted per extrema.

Fig. 1: Decomposition of watermarked audio frame by EMD

Fig. 2: Data structure \[ \{m_i\} \]

The number of IMFs and their number of extrema depends on the data of each frame. The number of bits to be embedded in the extrema varies from last-IMF of one frame.
to the next following frames. Digitalized watermark bit and following Synchronization Codes are not all embedded in extrema of last IMF of only one frame instead it will spread out in the extrema of last IMF of all consecutive frames since the binary sequence to be embedded is larger than the number of extrema per last IMF. This depends also on the length of the frame.

![Block diagram](image_url)

**Fig. 3: Block diagram**

Let \( N1 \) and \( N2 \) are the numbers of bits of Synchronization Code and watermark bit respectively, the bit length of binary sequence to be embedded to the host signal is equal to \( 2N1+N2 \). Thus, these \( 2N1+N2 \) bits are spread out on several last-IMFs or extrema of the consecutive frames. Further, this sequence of \( 2N1+N2 \) bits is embedded \( P \) times. Extraction of watermark bit from each last IMF is done by searching for SCs. The host signal is not used for watermark extraction and therefore the algorithm is blind.

Overview of the proposed algorithm is as follows:

**A. Embedding the Watermark:**

Before embedding, SCs are combined with watermark bits to form the binary sequence \( \{m_i\} \in \{0,1\} \). The process of watermark embedding is shown in Fig. 2 and the steps are detailed as follows:

1) **Segmentation** - The audio signal is sampled at 44.1 kHz and segmented into frames of size of 64 samples each.
2) **EMD** - On each frame, EMD is conducted to extract the IMFs.
3) **Embedding** - The binary sequence \( \{m_i\} \) is embedded \( P \) times into extrema of the last IMF by QIM [10].

\[
e_i^* = \begin{cases} 
\frac{2i}{S} \cdot S + \text{sgn} \left( \frac{12}{4} \right) & \text{if } m_i = 1 \\
\frac{2i}{S} \cdot S + \text{sgn} \left( \frac{2}{4} \right) & \text{if } m_i = 0
\end{cases}
\]

(3)

Where \( e_i \) and \( e_i^* \) are the extrema of \( \text{IMF}_c \) of the host audio signal and the watermarked signal respectively. \( \text{sgn} \) function is equal to “+” if \( e_i \) is a maxima, and “-” if it is a minima, \( \lfloor \ \rceil \) denotes the floor function, and \( S \) denotes the embedding strength chosen to be 0.98 to maintain the inaudibility constraint.

4) **Reconstruction of frames** - Using the modified IMFs, the frames are reconstructed by applying Inverse EMD.
5) **Concatenation** - The obtained frames are concatenated to get the watermarked audio signal.

**B. Extraction of Watermark:**

Here binary data is extracted then SCs in the extracted data is searched. 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 follow 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. Basic steps involved in the watermarking extraction are as follows:

1) **Segmentation** - The watermarked signal is split into frames.
2) **EMD** - Each frame is decomposed into associated IMFs.
3) **Extraction of bits** - Extrema \( \{e_i^*\} \) of IMF is extracted. Extract \( m'_i \) from \( e_i^* \) using the following rule:

\[
m'_i = \begin{cases} 
1 & \text{if } e_i^* - \frac{L}{S} \cdot S \geq \text{sgn} \left( \frac{2}{4} \right) \\
0 & \text{if } e_i^* - \frac{L}{S} \cdot S < \text{sgn} \left( \frac{2}{4} \right)
\end{cases}
\]

(4)

4) **Searching of SC** - The searching involves the following steps:

   1. **Step 1:** Set the start index of the extracted data, \( y \) to \( I=1 \) and \( L=N_1 \) select samples (sliding window size).
   2. **Step 2:** Evaluate the similarity between the extracted segment \( V = y_1 : L \) and \( U \) bit by bit. If the similarity value is \( \leq \tau \), then \( V \) is taken as the SC and go to Step 3. Otherwise proceed to the next step.
   3. **Step 3:** Increase \( I \) by 1 and slide the window to the next \( L=N_1 \) samples and repeat Step 2.
   4. **Step 4:** Evaluate the similarity between the second extracted segments \( V' = y_1 : L_4 \) and \( U \) bit by bit.
   5. **Step 5:** If \( I+N_1+N_2 \) of the new value \( I \) is equal to the sequence length of bits, go to next step (Extraction of P watermarks) else repeat Step 3.
5) **Extraction of P watermarks** - Extract the P watermarks and compare between these marks bit by bit, for the correction, and extract the desired watermark.
VI. FLOWCHART

Fig. 5: Flowchart of embedding and extraction process

VII. RESULTS

MATLAB is used as a tool to evaluate the performance of the proposed system. The embedded watermark, W, is a binary logo image of size bits (Fig. 8). We convert this 2D binary M*N 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 is set to 4. The 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.

A. Signal to Noise Ratio (SNR):

SNR (Signal to Noise Ratio) is used to analyze the imperceptibility and quality of the watermarked audio file in comparison to the original audio. The SNR measures the similarity between two audio files i.e. how close two audio files are with each other. According to International Federation of the Photographic Industry (IFPI) recommendations, a watermarked audio signal should maintain more than 20 dB SNR. Moreover, the higher the SNR value of a watermark object, the better the degree of hidden message imperceptibility. SNR can be described as:

\[
SNR(dB) = 10 \log_{10} \left( \frac{\text{Audio signal}}{\text{Watermarked audio signal}} \right)
\]

(5)

B. Bit Error Rate (BER):

Bit Error Rate is used to evaluate the watermark detection accuracy after signal processing operations. BER is defined as follows:

\[
BER(W, W') = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} W(i,j) \oplus W'(i,j)}{M \times N}
\]

(6)

Where \(\oplus\) is the XOR operator and M*N are the binary watermark image sizes. W and W' are the orthogonal and the recovered watermark respectively.

C. Normalized Cross-Correlation (NC):

To evaluate the similarity between the original watermark and the extracted one, we use the NC measure defined as follows:

\[
NC(W, W') = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} W(i,j) \oplus W'(i,j)}{\sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} W^2(i,j)} \times \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} W'^2(i,j)}}
\]

(7)

D. Robustness Check:

To assess the hardness of our approach, totally different attacks are performed:

1) Noise: White mathematician Noise (WGN) is extra to the watermarked signal till the ensuing signal has associate SNR of 20 dB.

2) Filtering: Filter the watermarked audio signal.

3) Cropping: Segments of 512 samples are removed from the watermarked signal at 13 positions and later replaced by segments of the watermarked signal contaminated with WGN.

4) Re-sampling: The watermarked signal, originally sampled at 44.1 kHz, is re-sampled at 22.05 kilohertz and restored back by sampling once more at 44.1 kHz.

5) Re-quantization: The watermarked signal is re-quantized all the way down to 8 bits/sample then back to 16 bits/sample.

E. Simulation Results:
VIII. CONCLUSION

There is a need for copyright protection of the audio signals particularly for the audio files on the internet. Digital watermarking system is a mean for copyright protection of audio files. In this paper a innovative adaptive watermarking theme using EMD algorithm is projected. Watermark is embedded in low frequency mode (last IMF), thus achieving smart performance against various attacks. Watermark is expounded to synchronization codes and thus the synchronized watermark has the ability to resist shifting and cropping. Information bits of the synchronized watermark area unit embedded in the last IMF of the audio signal using QIM. Intensive simulations over different audio signals indicate that the planned watermarking theme has larger lustiness against common attacks than the other recently planned algorithms. This theme has higher payload and better performance compared to earlier audio watermarking methods. The watermark does not introduce audible distortion. From the result it is seen that watermarked audio signals are indistinguishable from original ones. The watermarking method involves easy calculations and does not use the original audio signal. The simulation results show that this method has good Signal to Noise Ratio. In future, the proposed technique can be improved by the use of artificial intelligence. The technique can be applied to video files and can be used to hide audio file in video. The characteristics of the human auditory and psychoacoustic model can also be included in this method. And we can also investigate if this method supports various sampling rates with the same payload and robustness.

REFERENCES