Audio Watermarking Schemes and Embedding Algorithms

Dissertation submitted in the partial fulfillment of requirement for the award of degree of

Master of Engineering

in

Electronics and Communication

Submitted by:

PRABHJOT KAUR

Roll No: 801261014

Under the guidance of:

Dr. KULBIR SINGH

Associate Professor

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

(Established under the section 3 of UGC Act, 1956)

PATIALA – 147004 (PUNJAB)

DECLARATION

I hereby declare that the work which is being presented in the dissertation entitled, **"Audio Watermarking Schemes and Embedding Algorithms"** in partial fulfillment of the requirement for the award of degree of Master of Engineering in Electronics and Communication submitted in Electronicsand Communication Engineering Department of Thapar University, Patiala, is an authenticrecord of my own work carried out under the supervision of Dr. Kulbir Singh, Associate Professor, ECED and refers other researcher's work which are duly listed in the reference section.

The matter presented in this dissertation has not been submitted in any other University/Institute for the award of degree.

Date: 20/6/14

(PRABHJOT KAUR)
Roll No: 801261014

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

(Dr. KULBIR SINGH)
Associate Professor
ECED, Thapar University

Countersigned by:

Head
ECED, Thapar University
Patiala-147004

Dean of Academic Affairs
Thapar University
Patiala- 147004

ACKNOWLEDGEMENT

First of all, I would like to express my gratitude to **Dr. KULBIR SINGH**, **Associate Professor**, Electronics & Communication Engineering Department, Thapar University, Patiala for his guidance and support throughout this thesis work. I am really very fortunate to have the opportunity to work with him. I found his guidance to be extremely valuable.

I am thankful to the **Head of Department, Professor (Dr.) SANJAY SHARMA** of Electronics & Communication Engineering Department for their encouragement and inspiration for the execution of this thesis work.

I am also thankful to the entire faculty and staff of Electronics & Communication Engineering Department for the help and moral support which went along the way for the successful completion of this thesis work.

My greatest thanks are to all who wished me success especially my parents. Above all I render my gratitude to the Almighty who bestowed self-confidence, ability and strength in me to complete this work for not letting me down at the time of crisis and showing me the silver lining in the dark clouds. I do not find enough words with which I can express my feelings of thanks to my dear friends for their help, inspiration and moral support which went a long way in successful competition of the present study.

Prabhjot Kaur
Roll no. 801261014

ABSTRACT

Increasing amount of digital data sharing tools has created copyright infringement issues. This has become the threat to content providers. Digital watermarking remained the most powerful tool for intellectual property protection in literature. There are several applications of digital watermarking including copyright protection, fingerprinting, broadcast monitoring and content authentication.

Depending upon the requirements of an application, watermark can be robust, fragile or semi-fragile. Work is carried out on audio fragile watermarking schemes. Fragile audio watermarks possess the characteristic to get destroyed even after slightest modification and hence is used for temper detection. Fragility does not require watermark to possess robustness- one of three conflicting properties of watermark, i.e. robustness, imperceptibility and pay load. Thus robustness is not an important issue in this approach, so other remaining two parameters can be thought to improve to higher extent than usual. In proposed algorithm, imperceptibility is targeted and SNR = 68.89 dB and MSE of watermarked audio of order 10^{-9} is achieved. Achieved SNR and MSE proves to be better in MSE by 10,000 times and SNR by 30 dB, than algorithms reported recently.

Herein, a new signal adaptive fragile audio watermarking technique is proposed, in which audio signal is segmented and decomposed by Empirical Mode Decomposition. Mean of last Intrinsic Mode Function obtained for each segment is calculated and matrix is formed. Watermark bits are inserted into singular values of matrix via Quantization Index Modulation to assure some extent of robustness. Empirical Mode Decomposition and Singular Value Decomposition provide good degree of imperceptibility factor. Using all these techniques in one package provided us with better results than the recent researches.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER

1

INTRODUCTION

Introduction part introduces the concept in order to convey its background, importance, application and need to study. It clears the concept in thorough manner and is a key to growth of concept, available literature and future possibilities.

1.1 PREAMBLE

As broadband connects whole world, and whole data on the earth going digital, challenges and paths for innovations are being revealed. Elasticity, decreasing prices of digital gadgets and easiness in usage of their software facilitate people to create, have made it so natural to share the multimedia data over the entire network. This creates the possibility of perfect imitation in digital era, which has made the protection of intellectual property, ownership and prevention of tempering in multimedia data, an important research and technological issue [1]. Digital Watermark was proposed as a new and alternative method to administer intellectual property rights. Following emphasis is on audio digital data protection via audio watermarking. Various watermarking schemes are being proposed till date and different watermarking algorithm are being developed. This section is followed by electronic watermark introduction and various embedding algorithms, along with advantages and disadvantages of every algorithm.

1.2 DIGITAL WATERMARK

First electronic watermark was invented by Andrew Tirkel in 1992. It is based on bit manipulation of Least Significant Bits, which makes it very easy and rapid watermark embedding procedure [25]. Later on, different transforms were used to decompose the signal on various scales, and different new watermark embedding locations were formulated. Usage of different transform made the watermark to be embed in various domains- time domain [15], frequency domain [21] and time-frequency domain [26]. It is desired that transform should decompose the signal with any statistical properties, because audio signal source is purely random. Following this, watermark research grew on the steps made by transform evolution, i.e. as the new transform was evolved as a remedy of previous transform flaw, so does the watermark via using that transform. Also, along with different transforms, use of human perception curve, Non- uniformity of DFT coefficients and various mathematical tools like Singular Value Decomposition (SVD), Quantization Index Modulation (QIM) is implemented to enhance results.

1.3 WATERMARK EMBEDDING DOMAIN

Appropriate locations are chosen for embedding watermark to retain good quality like high imperceptibility, high robustness and etc. So watermark is embedded in various domains, when evaluated, gave a wide range of results. Depending on kind of result obtained, particular domain is used for embedding for particular application. For example some application require robustness at the most, so imperceptibility factor can be compromised in that case [1]. Others may require imperceptibility/audio quality on above and robustness, pay-load may be compromised. Different domains used for watermark embedding, are being discussed as follows-

1.3.1 TIME DOMAIN

Watermark is embedded in signal without decomposing the signal in other domain, or simply altering the domain of the signal. Very first method is Least Significant Bits (LSB) encoding method. LSBs of signal samples are altered according to watermark code bits. Obviously, if not cared, it may cause distortion. In worst cases it would be audible which is not acceptable. Thus watermark strength is restricted in this case. So, many other methods, like taking audio signal as watermark is considered and applied. It includes echo hiding, patchwork and etc. In echo hiding, echo is added into various segments of signal, thus is spread in time [17]. All these methods have positives and negatives. Negatives include their poor resistance to the attacks on signal. Positive point include the simplicity and easiness in implementation [13].

1.3.2 FREQUENCY DOMAIN

Watermark is embedded in selected range of frequencies. Once embedded, it is hard to delete watermark. Very first transform used in this domain is Fourier Transform (FT).

Signal is transformed from time domain to frequency domain. FT gives energy scales in frequency domain which are referred as Fourier coefficients. Watermark is embedded in Fourier coefficients using various watermark embedding algorithm [21]. However, there are some drawbacks of Fourier Transform (FT), due to which various modified versions of FT were proposed [30]. Modified version of FT include Short-Time Fourier Transform (STFT) and etc. Other transform like Discrete Cosine Transform, Discrete Wavelet Transform came one by one as a solution to problems faced in transform already in use. Embedding watermark using these transforms make it more robust and improves the result.

1.3.3 TIME-FREQUENCY DOMAIN

Watermark is spread in both time as well in frequency domain. It comprises mainly two methods, and those are Spread-spectrum method (SS) [11] and Empirical mode Decomposition (EMD) [20]. Spread spectrum is generally a linear method and is similar concept as used in communication, spreading watermark information in broader frequency range. On the other hand, EMD divides the signal into range of frequencies called Intrinsic Mode Functions (IMF). IMFs are generally amplitude modulated as well frequency modulated signals. Watermark is embedded in selected range of frequencies, here IMF using different embedding algorithm [26]. Watermark embedded in this domain is difficult to be removed and thus is highly robust.

1.4 INTRODUCTION AND EVOLUTION OF TRANSFORMS

Transform is way through which one can see and analyse a signal in one domain, using its other known domain. This is especially required when it is impossible to get required information from the known domain. For example Fourier Transform tells about frequency components along with their magnitudes, using time domain signal. Following transforms are described step by step as they evolved.

1.4.1 FOURIER TRANSFORM AND EMBODIMENTS

Fourier series suggests that any periodic signal can be represented as summation of different frequency sinusoids with varying amplitudes. It was named in the honour of French mathematician- Fourier, who made great contributions in study of Heat propagation and trigonometric theory [20]. Further, this idea was extended to aperiodic signals via Fourier transform. Fourier series/transform transforms vision from time domain to frequency domain of the signal. It is mathematically defined as –

$$X(F) = \int_{-\infty}^{\infty} x(t)e^{-jwt}\ dt \tag{1}$$

Limitations

Though great in concept, yet this idea has a shortcoming of restricting the signal source to be linear in nature. Being a kernel based or integral transform, Dirchlet conditions needs to be fulfilled. This restricts the signal properties. Thus, Fourier is unable to represent the signal with properties beyond its scope [20]. For example, one signal with one transient is resolved by FT. Then, to represent the transient, FT will need a number of frequency components. It leads to energy dispersion in frequency domain of the signal, and is not desired.

Motivation

Many other forms of transforms were used to solve for the gaps discussed. Any transient or non-stationarity originates from non-linearity of source. Hence, while considering a non-stationary signal, signal is divided into segments, so that every segment is stationary in itself. This too had a flaw, of assumption of segment to be stationary. Secondly, segmentation is done with same scale, hence not all segments are necessarily stationary.

1.4.2 DISCRETE COSINE TRANSFORM

Fourier Transform when used for a time domain transient, it uses so many frequency components to represent that abrupt transient. This causes the time-frequency-energy distribution difficult, or in simple word spreads the energy in frequency domain. A new transform evolved *Discrete Cosine Transform* (DCT) came up as a remedy of energy dispersion in Fourier Transform when dealt with signal containing abrupt transients. DCT in contrast deal with only cosine terms. It helps it concentrating the energy in frequency domain and thus is better transform than FT to analyse the signals with transients. Thus DCT offers frequency resolution of larger set of signals [20].

Limitations

Though time-energy-frequency distribution problem is solved to some extent, but not completely. It is still impossible to independently deal with non-linear signals.

Motivation

As Fourier is itself a integral transform and thus embodiments based on FT will have same kind of problems. So, rather than improvements on Fourier Transform, a novel approach is required for non-linear signals. Moreover, considering the positive points, watermarking was done in DCT coefficients and it proved better than the watermarking results via FT. Many embodiments are implemented in DCT watermarking. Simplest method is to decompose the signal by DCT, and embed watermark in DCT coefficients.

1.4.3 WAVELET APPROACH

Wavelet approach is essentially adjustable window Fourier Spectral Analysis. Its mathematical definition is as follows –

$$W(a,b;X,\psi) = |a|^{-1/2}\int_{-\infty}^{\infty} X(t)\psi^*\left(\frac{t-b}{a}\right)dt \qquad (2)$$

in which ψ^* is the basic wavelet function that satisfies very general conditions of a transform to be defined. Parameter a is scaling factor and b is translation of the origin. Generally ψ^* is not orthogonal for every value of parameter a for continuous wavelets. However, one can make the wavelet orthogonal by choosing a set of values of a and constraining parameter a to attain any value from that set. This discrete wavelet analysis will miss physical signals having scales dissimilar from values of a. Apart from this characteristic, discrete or continuous, wavelet approach is linear [20].

Limitations

Problem with most commonly used Morlet wavelet is the leakage generated due to limited time duration of basic wavelet function. It makes the quantitative definition of energy-frequency-time distribution difficult. In spite of all these drawbacks and difficulties, wavelet is still most commonly used in practical applications.

Motivation

Rather than fixed choice of filters, or functions; an adaptive approach is required. Approach should take its filters / functions, and should result as per the signal transition per unit time.

1.5 DIFFERENT METHODS AND TRANSFORMS IN AUDIO WATERMARKING

Optimum “holes” are peculiarly decided to embed information in them. Audio watermarking moves step by step following the steps in transform evolution. This follow up was because better resolution of new transform. Being a purely non-stationary signal, it is necessary for the transform to be decompose better, giving every information for each kind of signal. Every watermarking scheme is categorised according to transform used, and thus domain differs accordingly.

Figure 1.1 describes and categorises each of the watermarking scheme into its corresponding domain. Other tools and embodiments are also written for more clarity.

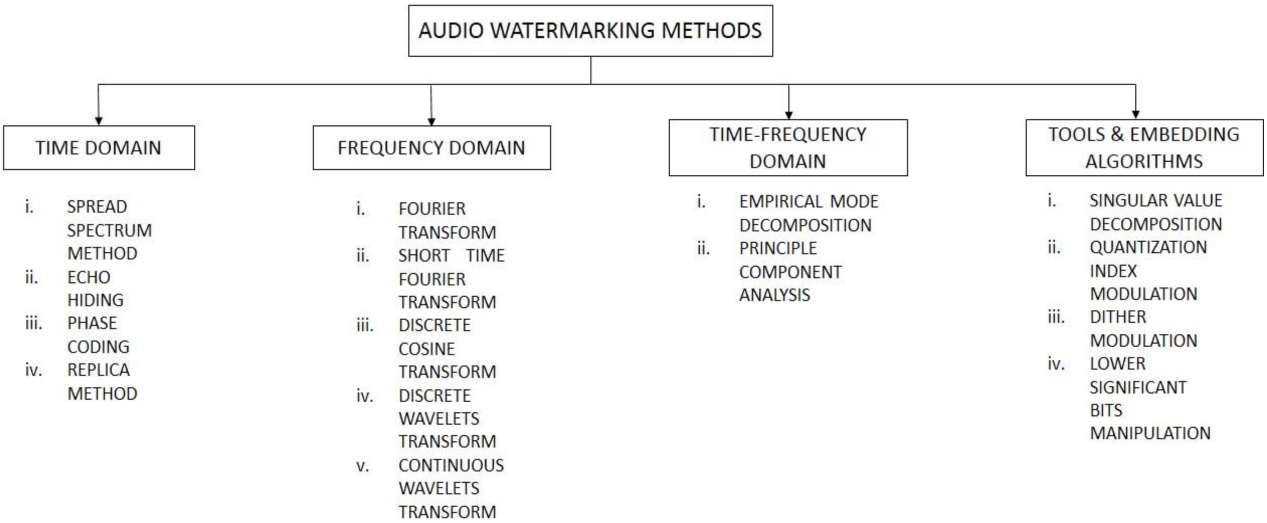


Figure 1.1: Different Audio Watermarking Methods

1.6 ORGANIZATION OF THESIS

Chapter – 2 starts with the very basic conditions need to be fulfilled for a transform to be complete and to be in usable form. It is followed by methods for analysis of non-stationary signals. The mathematical description, advantages and their flaws are being discussed along with description of every method. Different embodiments are been done on existing methods, to check best possible set of values of parameters. These all embodiments are listed following description of methods based on different transforms. Different algorithms result are listed, based on different transforms. One can conclude the best candidate for watermarking by comparing the results quantitatively. Finally, based on knowledge gathered and gaps discussed, problem for thesis is formulated.

The proposed method is described in chapter – 3, explores the need and idea of proposed algorithm. Watermark embedding and extraction algorithm, explained with flow chart is presented. Different performance analysis parameters are evaluated for justification of being better. Performance is checked even after attacks either malicious or not.

The applications of proposed method are also being discussed in chapter – 4. It has been shown and proved its betterment than other algorithms.

Finally, the conclusion of thesis along with future possibilities in the area of work of thesis is presented in chapter – 5.

CHAPTER 2

LITERATURE SURVEY

The literature review sums up, deduce and evaluates existing “literature” on specific concept. This helps to initialize research with a good start by developing whole knowledge of said concept field. Literature review can sequentially tell about researches being already done, gaps can be found out and thus new topics for research are defined.

2.1 ANALYSIS OF NON-STATIONARY SIGNALS

In audio watermarking, signal to be dealt is audio signal. Audio signal as defined by its characteristics is non-stationary. All kernel based approaches constrain the signal to be with finite discontinuities to make the integral finite and thus making the transform defined for that signal. Also, when one tries to represent any non-stationarity of signal in frequency domain, transform will output so many frequency component in order to reconstruct the signal accurately. This will disperse the energy of the signal in frequency domain and truncating the frequency domain will not reconstruct the signal with precision [20].

Summing up, all existing methods are based on decomposition of signal in time domain series into a component basis function satisfying –

a) Completeness of basis b) Orthogonality of basis

Its examples include Fourier Transform, Wavelet Analysis, Cosine Transform, Principal Component Analysis and etc. Before Proceeding to next topic, one should know the meaning of instantaneous frequency concept [20].

2.1.1 INSTANTANEOUS FREQUENCY

The concept of instantaneous energy and instantaneous envelop are well established. Concept of instantaneous frequency is rather controversial. Existing opinions ranges its definition in different manner but is only for ‘mono-component signal’ [20].

Let us consider a simple example, to illustrate the restrictions have to be imposed on data as discussed by Gabor (1946), Bedrosian (1963) and, more recently, Boashash (1992): for any function to have a meaningful instantaneous frequency, the real part of its Fourier transform has to have only positive frequency. This restriction can be proven mathematically as shown in Titchmarsh (1948) but it is still global. For data analysis, this requirement has to be translated into physically implementable steps to develop a simple method for applications. For this purpose, the restriction condition needs to be modified from a global one to a local one, and the basis has to satisfy the necessary conditions listed in the last section. Let us consider some simple examples to illustrate these restrictions physically, by examining the function, $x(t) = sin\ t$, a simple sinusoid [20].

Its Hilbert transform is simply cos t. The phase plot of $x\{y$ is a simple circle of unit radius as in figure 2.1a. The phase function is a straight line as shown in figure 2.1b and the instantaneous frequency, shown in figure 2.1c, is a constant as expected. If one moves the mean by an amount, say, then,

$$x(t) = a + \sin(t) \qquad (3)$$

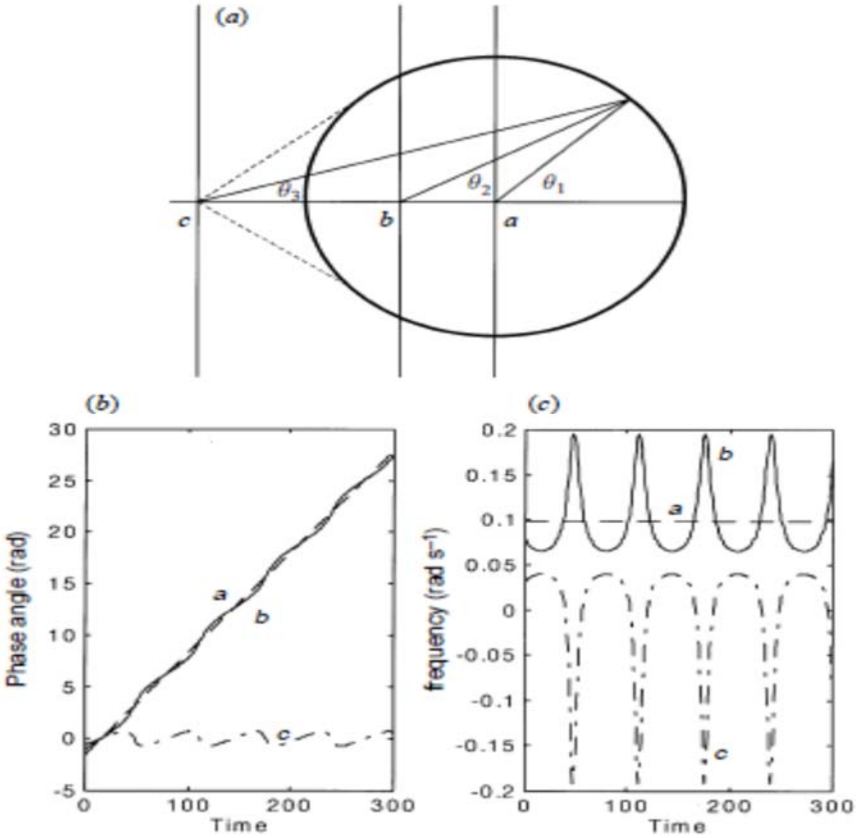


Figure 2.1 Physical interpretation of instantaneous frequency. (A) The phase plane for the model functions of $x(t) = a + sin(t)$. (A) $a = 0$; (B) $a < 1$; (C) $a > 1$. (B) The unwrapped phase function of the model functions. (C) The instantaneous freq. computed [20]

There are methods apart from transforms discussed in chapter 1. These methods are as follows:

2.1.2 THE WIGNER – VILLE DISTRIBUTION

Wigner-Ville distribution is also called as Heisenberg wavelet. From its definition, it is the Fourier transform of the central covariance function.

For any time-series X(t), central covariance can be defined as

$$V(w,t) = \int_{-\infty}^{\infty} C_c(\tau,t)\, e^{-iw\tau} \qquad (4)$$

The problem of this method is cross terms as indicated by negative powers for some frequency ranges. The result can also omit these negative powers by using window method, thus it follows the same shortcomings as that of Fourier analysis. Like fixed scale is used for segmentation of audio signal. So segment may or may not be stationary/linear. So adaptivity is required for segmentation as per the statistics of signal per unit time.

In nut-shell, though it was a new and unique idea, but still problem of decomposing non-linear signals persists [20].

Extension to this method was made by Yen, he used Wigner-Ville distribution to define wave packets to reduce a complicated set of data to finite number of simple components. This approach is widely accepted and is applied to various set of problems. It requires good judgment while dealing with complex set of data.

2.1.3 EVOLUTIONARY SPECTRUM

Priestley proposed a basic idea to extend Fourier concept, to analyse spectrum of signal, to more general generalized basis: from trigonometric functions to orthogonal functions indexed by time t and defined for all real w, frequency parameter.

According to this concept, any random variable, (t) , can be expressed as

$$X(t) = \int_{-\infty}^{\infty} \varphi(w,t) dA(w,t) \qquad (5)$$

where $dA(w,t)$ is the Stieltjes function for the amplitude and is related to the spectrum as follows-

$$E(|dA(w,t)|^2) = d\mu(w,t) = S(w,t)dw \qquad (6)$$

where $\mu(w,t)$ is the spectrum, and $S(w,t)$ is the spectral density at a specific time t, also designated as the evolutionary spectrum.

The evolutionary spectrum gained popularity in earthquake community. The main problem with this method is to find a method to define the basis function. For this method to work, basis function is to be decided posteriori. No systematic way is offered till date, so it was a little difficult to construct a signal spectrum for a given data set. Following this difficulty, earthquake community considered it as data simulation process rather than data analysis. An evolutionary spectrum is assumed at first hand, then possible set or signal will be reconstructed based on assumed spectrum [20].

Although evolutionary spectrum was never useful and EMD replaces it with a truly adaptive representation for non-stationary processes.

2.1.4 EMPIRICAL ORTHOGONAL FUNCTION EXPANSION

The empirical orthogonal function expansion (EOF) is also known as principle component analysis, or singular value decomposition method. The principle of EOF is summarized as follows –

For any real signal $z(x,t)$, the EOF will reduce to

$$z(x,t) = \sum_{1}^{n} a_k(t) f_k(x) \tag{7}$$

in which,

$$f_j . f_k = \delta_{jk} \tag{8}$$

The orthogonal basis, $\{f_k\}$, is the collection of empirical eigen functions defined by

$$C. f_k = \lambda_k f_k \tag{9}$$

where C is sum of inner products of variable.

EOF represents a sweeping departure from all above mentioned methods, for expansion basis is derived from data. Therefore, being a posteriori, it is highly efficient. The acute flaw of EOF is that it only gives a distribution of the variance in modes defined by $\{f_k\}$, but this distribution by itself does not recommend scales or frequency content of signal. One should deal each mode as independent variations, though if viewed with care, these decompositions/ variations are not unique. A single component even if basis is orthogonal, doesn’t usually contain physical meaning [20].

Vautard and Ghil *et. al* in 1989 proposed singular spectral analysis method, in this Fourier transform of the EOF. Again here, it is to be assured that each EOF component is stationary. Otherwise Fourier Transform of EOF will make less sense and purpose will not be solved. Subsequently, singular spectral analysis is not a real improvement.

Regardless of its drawbacks, it is most popular especially because of its adaptive nature. Applications including oceanography and meteorology use EOF at most.

2.1.5 EMPIRICAL MODE DECOMPOSITION & HILBERT TRANSFORM

Following methods of analysing non-stationary signal, a new approach is defined using concepts already defined. It mainly requires two steps – First is to pre-process the data using Empirical Mode Decomposition, obtain Intrinsic Mode Functions. Second is to apply Hilbert transform to the obtained IMFs. This will give out the energy-frequency-time distribution, referred as Hilbert spectrum. Existence of Hilbert transform is not only the limitation here, but also the existence or definition of a meaningful instantaneous frequency which is restrictive and local [20].

There are few conditions for IMF to be defined-

a) Number of extreme and zero crossing should differ by at most one.

b) At any point, the mean value of the envelope should be zero. Envelope here in is defined by local maxima and minima.

First condition makes sense in terms of making signal decomposition narrow in frequency domain. Second condition is new and is for non-stationary processes. Fulfilling second condition, no other frequency component can exist in any of the mode. Following mentioned steps are followed while decomposing via EMD [20]-

a) Identify all extreme of $x(t)$

b) Interpolate the local maxima to form an upper envelope $u(x)$.

c) Interpolate the local minima to form a lower envelope $l(x)$

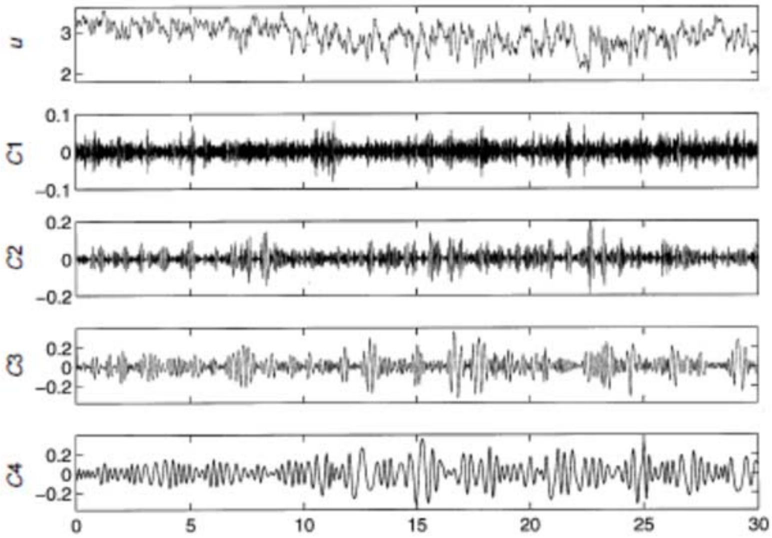
d) Calculate mean envelope: $m\ (t) = [u(x) + l(x)]/2$.

e) Extract the mean from the signal: $h\ (t) = x\ (t) - m\ (t)$

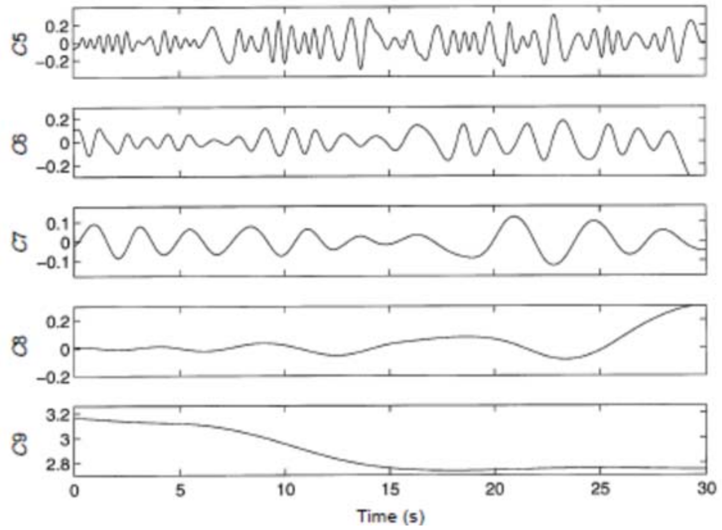
f) Check whether $h\ (t)$ satisfies the IMF condition.

YES: $h\ (t)$ is an IMF, stop sifting; NO: let $x\ (t) = h\ (t)$, keep sifting

Repeating steps d), e) and f) is called sifting procedure. It is done to choose correct IMF so as to fulfil the conditions. It is called sifting procedure because every time fast oscillations are extracted and slow ones are remained. So in essence actually sifting a range of oscillations and extracting them, is happening. It is carried out until the later can be considered as zero mean



a)



b)

Figure 2.2 : The resulting emd components from the wind data: a) The original data and the components C1-C4; b) The components C5-C9. Notice the last component, C9, is not an imf; it is a trend [20].

according to some tolerances or stop criterion. Until this is achieved effective IMF is achieved and corresponding slow oscillations residual is computed [29].

Figure 2.2 illustrates the decomposition of wind signal via EMD. It is been shown that number of extremes are decreasing as going from one IMF to next IMF. Last one is not an IMF so is called as trend or residual. In figure 3.2, each column, signals in bottom row are obtained as superposition of waveforms plotted in the middle and top rows.

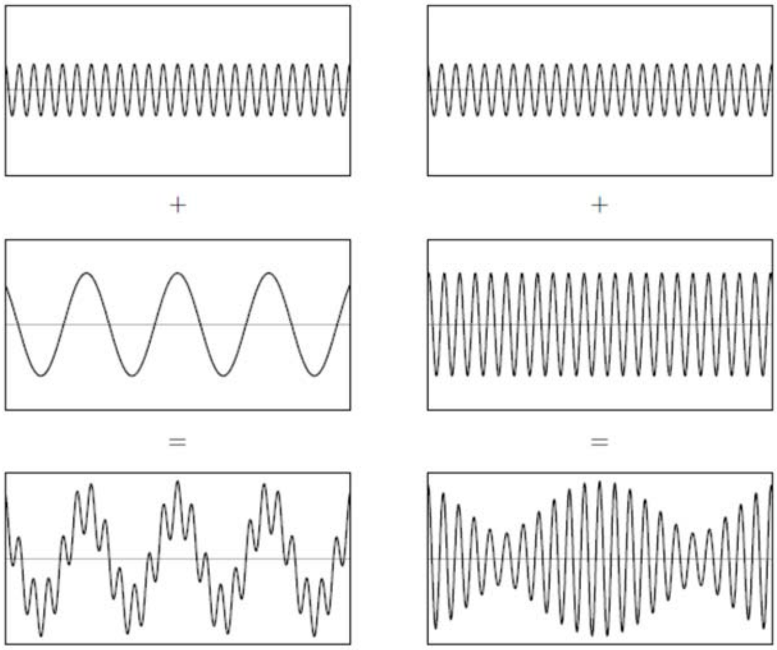


Figure 2.3 : EMD Resolution- Beat Effect [21]

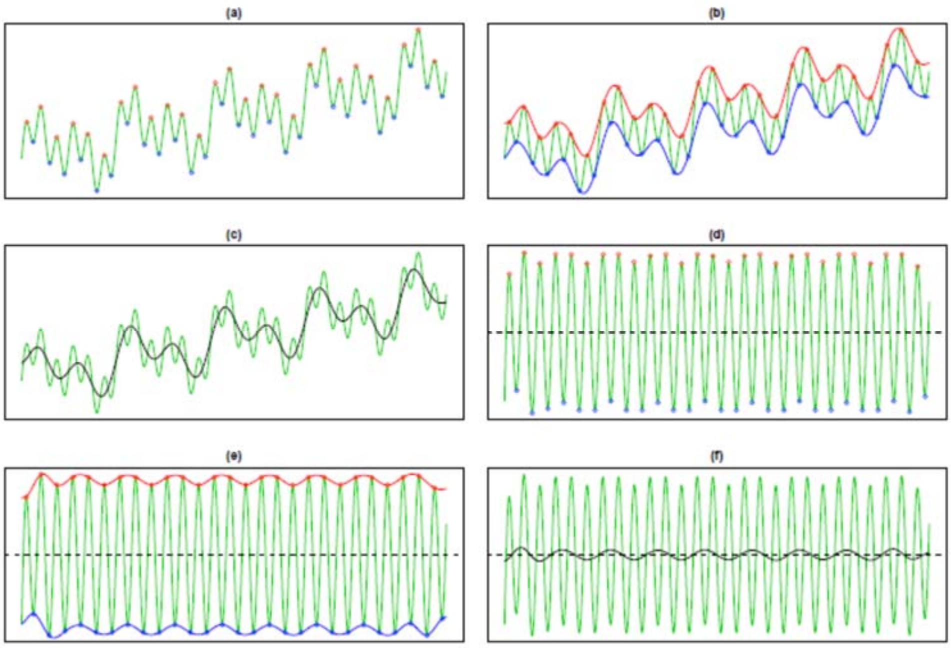


Figure 2.4 : EMD Sifting Procedure [21]

When frequencies of the two superimposed tones are sufficiently far apart (left column), two-tone interpretation is meaningful whereas, when they get closer (right column), interpretation in terms of single tone is favourable. Figure 3.3 illustrates the whole process of sifting procedure used in EMD implementation.

Summarizing, EMD can be summarized by the motto “fast oscillations superimposed to slow oscillations”. However, decomposition is carried out on local oscillation basis. Local oscillations are obtained by interpolating local maxima and local minima.

2.1.6 OTHER MISCELLANEOUS METHODS

Other than above discussed methods, also some miscellaneous methods exist. Example of these methods are- least square estimation of the trend, differencing to generate stationary data, and smoothening by moving averaging.

All above methods are designed to modify global representation of Fourier analysis, but they all failed in one way or other. Having studied above methods, summarizing the following conditions for the basis to represent a non-linear, and non-stationary time series-

a) Complete

b) Orthogonal

c) Local

d) Adaptive

The first condition guarantees the degree of precision of the expansion; second condition guarantees positivity of energy and avoids leakage. These are standard requirements defined for all linear expansion methods. For non-stationary signals orthogonality condition needs to be modified. Requirement for adaptivity is crucial for both non-linear and non-stationary data, for only adapting to the local variations of the data for describing the physics of the process and not just to fulfil the mathematical requirements for fitting the data. It’s been used in Audio watermarking schemes [20].

2.2EXISTING AUDIO WATERMARKING SCHEMES

Many researchers have modified, enhanced the previous researches with mathematical tools. It can be done in spatial domain, frequency domain and time-frequency domain.

2.2.1 TIME DOMAIN

Watermark is distributed directly into amplitude on time basis. Intensity of watermark is kept low for imperceptibility and making it less robust.

Low Frequency Amplitude Modifications- Very basic idea to embed watermark retaining every condition is to embed data in audio signal information region. As audio information is present in low frequency region, so watermark is embedded in low frequency region too.

A new method exploits differential average-of absolute-amplitude relations within each group of audio samples to represent one-bit information, is proposed by Wen-Nung Lie *et. al* [16].

Time Spread Echo- Other idea is use other audio signal as watermark. So that audio frequencies mix with each other in a way that no one can separate them. Echo is spread to reduce its energy and thus make inaudible as compare to single echo and its embedding algorithm is signal adaptive, by Byeong-Seob Ko *et. al* [12]

Adaptive psychoacoustic Masking- Audio will be listened by a human and evaluated by same being. So why not to check human auditory characteristic and use them for choosing watermark embedding location. A blind watermarking scheme in which watermark generated is masked according to Human Auditory System, is proposed by Herkiloglu K. *et. al* [6].

2.2.2 FREQUENCY DOMAIN

Researches in frequency domain are being carried out taking different transforms. In this case watermark is distributed in selected frequency/frequency-range. If it is once embedded, it is hard to be deleted. Frequency domain approaches are more robust than time domain approaches. Because it is more noticeable when signal sample values are changed, than changing magnitude value of a particular/range of frequencies. So it is better to choose particular frequencies and their magnitude values, encode them, replace them or simply embed watermark in these values.

The followings present the related researches. First Transform came up in this is Fourier Transform. Most of researches chose particular frequency or frequency range as watermark location. This frequency is chosen as low frequency to increase robustness.

FFT and HAS – FFT of signal is computed and Watermark is embedded in phase coefficient as per Human Auditory System, is proposed by Xiumei Wen *et. al* [28].

FFT and QIM - Watermark is embedded in frequency domain by Quantization Index Modulation, by Nima Khademi *et. al* [8].

Non-Uniform DFT – Signal frequency domain is evaluated by computing DFT. Watermark embedding frequency points are selected by a secret key, by Ling Xie *et. al* [30].

Quadratic form of watermark & FFT - Quadratic form taken as key matrix and watermark data is first encrypted using key matrix. Encrypted data is embedded in FFT coefficient of audio signal by Dr. A. Chandra Sekhar *et. al* [21].

DCT is popular in image processing, also it plays wide role in audio watermarking. DCT of blocks of audio signal is compaction of energy and watermark is embedded in these coefficients. Inverse DCT will spread the whole energy again and thus watermark is distributed.

Non Uniform DCT - Watermark embedding in chaos selected frequency coefficient of DCT, by Huang Xiong-Hua *et. al* [7].

DCT & Vector Quantization – Blocked DCT is computed and Vector Quantization of middle frequency (MF) coefficients according to watermark bits is performed, by Ji-Xiu Liu *et.al* [18].

DCT & Neural Network & Hamming Code – Neural Network Trained for searching MF positions for embedding watermark. This made it adaptable to signal. Watermark is coded via Hamming Code & Random Sequence, by Charfeddine Maha *et. al* [19]

Considering failure of Fourier as well Cosine Transform to represent non-stationary signal, new multi-resolution Transform evolved named Wavelets. It decomposes the signal into sub-bands of user defined resolution. Thus one can model audio signal and choose watermark embedding location in much optimum way.

Self-Synchronized DWT Watermarking – DWT of audio signal Synchronization codes and watermark bits are embedded in low frequency components obtained from DWT, by Shaoquan Wu *et. al* [29].

DWT, DCT & HAS – Multi-resolution capability of DWT and compact energy by DCT; Watermark is embedded using Adaptive Quantization according to Human Auditory Masking, proposed by Xiang-Yang Wang *et. al* [27].

DWT & Mean Quantization Principle – Blocked DWT is computed. Embedding watermark bits in wavelets coefficients via Mean Quantization Principle, by Wang Lanxun *et. al* [14].

DWT & Group Amplitude Quantization based on Watermark Cost Function – Equation is proposed connecting Watermark Cost Function and Group Amplitude Quantization Legrange Principle is used to derive optimisation solution. Optimised results are then applied to embed the watermark, by S.-T. Chen *et. al* [4].

DWT & Characteristic Curve of Energy Proportion Function (CCEP) – Characteristic Curve of Energy- Proportion Function is obtained. Based on CCEP and properties of energy-proportion function, watermark is embedded. It is blind in nature; proposed by S.-T. Chen *et. al* [3].

2.2.3 TIME FREQUENCY DOMAIN

A band of frequency is chosen and then distribute watermark in it. All kernel based transforms used in frequency domain are not signal adaptive. Although wavelet possess multi-resolution capability but filters defining its resolution are fixed. So once resolution in decided it will not change for the whole signal. EMD is signal adaptive process, decomposing signal into Intrinsic Mode Functions (IMF) – band of frequencies. Watermark is embedded in selected IMF.

Empirical Mode Decomposition (EMD) – EMD of audio frames is evaluated. Watermark embedding in extremes of highest energy IMF; Imperceptible & Robust in nature; proposed by A N K Zaman *et. al* [32].

EMD & QIM – EMD of audio frames is evaluated and watermark embedding via QIM in last IMF extremes is done; by Kais Khaldi *et. al* [9].

Spread Spectrum Method – Basic approach to watermark, Watermark is embedded in Direct Spread Spectra of audio, by Darko Kirovski *et. al* [11]. Also, further improvement is done in this algorithm [12].

Time-frequency operations are comparatively more robust and imperceptible than frequency domain and time domain approaches.

2.3EXISTING EMBODIMENTS

Once a method is defined, results are also analysed by changing its parameters. Hence, method along with its varying parameters is fully defined for a particular application. Such variations developed in audio watermarking era is discussed as follows headed by parameter into consideration-

2.3.1 HUMAN AUDITORY SYSTEM CHARACTERISTICS

HAS is used to mask and embed watermark in appropriate manner, so as to insure inaudibility. Firstly it was implemented along with Fourier Transform [28] and was checked for improvement achieved. Charfeddine Maha *et. al* implemented a technique in which neural network is trained to find out middle frequency positions of DCT coefficients for watermark embedding [33]. Also, to enhance robustness, watermark is coded via Hamming code and random sequence. He continued the same technique by just changing transform from DCT to Wavelet.

2.3.2 HIERARCHY OF TRANSFORMS

It is one of the methods researches opt for. Very simple is a double DCT transform. It was implemented to form a hierarchy of transforms and was evaluated [31]. It outperforms the simplest approach. Xiang-Yang Wang *et. al* suggested a hybrid way comprising DWT and DCT transforms [27]. It was better as it possess multi-resolution capability of DWT and energy compaction via DCT transform. Chundong Wang *et. al* however checked results for method which used DWT along with Fourier Transform [25].

2.3.3 WATERMARK EMBEDDING ALGORITHM

Instead of searching for best location for watermark embedding, researchers were motivated to propose new algorithm for watermark embedding. It resulted in Quantization Index Modulation (QIM) proposed Brian Chain *et. al* [2]. He suggested QIM to be host signal rejection method and non-linear in nature. In contrast, Spread Spectrum method is linear and host signal non-rejecting method. Researches are carried out using QIM along with different transforms and results are analysed. Results obtained using QIM proved better than the watermarking using only transforms. For example, QIM along with FT is implemented [8], Gaorong Zeng *et. al* [33] implemented QIM along with DCT, Kais Khaldi *et. al* [9] done watermarking via Empirical Mode Decomposition and QIM. Their results outperforms the previous researches when compared.

2.3.4 MATHEMATICAL TOOLS

It was advisable to resolve the signal to more extent, using some other mathematical tool in conjunction with transform. Ji-Xiu Liu *et. al* suggested to decomposition of DCT coefficients in vector space using vector quantization [18]. It represented the middle frequency coefficients in vector space. Most popular mathematical tool is Singular Value Decomposition (SVD). Gulivindala Suresh *et. al* proved DCT in conjunction with SVD to be better than previous researches based on simple DCT transform [13]. Same was implemented on DWT transform along with SVD on lower frequencies in [25] and it gave better results too. It is because SVD has several advantages- like changing eigen values will change the matrix value by very negligible amount. Secondly, it can be applied on any sized matrix. So SVD explores the coefficient into more detail, and altering its value will not affect sample values. Thus improves imperceptibility as well robustness. QIM and SVD are discussed in more detail because of their importance and advantages-

2.3.4.1 Quantization Index Modulation

Important uses of data hiding schemes in digital media are to provide resistance to copyright, and reassurance of content integrity. Hence, the data should stay embedded in host signal, even if that signal is subjected to attacks/manipulation degrading its quality, for example- filtering, cropping, resampling or lossy data compression. Thus, the techniques used for embedding the hidden data vary depending on the quantity of data being hidden and the required persistence of those data to attacks/manipulation. Since no one method can achieve all described goals by standing alone, a class of processes/algorithms is needed to cover the range of possible applications.

The challenges of data hiding are technically forbidden. Any “holes” to fill with secret data in host signal, either perceptual or statistical, are likely ultimate goal for removal by lossy signal compression. The important key to hiding data successfully, is the finding of the holes that are not suitable for misuse by attacks or other compression algorithms. Further challenge is to fill these holes with secret data in a way that remains same to a large class of attacks/ manipulations/ host signal transformations. [2]

Host Signal Interference Non-Rejecting Methods

A large number of data embedding algorithms are designed based on the principle that the host signal is similar to a source of interference or noise. This view arises when one ignores the fact that the encoder has access to, and hence can use knowledge of, host signal x. The simplest of this class have embedding functions in which it is simply added,

$$s(x,n) = x \mp w(n) \tag{10}$$

where $w\ (n)$ is typically a pseudo-noise sequence. Such kind of embedding algorithms are referred to as Spread Spectrum Method. (“Patchwork” algorithm of Bender, comprises adding small amount $\pm$ to some pseudo-randomly chosen host signal sample values, subtracting a small amount $\pm$ from other samples [2].

As this method is based on adding or subtracting a little value, thus is equivalent to adding a pseudo-random sequence w (n) of $\delta \mp$ to the host signal. Hence, Patchwork algorithm to be as spread spectrum method is studied.)

One can observe that for SS kind of additive methods, can inhibit the ability of decoder to estimate n. Also, even in case of less severe conditions, one can embed lesser amount of data.

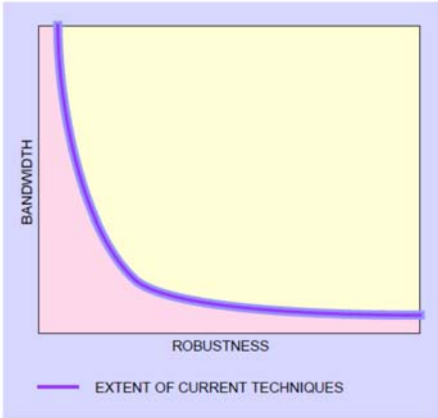


Figure 2.5 : Conceptual data hiding problem space [2]

Thus, these methods are useful when either host signal is available at the decoder or host signal interference is smaller than the channel interference. Problem of host signal interference is very common in linear schemes, is shown in fig 2.3.

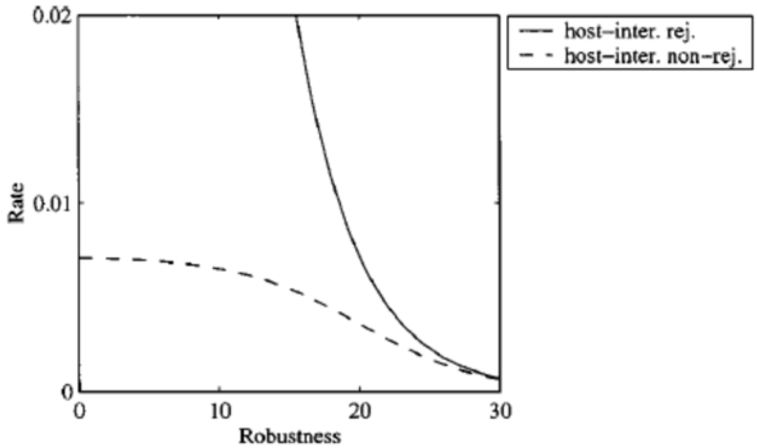


Figure 2.6 : Qualitative behaviour of host-signal rejecting and non-rejecting embedding methods [2]

Host Signal Interference Rejecting Methods

It is observed that one of the inherent limitations of embedding methods is that they do not reject host-interference rejecting methods.

This QIM class of embedding methods reveals the type of behaviour shown by the solid line in Fig. 2.6, while providing enough reliability for a design to hold trade-off among rate, robustness, and distortion, means to move from one point to the other on the solid curve of Fig. 2.6.

One can view $s(x, n)$ as an ensemble of functions of variable x, each function in ensemble indexed by n. Denote functions in ensemble as $s(x; n)$ to highlight this view. If the distortion induced by embedding is to be small, then each ensemble function must be close to identity function in some sense so that if it assumed all of these identity functions are quantizers, then the embedding method is referred to as QIM method. Thus, quantization index modulation refers to embedding informative data by first modulating index or sequence of the indices according to embedded information and then quantization of the host signal with the associated sequence of / a quantizers. Figure 2.7 illustrates QIM in the case where just one bit is embedded. Thus, two quantizers (for 0 or 1) are required, and their corresponding sets of reconstruction points in $< \ N$. If $\ n\ D\ 1$, for example, host signal is quantized with A- quantizer, i.e., s is chosen to be the A closest to x. If $n\ D\ 2$, x is quantized with the B-quantizer. The sets of reconstruction points are all different, thus offers great host signal rejection property, shown in fig 2.7.

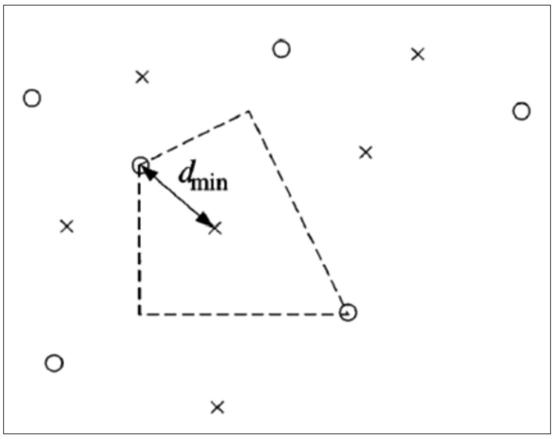


Figure 2.7: Quantization index modulation for information embedding [2]

It can be visualized, the design of QIM based systems as simultaneous design of an ensemble of quantizers (source codes) and defined signal constellations (channel codes). [2]

In linear algebra, the singular value decomposition (SVD) is a factorization/decomposition of a real or complex, square/rectangular matrix. It is being common in many useful applications in statistics and signal processing.

2.3.4.2 Singular Value Decomposition

Mathematically, Singular Value Decomposition of a matrix (say M) of size mxn, is of the form-

$$M = U \sum V^* \tag{11}$$

where U is a $m \times m$ real or complex upper half matrix and V is lower half matrix. $\sum$ is the diagonal matrix with non-negative real numbers. V* here represents the transform of V matrix, and is of size $n \times n$. Diagonal values of $\sum$ are called singular values of matrix M. The singular value decomposition and Eigen vectors or Eigen decomposition are closely related. Named as -

- The left singular vectors of matrix M are Eigen vectors of matrix MM^*.
- The right singular vectors of matrix M are Eigen vectors of matrix M^*M.
- The non-zero singular values of matrix M, diagonal values of matrix sigma, are square roots of the non-zero Eigen values of both MM^* and M^*M.

Its application include -

- Solving homogeneous linear equations
- Total least square minimization
- Range, null space and rank
- Low rank matrix approximation
- Separable models
- Nearest orthogonal matrix
- The Kabsch algorithm
- Signal Processing

Historical perspective of SVD

The singular value decomposition (SVD) was basically developed by differential geometers, who wanted to determine whether a real and bilinear form could be made equal to other by independent and orthogonal transformations of the two spaces it acts upon. Camille Jordan and Eugenio Beltrami discovered independently, in 1874 and 1873 respectively, that the singular values of the bilinear forms, represented as matrix, form complete set of invariant for bilinear forms when is under orthogonal substitutions. James Joseph Sylvester also started his work at the singular value decomposition for real and square matrices in 1889, independently of both Jordan and Beltrami. Sylvester called the singular values as the canonical multipliers of the matrix.

2.3.5 OTHER MISCELLANEOUS METHODS

Various methods were implemented by changing watermark embedding locations. Like in Fourier Transforms- Non- uniform DFT [7], Quadratic form of watermark and FFT [21] is used. Also, in case of DCT transform, instead of embedding watermark code in DCT coefficients, alternatives are proposed. For example Yan-Yuang *et. al* suggests the watermark embedding in AC-DCT coefficients [31] or in chaos selected frequency coefficient resulting DCT into non-uniform DCT [7]. Gulivindala Suresh *et. al* proposed other encoding scheme for watermark image, i.e. by using Arnold transform. It helped it in improving its robustness [13].

2.4PERFORMANCE ANALYSIS PARAMETERS

Every systems in analysed using various parameters defining its performance. Performance of watermark is checked on the basis of quality of three parameters being discussed. Those are- robustness, imperceptibility and pay-load. So following various calculations used to define quality of these parameters are considered-

2.4.1 IMPERCEPTIBILITY TEST

The imperceptibility test is performed by both subjective and objective means. Also, it can be measured by mean square error and signal to noise ratio. Low value of MSE and high value of SNR is required for imperceptible audio.

2.4.1.1 Subjective Listening Test

Subjective listening tests are essential because ultimate judgement is made by human perception. In subjective listening test five participants are provided with original and watermarked audio signal and are asked to report the dissimilarity among them as per five-point subjective grade (SG). Their average is chosen afterwards to get final SG. Mapping among value, SG and ODG is given in table 2.1.

Table 2.1: Subjective and objective difference grades

SG	ODG	Description of Impairments	Quality
5.0	0.0	Imperceptible	Excellent
4.0	-1.0	Perceptible, but not annoying	Good
3.0	-2.0	Slightly annoying	Fair
2.0	-3.0	Annoying	Poor
1.0	-4.0	Very annoying	Bad

2.4.1.2 Objective Listening Test

The ultimate goal of objective measurement algorithms is to substitute subjective listening tests by modelling human listening response. The objective measurement metric namely

Objective Difference Grade (ODG) does not always correlate with subjective listening test very well. However a final judgement of audio quality has to be based on subjective tests.

2.4.1.3 Mean Square Error

It is mean of the square of difference of original signal and watermarked audio signal [10].

$$MSE = \sum_{i=1}^{laudio} \frac{(audio - watermarked\ audio)^2}{Number\ of\ audio\ samples} \qquad (12)$$

Lower the value of Mean Square Error (MSE) better the imperceptibility factor.

2.4.1.4 Signal to Noise Ratio

Signal to Noise Ratio (SNR) of watermarked audio is computed using following formula [10],

$$SNR = 10 \times log_{10}\left(\frac{\sum_{j=1}^{laudio} X(j)^2}{\sum_{j=1}^{laudio}(X(j) - WX(j))^2}\right) \qquad (13)$$

2.4.2 ROBUSTNESS TEST

Robustness is important parameter when one deal with robust watermarks. It is carried out by attacking the watermarked signal, and then checking out the quality of watermark.

2.4.2.1 Bit Error Rate

It is the number of bits of watermark extracted which are not same as that of original watermark embedded in original signal. It is used after the signal being attacked [9].

$$BER\ (W, W^*) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} W(i,j) \oplus W^*(i,j)}{M \times N} \qquad (14)$$

where W is the original watermark and $W^*$is watermark extracted. $M \times N$ is the original watermark image size.

2.4.2.2 Normalized Correlation

Evaluation of similarity between original watermark and extracted watermark is given by Normalized correlation (NC) [9].

$$NC(W, W^*) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} W(i,j)W^*(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} W^2(i,j)}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} W^{*2}(i,j)}} \qquad (15)$$

Large value of NC indicate the watermark presence and smaller value indicates absence of watermark.

2.5COMPARATIVE ANALYSIS

Comparison among various approaches is important when to pursue in a particular subject. It summarizes all positives and negatives, at once, along with quantitative analysis & results. To have better understanding of ups-downs of all techniques, results are summarised in table 2.2.

It can be seen that performance of DCT and DWT is comparable, yet DWT is more popular because of its multi-resolution characteristic. As it is previously discussed there are signals those are beyond the eye of DWT, to analyse non-stationary signals. So new techniques like EMD are also been used as discussed.

Attack Type	DCT + DWT [15]		DCT		DWT		EMD [20]	
	BER %	NC	BER %	NC	BER %	NC	BER %	NC
Without attack	0	1.0	0	1.0	0	1.0	0	1.0
MP3 (32 kbps)	0	0.9304	9.86	1.0	11.82	1.0	0	1.0
AWGN(20 dB)	1.71	0.9700	6.18	0.8965	1.93	0.9662	0	1.0
Resampling (22.05 kHz)	0.61	1.0	0.93	0.6884	0.89	0.7067	1	0.9983
Filtering	1.09	0.9418	3.41	0.9683	0.07	0.9842	0	1.0
Cropping	0	1.0	0	1.0	0	1.0	0	1.0
Requantization	0	1.0	0	1.0	0	0.9987	0	1.0

Table 2.2 : Comparative analysis of different transforms

2.6 GAPS

Gaps are important part in research work, as they direct researchers to proceed their work. Firstly gaps come from transform evolution. As discussed in transform evolution section, every kernel based approach restricts the signal to be linear in nature. However this is not the case if one works in audio era. Audio signal is itself a non-linear / non-stationary signal. Secondly, when fragile audio watermark is concerned, one parameter is compensated i.e. robustness. So other parameters can be thought to improve. However in literature review no one tried to fix the value of robustness and improve imperceptibility or pay-load.

2.7 OBJECTIVES

Outlining of gaps clearly gives objectives, which can be thought to work upon. Following the objectives deduced from gaps are discussed-

2.7.1 AUDIO SIGNAL PROCESSING VIA EMD

Transform evolution resulted in wavelet transform in the end. Though wavelet offers a variety of basis function to analyse a given signal, yet user is allowed to choose from those fixed basis functions. Also as discussed before, every chosen scaling/dilation factor does not represent every transition in signal. So a new approach is required which is adaptive to every signal transition and can represent that accurately. One such approach is discussed and that is empirical mode decomposition and Hilbert transform already discussed in analysis of non-stationary signals. So EMD can be thought to decompose audio signal and watermark it, to get possible improvisation.

2.7.2 INCREASE IMPERCEPTIBILITY AND/OR PAY-LOAD IN FRAGILE AUDIO WATERMARK

There are three conflicting parameters of watermark- robustness, imperceptibility and pay-load. In fragile audio watermarking, robustness can be kept at fixed level. This will improve the quality of watermark as fragility doesn't require watermark to be robust.

2.7.3 MATHEMATICAL TOOLS

Any method implemented can give better results by using mathematical tools in conjunction with that method. It is been noticed in literature review that there are fewer schemes which used QIM and SVD in conjunction. Particularly QIM and SVD are said because of their potentially good advantages.

2.8 PROBLEM FORMULATION

Very first objective is considered and other two objectives are achieved in implementing first objective. So, in proposed work Empirical Mode Decomposition is used to decompose the audio signal and to watermark it. However, EMD too possess some drawbacks when implemented. EMD causes a reconstruction error though negligible of order 10^{-17}. This creates problem when again decomposition of the reconstructed signal is done, the obtained IMFs are not same as which earlier got by decomposing the original signal. So if IMFs obtained are not same even if no watermark is added, then how can one add watermark and extract it? So, it is a big issue when one work with EMD in watermarking era. One cannot extract the original watermark until one get the similar IMF. So it needed a way to mask or avoid erroneous methods which use EMD, retaining quality parameters.

Three parameters- robustness, imperceptibility and pay-load, are in confliction. One need to maintain a trade-off among them. In fragile watermarks, robustness is not an issue. So other two parameters can be thought to improve. Idea in background of proposed method is, to enhance imperceptibility to its maximum.

So a method is proposed in which problems of EMD are avoided, along with advantageousness of EMD (Objective 2.7.1). Approach of this method is to maximize the imperceptibility factor (Objective 2.7.3). QIM and SVD are used to enhance obtained results.

CHAPTER

3

PROPOSED AUDIO WATERMARKING ALGORITHM

As discussed in previous chapter, there has been several watermarking algorithms for improving robustness, fragility and etc. as per the application going to use them. Spread Spectrum remained in literature review for long time, until QIM was invented. QIM is in contrast non-linear and host signal non-rejecting method.

IMFs are approximately orthogonal to each other. Audio information lies in low frequency regions and so are better locations for being robust. EMD is also been used in few previous researches like in [5]-[7]. In [5], IMF having highest energy is chosen to embed watermark and reason is not known. In [6], watermark bits are embedded in mean trend obtained after decomposition of audio segments. There is domain transform being used along with EMD in both [5] and [6]. IMFs can be reconstructed via their local maxima and minima. Using this in [7], watermark is embedded in extremes of last IMF and then modified extremes can be interpolated to form watermarked IMF. This may be erroneous because EMD is very sensitive to numerical errors. While interpolating new extremes, there may exist a slight slope change, leading to slightest change in frequency. This slightest change can alter IMFs obtained after decomposing new reconstructed audio signal. Example- number of extreme points per IMF can vary and in worst cases, number of IMFs obtained can vary. If one force to fix number of IMFs to be same, resolution needs to be signal adaptive. To avoid this one need to know the method to detect erroneous extreme points and other possible erroneous cases, and then mask/ avoid such cases, is not discussed.

3.1 BASIC IDEA

In proposed method audio signal is segmented and decomposed by EMD. Audio information is mainly present in lower frequencies. So in proposed method the mean of last IMF is chosen. Taking mean avoids any error due to frequency change or any attackable position. It also makes audio more imperceptible and increases robustness. Further, to spread the watermark firstly decompose the matrix, formed by mean of last IMFs of each audio segment, into singular values (SVs) and then embed watermark bits into singular values. Watermark embedding is done via Quantization Index Modulation due to its good robustness and blind nature. Parameters of QIM is chosen to assure inaudibility. Experimental results show that this algorithm is imperceptible even in worst conditions.

3.2 MATHEMATICAL FORMULATION

The idea of proposed method is to hide and spread the watermark in time domain in a peculiar way. Audio signal is segmented into frames and EMD is conducted on every frame to extract associated IMFs. Take mean of consecutive last IMFs, represent it into a matrix. Decompose the matrix via Singular Value Decomposition (SVD) for singular values.

Let $X = \{ x(i), 1 \leq i \leq L\}$ represent the host audio signal of length L samples. $W = \{w(i,j), 1 \leq i \leq M, 1 \leq j \leq M\}$ is binary image, fig 4.1, to be embedded within the host audio signal, and $W(i,j) \in (0,1)$ is the pixel value at (i,j). Make $W(i,j)$ in form of 1-D as $wc(i) \in (0,1), 1 \leq i \leq M \times M$. Choose a watermark enhancement factor say S experimentally to achieve maximum imperceptibility and robustness.

3.2.1 WATERMARK EMBEDDING ALGORITHM

The block diagram of watermark embedding in proposed scheme is shown in figure 4.2. The main steps comprising in this algorithm are described below.

Step 1 The audio signal is partitioned in non-overlapping 1-D frames F_j , $j = 1, 2, 3, \ldots , N$. Here N is the number of frames.

Step 2 Take EMD of individual frame and extract mean value of last IMF.

$$EMD(F_j) = \sum_{i=1}^{limf} IMF_i + residue \tag{16}$$

where, $limf$ is number of IMFs obtained. Reconstruction error, though negligible, may exist here as defined below:

$$error = F_j - \left(\sum_{i=1}^{limf} IMFs + residue\right) \tag{17}$$

Represent mean of last IMFs in 2-D matrix format $P(i,j), i = 1, 2, 3, \ldots, M, j = 1, 2, 3, \ldots, M$. Make sure $M \times M = N$.

Step 3 SVD is applied on this matrix and singular values are obtained. Let $\lambda = (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_u)$ be SVs of block matrix $P(i,j)$.

Step 4 Embed watermark bits obtained from $wc(i) \in (0,1)$, into SVs obtained above as follows by QIM:

$$\lambda_i^* = \begin{cases} \lfloor \lambda_i/S \rfloor * S + sgn(3S/4) & if\ wc(i) = 1 \\ \lfloor \lambda_i/S \rfloor * S + sgn(S/4) & if\ wc(i) = 0 \end{cases} \tag{18}$$

Step 5 New matrix can be obtained by taking inverse SVD using modified SVs. Say it as $P^*(i,j)$. Represent it into 1-D format as $P^*{}_i, 1 \le i \le M \times M$.

Reconstruct watermarked frames taking inverse EMD as follows:

$$F_j^* = \sum_{j=1}^{limf} IMFs_j + residue + error + P_j^* - P_j \tag{19}$$

where, P_j represents 1-D matrix of original means of last IMFs, P_j^* represents 1-D format for modified means.

Step 6 Concatenate consecutive frames and watermarked audio signal $(X^*)$is obtained. Watermarked signal and audio signal are shown in figure 4.4.

3.2.2 WATERMARK EXTRACTION ALGORITHM

The block diagram of watermark extraction procedure of proposed scheme is shown in Figure 4.3. The main steps of extraction algorithm are as follows:

Step 1 The watermarked audio signal (X^*) is segmented into frames $RF_j, j = 1, 2, 3, \dots, N$

Step 2 Decompose each of frame F_j^* using EMD, extract last IMF and calculate mean. Represent means of consecutive last IMFs in 2-D matrix form. $RP(i,j), i = 1, 2, 3, \dots, M, j = 1, 2, 3, \dots, M$

Step 3 Apply SVD on matrix $RP(i,j)$, obtain SVs $(\lambda^*)$and extract watermark code bits (ewc_i^*) as follows:

$$ewc_i^* = \begin{cases} 1 & if\ \lambda_i{}^* - \lfloor \lambda_i{}^*/S \rfloor * S \ \ge \ S/2 \\ 0 & if\ \lambda_i{}^* - \lfloor \lambda_i{}^*/S \rfloor * S < \ S/2 \end{cases} \tag{20}$$

where $rc_i^*$is received/extracted code. Decode it into 2-D image$EW(i,j)$.

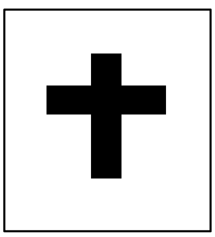


Figure 3.1 : Original Watermark Image

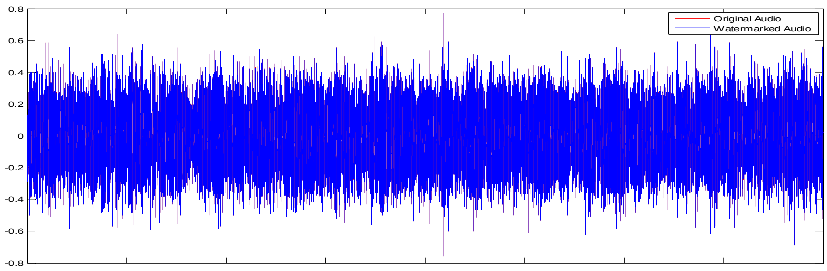


Figure 3.2 : Original and Watermarked Audio Signal

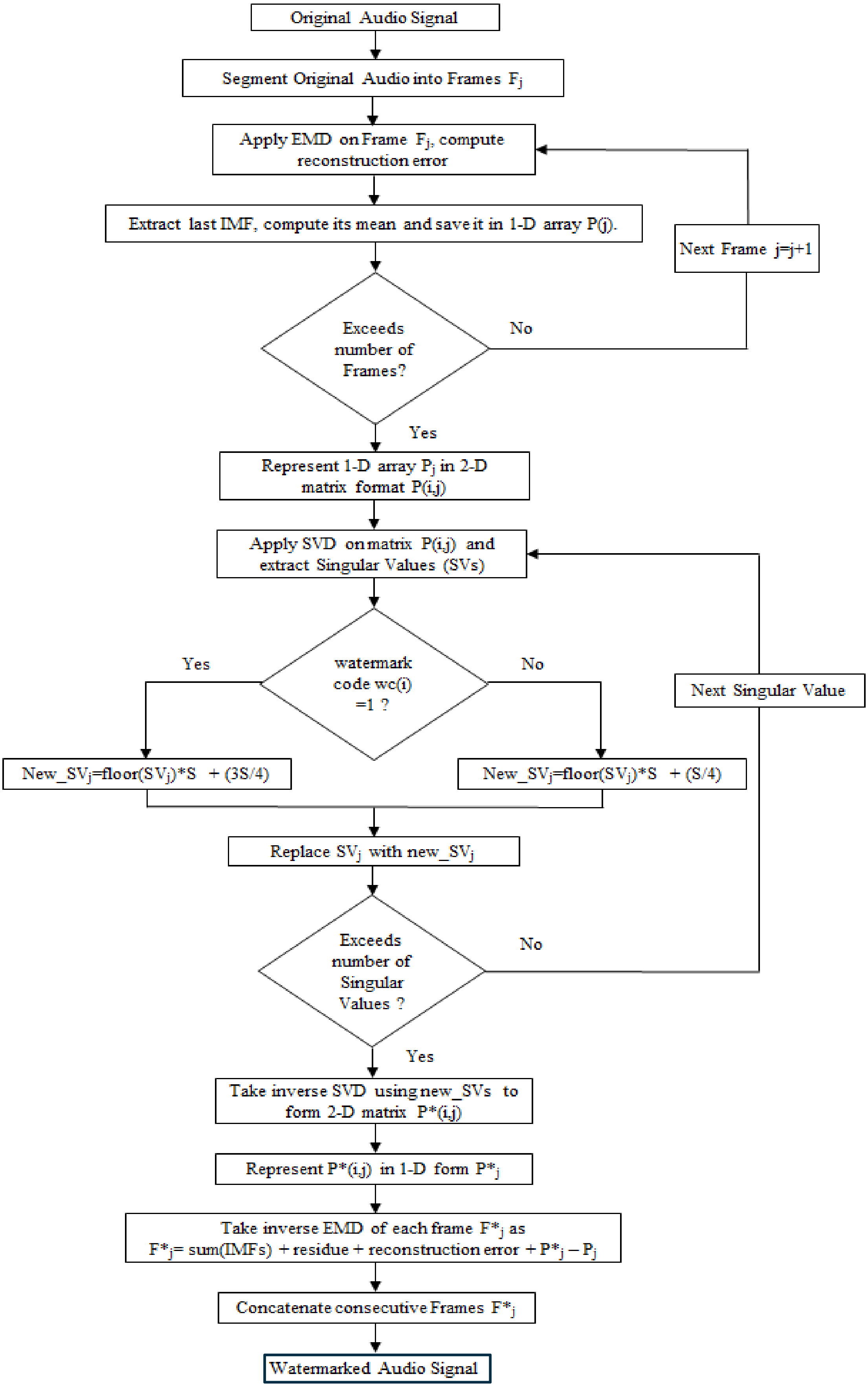


Figure 3.3 : Watermark Embedding Algorithm

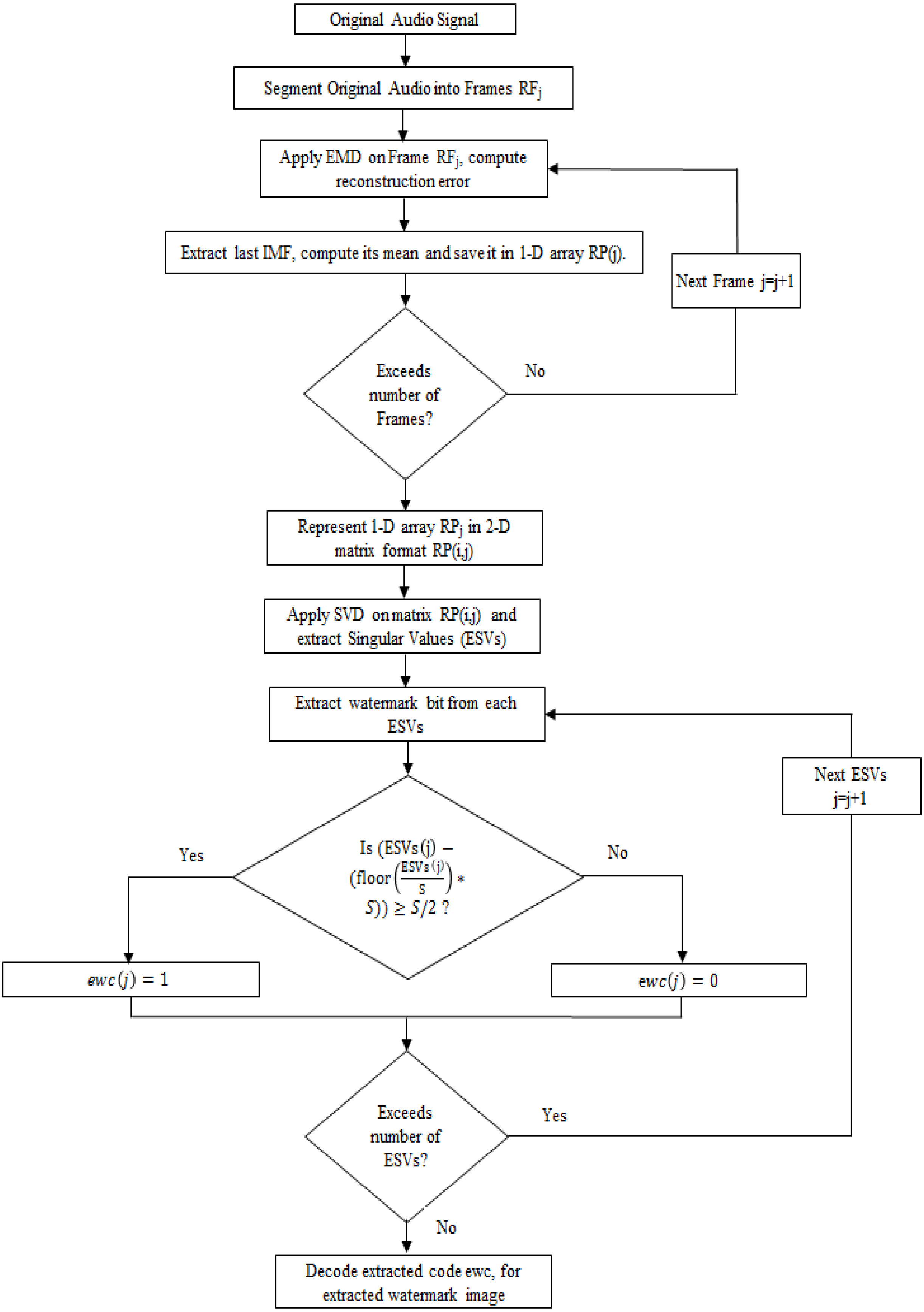


Figure 3.4 : Watermark extraction algorithm

CHAPTER 4

RESULTS & DISCUSSIONS

Simulation code is written in MATLAB language and is applied to pop music comprising speech notes, classical notes as well. EMD is applied to the said signal after segmentation. Various performance analysis parameters are discussed along with their value obtained after executing the code. Firstly, imperceptibility is checked and both subjective and objective tests are performed. To check the audio quality, mean square error is calculated. Fragile watermark is generally used for temper detection and thus temper locations are calculated. This detection of temper locations is performed even after content persistent non-malicious attacks. Experimental results proved that audio quality is retained to maximum, and can tell tempering locations even after non-malicious attacks.

4.1 IMPERCEPTIBILITY TEST

The imperceptibility test is performed by both subjective and objective means. Also, it is measured by mean square error and signal to noise ratio experimentally. Low value of MSE and high value of SNR is required for imperceptible audio.

4.1.1 SUBJECTIVE LISTENING TEST

In subjective listening test five participants are provided with original and watermarked audio signal and are asked to report the dissimilarity among them as per five-point subjective grade (SG). Their average is chosen afterwards to get final SG.

4.1.2 OBJECTIVE LISTENING TEST

The objective measurement metric namely Objective Difference Grade (ODG) does not always correlate with subjective listening test very well.

Audio File	Average SG	ODG
Blues	4.9	-0.023
Pop	5.0	-0.034
Country	4.9	-0.057
Classic	4.8	-0.044
Jazz	5.0	-0.018

Table 4.1 : Average SG and ODG scores for different audio signals

ODG is output obtained from Perceptual Evaluation of Audio Quality (PEAQ) measurement algorithm specified in ITU-R BS.1387 (International Telecommunication Union-Radio-communication Sector). It corresponds to subjective grades obtained in subjective listening tests. The ODG ranges from 0.0 to -4.0 corresponding to imperceptible to very annoying. On an average, ODG values are in lesser than -1.0, hence can be remarked as imperceptible. Results of subjective and objective tests are in table 4.1

4.1.3 MEAN SQUARE ERROR

It's been checked for proposed scheme and came out to be 1.1694×10^{-9}. It is more imperceptible than other techniques reported recently, as in other algorithms it is approx. up to 10^{-5}.

4.1.4 SIGNAL TO NOISE RATIO

Signal to Noise Ratio (SNR) of watermarked audio is computed using above mentioned formula. SNR is calculated for proposed scheme and came out to be 68.98 dB. It is comparatively highest SNR than other techniques reported recently, shown in table 4.3.

4.2 PERFORMANCE UNDER MALICIOUS ATTACKS

Performance of watermark to attacks is checked by BER and NC. In order to check impact of attacks on imperceptibility, SNR is also measured after every attack. Different malicious attacks are performed:

4.2.1 ADDITIVE WHITE GAUSSIAN NOISE

AWGN attack is defined by simply adding random sequence, scaled appropriately, to signal samples. It is added in watermarked audio signal till SNR of signal left is 20dB.

4.2.2 RESAMPLING

Frequency of watermarked audio signal is 44.1 KHz and is sampled at 11.2 KHz. Again it is sampled to 44.1 KHz. Then different parameters are measured.

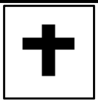
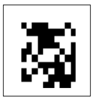
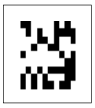
Attack Type	Normalized Correlation (NC)	Bit Error Rate (BER(%))	Signal to Noise Ratio (SNR (dB))	Extracted Watermark
No Attack	1	0	68.9829	
AWGN	0.4077	51	38.5610	
Resampling	0.3819	48	32.5631	
MP3 Compression	0.2689	50	28.5607	

Table 4.2 : Performance to malicious content tempering attacks

4.2.3 MP3 COMPRESSION

Compression is performed on watermarked audio, bit rate is 64 bps and audio file is converted to mp3 format. Again, obtained mp3 file is converted back to .wav format for analysis and measurement of parameters.

From table 4.2 one can observe, imperceptibility factor remains above 25 dB which is acceptable as per recommendations.

4.3 TEMPER LOCALIZATION TEST

In order to check temper location ability, some malicious tempers on watermarked audio signal are performed. Two attacks discussed are defined as:

Attack type 1 represents randomly deleting samples from audio. Here for simplicity, put the values of deleted samples equal to 0. The temper attack type 2 represents, randomly pick up 500 samples and then replace them with another audio signal. This is repeated 5 times on different samples to attack whole signal. Figure 5.1 (a) shows the maliciously tempered watermarked audio signal under temper attack type 1, and samples taken in number are 5,000 spread over whole 16,000,000 sample of audio. Figure 5.1(b) represents audio signal attacked by temper attack type 2, and samples from 1870 to 2370 are replaced by samples 2000 to 2500 of watermarked audio signal, 19,860 to 20,360 by 9,500 to 10,000 of watermarked audio samples, 78,800 to 79,300 by 15,000,000 to 15,100,500 of watermarked audio, 1,000,000 to 1,000,500 by 12,000,000 to 12,000,500 of watermarked audio and 15,800,000 to 16,300,000 by 1,200 to 1,700 of watermarked audio signal. Figure 5.2(a) and 5.2(b) show results for temper detection. A function $H\ (j)$ is defined as follows:

$$H\ (j) = \begin{cases} 1, & if\ audio\ frame\ is\ maliciously\ tempered \\ 0, & if\ audio\ frame\ is\ not\ maliciously\ tempered \end{cases} \tag{21}$$

Figure 5.2(a) and 5.2(b) show that proposed scheme has excellent ability of temper location.

4.4 TOLERANCE AGAINST NON-MALICIOUS ATTACKS

Practically, watermarked audio signal undergoes many content persistent non-malicious attacks, it is essential for watermark to retain audio quality. That is to say, apart from keeping a hold on audio quality, watermarking authentication technique should be capable to locate temper location accurately. Also, this is to inhibit the case if one tempers maliciously first, and then knowingly try to randomize the tempering locations to keep one's strategy. To check, non-malicious attacks or common signal processing is performed on above signals attacked by attack type 1 and 2. Signal in Figure 4.1(a) is subjected low pass filtering operation, cut off

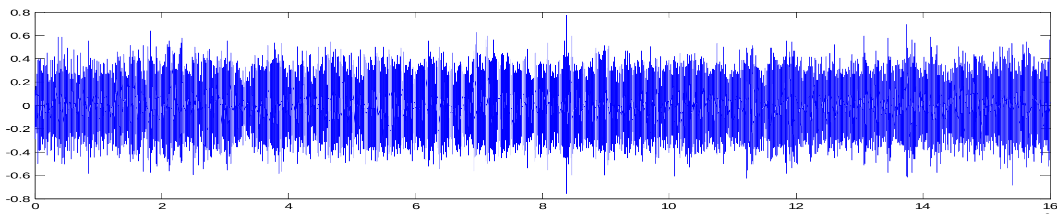


Figure 4.1(a) : Audio waveform after being attacked by attack type 1

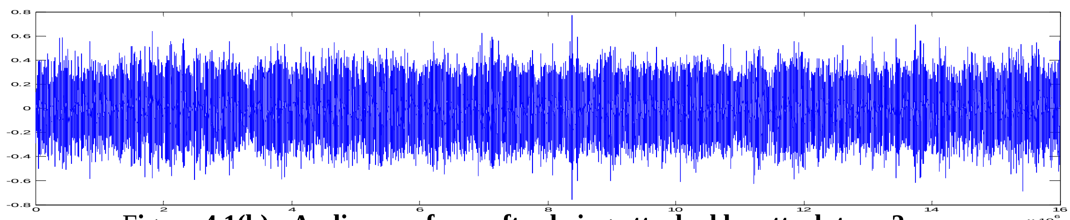


Figure 4.1(b) : Audio waveform after being attacked by attack type 2

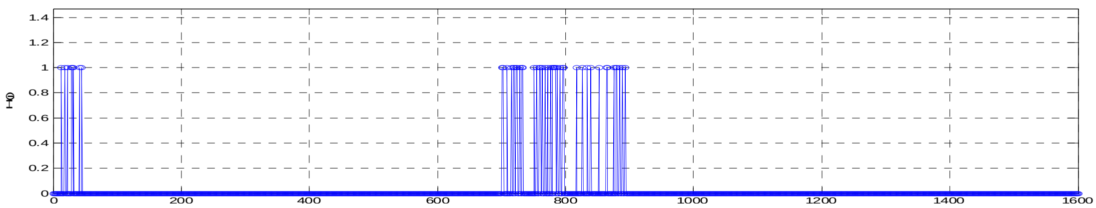


Figure 4.2(a) : Temper localization results of temper attack type 1

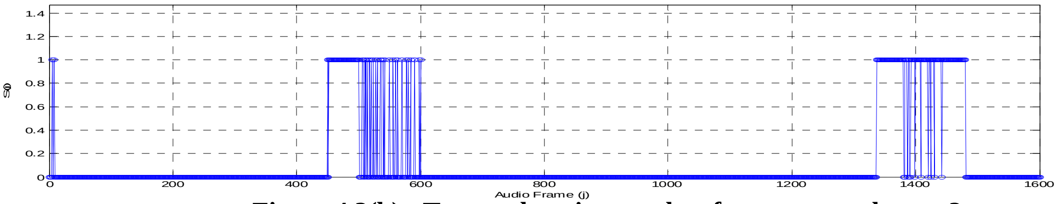
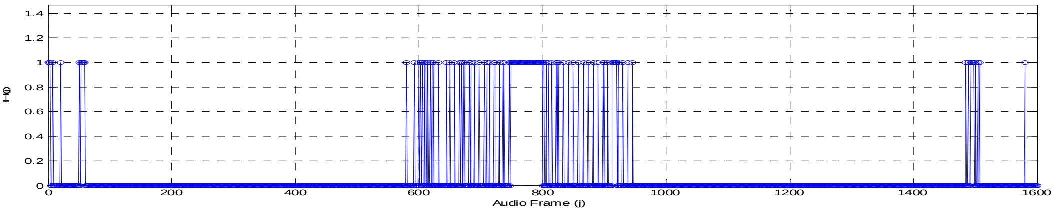
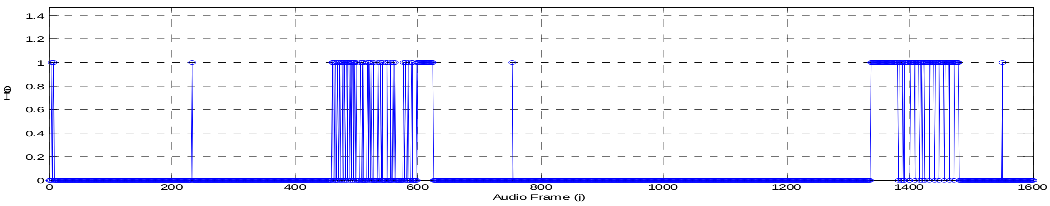


Figure 4.2(b) : Temper location results of temper attack type 2

frequency is 10 KHz and is checked for tempering locations. Result is shown in Figure 4.3(a). Signal in Figure 4.1(b) is subjected to MP3 compression taken at 64 bps, and result for tempering location after attack is shown in Figure 4.3(b). All these effects are done on



(a)



(b)

Figure 4.3 : Temper Detection Results (under non-malicious tempers and common signalling processing)- a) Temper location of performing low pass filtering of the audio signal in Figure 5.1(a); b) Temper location of performing MP3 compression to 64 bps of the audio signal in Figure 5.1(b)

Audio Editor Software (AVS) available at [12]. Figure 4.3(a) and 4.3(b) represents that proposed watermark authentication technique can still tell tempering locations, tolerating attacks. However, as there are some discontinuities in detected erroneous audio frames, one can conclude that it tolerates up to a certain degree of content persistent non-malicious attacks or common signal processing operations.

4.5 COMAPARATIVE ANALYSIS

SNR for various methods have been calculated and compared with proposed watermarking algorithm. It proved that proposed scheme is better than others in SNR. Difference in their value shows the improvement attained. Table 4.3 proves this statement.

Khaldi proposed audio watermarking using EMD and QIM, and gave better results in robust watermark. As far its imperceptibility is concerned, it is around 25-27 dB in SNR. In proposed algorithm, EMD and QIM are used, along with SVD and watermark is embedded

Methods	Algorithm	SNR (dB)
Proposed Method	EMD, SVD and QIM	68.98
Khaldi (2013)	EMD and QIM	25.67
Nematollahi (2012)	Multiple Wavelets and SVD	23.03
Uludaag (2001)	DC – Level Shifting	21.24
Bender (1996)	Echo	21.47
Bender (1996)	Phase	12.20
Bender (1996)	LSB	67.91
Cox (1997)	Spread Spectrum	28.59
Swanson (1998)	Frequency Masking	12.87

Table 4.3 : Comparison of different schemes with proposed scheme

in a way so that erroneous ways are avoided. Thus proposed algorithm gives better results as well is adaptable for a variety of signal variations.

CHAPTER 5

CONCLUSION & FUTURE POSSIBILITIES

5.1 CONCLUSION

In this dissertation, a novel watermarking scheme for authentication based on EMD, SVD and QIM is performed. Watermarking is embedded, spread in time domain using last IMF. Proposed algorithm being fragile, concentrates on imperceptibility and temper location tolerance to non-malicious attacks. Imperceptibility proves to be around 30 dB better in SNR and 10,000 times better in MSE than others in non-attacked as well in attacked cases. Its experimental results include 68.98dB in SNR and MSE of order 10^{-9}. Secondly, it is proved experimentally that it can locate temper location after malicious attacks, whether gone through non-malicious content persistent attacks or not. This algorithm, experimentally proves to be fit for audio signal integrity authentication but not for copyright protection. This is because the fragile watermark can be destroyed by an attacker. The possible improvisation of this method is to increase its watermark embedding capacity. It may be increased by using different coding schemes.

5.2 FUTURE WORK

For future work, possible flaws of EMD can be improved, to make watermarking algorithm robust in every case. Data pay load can be increased by several ways, like

- Taking say m number of watermark bits at a time, and then using 2^m number of quantizers to embed m watermark bits in audio signal.
- Using information coding theory for encoding original watermark bits.

Also, reliability in choosing watermark may have impact on the performance. Mathematical model can be built to create a best watermark for particular watermarking algorithm.

[1] Bassia P. and Pitas I., "Overview of Audio Watermarking Techniques," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, issue 2, pp. 32-34, 2002.

[2] Chen B. and Wornell G.V., "Quantization Index Modulation Methods for Digital Watermarking and Information Embedding of Multimedia," *Journal of VLSI Signal Processing Systems*, vol 27, no. 3, pp. 7-33, 2001.

[3] Chen S.T., Huang H.N., Chen S.J. and Wu1 G.D., "Energy-proportion based scheme for audio watermarking," *IET Signal Processing*, vol. 4, no. 4, pp. 576-587, 2009.

[4] Chen S.T., Wu G.D. and Huang H.N., "Wavelet domain audio watermarking scheme using optimisation based quantisation," *IET Signal Processing*, vol. 4, no. 2, pp. 720-727, 2010.

[5] Guo Q., Zhao Y., Cheng P. and Wang F., "An Audio Digital Watermarking Algorithm against A/D and D/A Conversions Based on DCT domain," *Proc. IEEE CECNet' 2012*, pp. 871-876.

[6] Herkilglo K., Yaslan,Y. Sener, S. and Gunsel, B, "Robust audio watermarking by adaptive psychoacoustic masking," *Proc. IEEE SPCA*, 2004, pp. 29 -32.

[7] HuaI H.X., Zhen J.W. and Xinl J.X., "Robust Audio Watermarking Based Non-Uniform DCT" , *Proc. IEEE ICIS*, 2010, vol 1,no. 1, pp. 585-588.

[8] Khademi N., Akhaeet M.A., Ahadi S.M., Moradi M. and Kashi A., "Audio Watermarking based on Quantization Index Modulation in the frequency domain," *IEEE International Conference on Signal Proces. & Comm. ICSPC 2007*, pp. 1127-1130.

[9] Khaldi K. and Boudraa A.O., "Audio Watermarking via EMD," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 21, no. 3, pp. 675-680, 2013.

[10] Khaldi K., Boudraa A.O., Turki M., Chonavel T., and Samaali I., "Audio encoding based on the EMD," *Proc. EUSIPCO*, 2009, pp. 924–928.

[11] Kiroveski D. and Malvar H.S., "Spread-spectrum watermarking of audio signals," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1020–1033, 2003.

[12] Ko B.S., Nishimura R., and Suzuki Y., "Time-Spread Echo Method for Digital Audio Watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 212-221, 2005.

[13] Lalitha N.V., Suresh G. and Telagarapu P., "Audio Authentication Using Arnold and Discrete Cosine Transform," *International Conference on Computing, Electronics and Electrical Technologies ICCEET*, 2012, pp. 530-532.

[14] Lanxun W., Chao Y. and Jiao P., "An Audio Watermark Embedding Algorithm Based on Mean-Quantization in Wavelet Domain," *Eighth International Conference on Electronic Measurement and Instruments ICEMI'2007*, vol. 2, no. 2, pp. 424 –425.

[15] Lei B., Soon I.Y. and Tan E.L., "Robust SVD-Based Audio Watermarking Scheme with Differential Evolution Optimization," *IEEE Trans. Audio, Speech, And Language Processing*, vol. 21, no. 11, pp.7-33, 2013.

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

[17] Tachibana R., Shimizu S., Kobayashi S., and Nakamura T., "An audio watermarking method using a two-dimensional pseudo-random array," *Signal Process*., vol. 82, no. 10, pp. 1455–1469, 2002.

[18] Liu J.X., Lu Z.M. and Shyang J.P., "A Robust Audio Watermarking Algorithm Based on DCT and Vector Quantization," *Eighth International Conference on Intelligent Systems Design and Applications*, 2008, vol. 3, pp. 541-544.

[19] Maha C., Maher E., Mohamed K. and Amar Chokri B., "DCT Based Blind Audio Watermarking Scheme," *Proc IEEE ICIP*, 2010, pp. 139-144.

[20] Norden E.H., Shen Z., Steven R. L., Manli C. W, Hsing H. S., Quanan Z., Nai-Chyuan Y.C., and Henry H. L., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903-995, 1998.

[21] Sekhar A.C., Suneetha C., NagaLakshmi G. and Kumar B.R., "Fast Fourier Transforms and Quadratic forms for Digital audio Watermarking," *International Conference on Advances in Recent Technologies in Communication and Computing*, pp. 449-452.

[22] Suresh G., Lalitha N.V., Rao C.S. and Sailaja V., "An Efficient and Simple Audio Watermarking Using DCT-SVD," *International Conference on Information Management, and Industrial Engineering 2012*, pp. 177-181.

[23] Chang L.N. and Lie W.N., "Robust Audio Watermarking Based on Low-Frequency Amplitude Modification," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 111-114, 2006.

[24] Tirkel A.Z., Schyndel R.G., Hall T.E. and Osborne C.F., "Secure Arrays for Digital Watermarking" 1998, available at http://goanna.cs.mit.edu.au~/~ronvs/papers/DICTA93.pdf.

[25] Wang C. and Guangming B.O., "A novel approach to generate and extract audio watermark," *Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2013, pp. 405-409.

[26] Wang L., Emmanuel S., and Kankanhalli M.S., "EMD and psychoacoustic model based watermarking for audio," *Proc. IEEE ICME*, 2010, pp. 1427–1432.

[27] Wang X.Y. and Zhao H., "A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 712-715, 2006.

[28] Wen X., Ding X., Jianhua L., Gao L. and Sun H., "An Audio Watermarking Algorithm Based on Fast Fourier Transform," *International Conference on Information Management, and Industrial Engineering* 2009, pp. 363-366.

[29] Wu S., Huang J., Huang D. and Yun Q. S., "Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission," *IEEE Trans. Broadcasting*, vol. 51, no. 1, pp. 213-315, 2005.

[30] Xie L., Zhang J. and Hongjie H., "Robust Audio Watermarking Scheme Based on Non uniform Discrete Fourier Transform," *Sichuan Key Lab of Signal and Information Processing Southwest Jiaotong University Chengdu*, 2009, China, pp. 1-5.

[31] Yan Y., Huang R. and Mintao X., "A Novel Audio Watermarking Algorithm for Copyright Protection Based on DCT Domain," *Second International Symposium on Electronic Commerce and Security* 2009, vol. 1, pp. 184-188.

[32] Zaman A.N.K, Ibrahim Khalilullah K.M., Islam W.I., and Khademul I.M., "A Robust Digital Audio Watermarking Algorithm Using Empirical Mode Decomposition," *Proc. IEEE ICEMI'2010*, pp. 1-4.

[33] Zeng G. and Qiu Z., "Audio Watermarking in DCT Embedding Strategy and Algorithm," *Proc. ICSP 2008*, pp. 2193-2196.

LIST OF PUBLICATION

1 “Fragile Audio Watermarking using EMD” communicated in AEU Elsevier *International Journal of Electronics & Communication*.