A non-blind multiple transform based audio watermarking based on highest entropy sub-bands

Jeebananda Panda, Kaberi Nath, Shrikishan Choudhary, Sunil Kumar

Abstract— With the rapid development of technology, digital format of information has become very popular in the recent era. People are tending more and more towards the use of mobile phones, laptops, computers etc. But the problem comes when fraudulent attacks are done on this digital data. So the concept of digital watermarking comes to protect these data. The holder of the digital content can hide some kind of information into the signal and can claim his ownership after the attacks. In this paper a robust and efficient watermarking algorithm is presented ensuring the copyright protection of the digital data. In digital audio watermarking some information is hidden into an audio file in a way such that it is transparent to the human ears as well as resistant against the removal of the hidden data from the audio. The possible usage of this is copyright protection, temper proofing of any multimedia file etc. In this algorithm a specific synchronization code followed by a binary image is inserted into highest entropy sub-bands of each frame of the host audio. Synchronization codes are generated by comparing the energy values of each sub-band to its subsequent one for a particular frame. The watermarked audio is indistinguishable from the host audio and thus it ensures the imperceptibility. The performance evaluation is done in terms of BER, Normalized Correlation (NC) and SNR values. Experimental results show that this scheme is robust against common signal processing attacks like noise addition, re-sampling, re-quantization, lowpass filtering, MP3 compression etc. and it can also resist some of the de-synchronization attacks like cropping.

Index Terms— Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Entropy, Singular Value Decomposition (SVD).

1 Introduction

Development of technology has introduced digital way of storing a great amount of data. Now a days, almost all the civilized population is using computers and internet which help to store a large amount of data for a long time without any kind of physical modification or damage. But intentional or unintentional manipulation of these data causes a huge threat to the ownership of that digital content. This situation calls for a solution called Digital Watermarking where the digital content endures any kind of intentional and unintentional modification or removal. This technique ensures the copyright protection of the digital contents. The embedded watermark contains some information which helps the owner to claim copyright on their intellectual property. Here in this paper Digital Audio Watermarking is done in order to ensure protection to the copyright holder against any malicious and signal processing attacks. In this technique watermark information is hidden in a host audio file in such a way that the quality of the host audio does not degrade and the watermark is recoverable even after the fraudulent attacks or removal on the watermarked audio. The watermark is known only to the owner, thus making this method highly robust and imperceptible. However, conflict arises while ensuring both robustness and imperceptibility. There is a tradeoff between these requirements while establishing an efficient algorithm. Here are some major requirements [1], [2], [9]:

- ☐ Robustness: It is the basic requirement of any watermarked signal. The signal should sustain its original form after common signal processing attacks or modification so that the watermark can be extracted from it.
- ☐ Perceptual Transparency: The additional information which is inserted into the host audio, should not distort the perceptual quality level of the audio. Therefore the imperceptibility of the watermark after embedding should also be ensured.
- ☐ Security: Security of the watermarked audio must be ensured in such a way that the attacker can not extract the watermark without knowing the secret key even if the extraction algorithm is known to him. So the secret key should be known only to the owner.
- ☐ Payload: Number of bits embedded per second in the host audio is termed as payload of the watermarking. Depending on the requirements of the watermarking algorithm, it can be high as well as low.

Imperceptibility and robustness are given priority and thus different algorithms are proposed with different properties according to the applications.

The sensitivity of Human Auditory System (HAS) towards additive white gaussian noise (AWGN) is more as compared to the Human Visual System (HVS) with a wider dynamic range. This adds up complexity to the audio watermarking as compared to image and video watermarking. So it is seen that, less works are done on audio watermarking than video or image watermarking.

Most of the watermarking techniques are classified mainly into two categories: i) Temporal Watermarking and ii) Spectral Watermarking. In temporal watermarking, watermark is directly embedded in time domain. It is easy to implement but less resistant against different attacks. On the other hand spectral watermarking has the advantage of several transforms in frequency or time-frequency domain. Thus it makes the embedding robust against the common attacks. Some fragments of the audio file is transformed and the embedding is done in these fragments. Commonly used transformations for this are Discrete Cosine Transform (DCT), Discrete Sine Transform (DST), Discrete Wavelet Transform (DWT), Fast Fourier Transform (FFT) etc.

There are two ways of extraction of the watermark from the watermarked audio. One is blind extraction and the other is non-blind or informed extraction [2]. In the blind extraction, host audio file is not needed for the retrieval of the watermark. In non-blind method the extraction of watermark is not possible without the host audio file. Here we propose an embedding algorithm which embeds watermark in highest entropy sub-bands of the host audio. In this paper, the algorithm is Non-blind in nature. In this scheme, 1-level DWT transform is performed on the host signal, resulting approximation coefficients and detail coefficients [1], [2]. In order to ensure robustness against the attacks, watermark is inserted in the highest entropy sub-bands [7] of the low frequency coefficients of each frame. Synchronization code [4], [10], [11] is also made a part of the watermark itself and it is generated from the host audio itself by energy comparison. There is a unique synchronization code for each frame. We will discuss about the embedding and extraction process in the upcoming sections of this paper. Rest of the paper is organised as follows:

Section II explains the related works that are done in this particular area, Section III explains the transformations used in the algorithm, Section IV explains the embedding algorithm, Section V explains the extraction algorithm, experimental results are shown in Section VI and Section VII contains the conclusion.

2 Related Works

There are several hybrid algorithms of embedding have been proposed which implements more than one transforms in order to ensure higher robustness. Mustapha Hemis and Bachir Boudraa have given an hybrid algorithm implementing both DWT and SVD [1]. Multiresolution decomposition of wavelet transforms and optimal matrix decomposition of SVD ensures high robustness, high payload and performance. Prof. Sujit M. Deokar and Bhaveek Dhaigude have shown an efficient watermark embedding algorithm using both DWT and DCT [2]. Arnold transform is used for scrambling of the watermark image. This scrambling makes the watermark almost undetectable for the attacker thus increasing the robustness. Yang et al. have shown a new method called zero-watermarking [3]. The embedding is done here into a secret key, not in the host signal itself. XOR operation is been done between the binary image and the relational value array in order to generate this secret key. This relational value array is generated by comparing the energy values of the frames of the host audio after doing 1-level DWT. Vivekananda Bhat K et al. have demonstrated a secure, efficient and robust watermarking algorithm using SVD in the wavelet domain [4]. Here embedding is done by applying Quantization index modulation (QIM) on the singular values (SVs) of the wavelet domain blocks along with the synchronization code. Mitchell D. Swanson et al. have proposed another embedding algorithm using perceptual masking [5]. This algorithm exploits temporal and frequency perceptual masking property and the watermark embedded here is generated from a pseudo-random sequence to ensure high robustness. Fathi E. Abd El-Samie again implements SVD to have a robust watermarking scheme [6]. Pranab Kumar Dhar and Tetsuya Shimamura have presented another efficient scheme where DCT, SVD and log-polar transformation is used [7]. Watermark is embedded into the highest entropy bands of the host audio. Ergun Ercelebi and Leyla Batakcı have shown a new algorithm by using LBWT (Lifting Based Wavelet transform) [8]. In this algorithm a pseudo random number matrix is generated using pseudo random number generator, which is embedded into the host audio depending on the values of the binary image. This method increases the tamper resistance of the watermark.

3 Transformations Used

Discrete Cosine Transform (DCT): In this transform, a finite sequence of summation of cosine functions can represent an audio signal at different frequency values. Mathematically it can be represented as follows:

$$X(k) = w(k) \sum_{n=1}^{N-1} \mathrm{x(n)} \cos\{\pi \frac{(2n-1)(k-1)}{2N}\},$$

$$k = 0,1 \ldots\ldots N-1$$

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{2} \times \frac{1}{\sqrt{N}}, & k = 1,2 \ldots\ldots (N-1) \end{cases}$$

Here, N gives the total number of samples in the sequence and x(n) is the original audio signal. DCT reduces the distortion occurring due to the embedding of the watermark because the energy is congested in a small portion of the sequence in DCT coefficients.

Discrete Wavelet Transform (DWT): Hierarchical decomposition of an input signal into a series of approximation coefficients and detail coefficients is known as the DWT or Discrete Wavelet Transform of that signal. The low frequency approximation coefficients can further be decomposed into two parts of approximation and detail coefficients. Fig1. Shows a 2-level DWT input signal A, cA and cD are low frequency approximation coefficients and detail coefficients respectively.

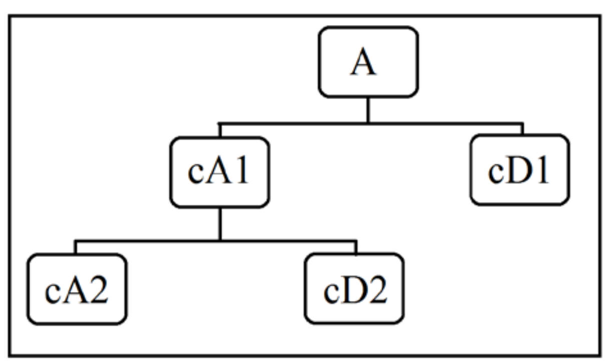


Fig1: 2-level DWT

Singular Value Decomposition(SVD): SVD of a matrix A of order $p \times p$ is defined as follows:

$$A = USV^T$$

Here, U and V are orthogonal matrices of order $p \times p$ and S is a diagonal matrix of order $p \times p$ with non negative elements. The diagonally placed values of S are known to be singular values (SVs) of A. The columns of U are known as the left singular vectors of A. The columns of V are known as the right singular vectors of A.

4 Embedding Algorithm

The proposed scheme is based on dual transformation DWT-DCT [2]. The main idea of embedding is that watermark should be embedded in the highest entropy low frequency DCT coefficients [7] of the approximation coefficients obtained after performing 1-level DWT of the original host signal. The Embedding method is shown in Fig.2.

The watermark is embedded in the following steps:

- Perform 1-level DWT to the original audio signal.
- Perform DCT to the approximation coefficients obtained from the 1-level DWT of the original signal.
- Partition the signal into N frames F={f_1, f_2, f_3 .... ,f_N}. These frames are non-overlapping to each other.
- Select the former L low-frequency components of each frame.
- Divide the low-frequency components in M sub-bands, each containing N elements. Reshape them in N×M order matrix $L_{j,NxM}$.
- Calculate the entropy of each sub-band of matrix $L_{j,NxM}$. [7]:

$$E_j = entropy(L_j) = -\sum_{i=1}^{K} s_i log_2 s_i$$

Where, i, s, E_j are the index number of the DCT coefficient, amplitude of DCT coefficients and entropy of the sub-band L_j respectively.

- Select the sub-band $B_{j,Nx1}$ of each frame f_j having maximum entropy and diagonalise it to matrix $B_{j,NxN}$.

$$E_{max} = \max\{E_1, E_2, \dots, E_M\}$$

'max' operation returns the maximum of the $E_1, E_2, \dots, E_M$.Thus E_{max} represents the maximum value among the $E_1, E_2, \dots, E_M$ [7]

- Store the positions of highest entropy sub-bands of all the frames in Key1.
- Take the watermark image of order (N-M)×(N-M) , and take SVD.

$$I = USV^T ,$$

Where, I is the watermark image, S is a diagonal matrix of order (N-M)×(N-M). U and V are used at the time of extraction.

- Concatenate the M bit Synchronization code (generation of Synchronization code is explained in the later section) with the diagonal matrix generated by SVD of image. The synchronization code is made a part of watermark $W_{j,NxN}$. [10]

Synchronization code	Watermark	....	Synchronization code	Watermark

- Embed the watermark $W_{j,NxN}$ in the maximum entropy subband of each frame $B_{j,NxN}$.

$$\hat{B}_{j,NxN} = B_{j,NxN} + \alpha W_{j,NxN}$$

Where $W_{j,NxN}$ is the watermark consisting of M synchronization code bits and (N-M) bits of image. α defines the embedding strength of the watermark and $\hat{B}_{j,NxN}$ is the modified highest entropy sub-band matrix of each frame.

- Substitute the original sub-bands by the modified sub-bands in the original frames.
- Concatenate all frames.
- Apply Inverse DCT (IDCT).
- Apply Inverse DWT (IDWT) to obtain the watermarked audio.

Steps for generating Synchronization code:

- Calculate the energy of each sub-band of matrix $L_{j,NxM}$ for each frame.
- For every frame, compare the $Energy(i)$ of each sub-band to its subsequent one. A Relational Value Array is generated of size M for each frame which is used as synchronization code.
- Relational Values can be obtained by using these conditions [3] :

$$Relational\ Value = \begin{cases} 1, & Energy(i) > Energy(i+1) \\ 0, & Energy(i) \le Energy(i+1) \end{cases}$$

- These Relational Value Arrays are used as synchronization codes for the corresponding frames.

Fig3. shows the steps for generation of synchronization code.

ISSN 2229-5518

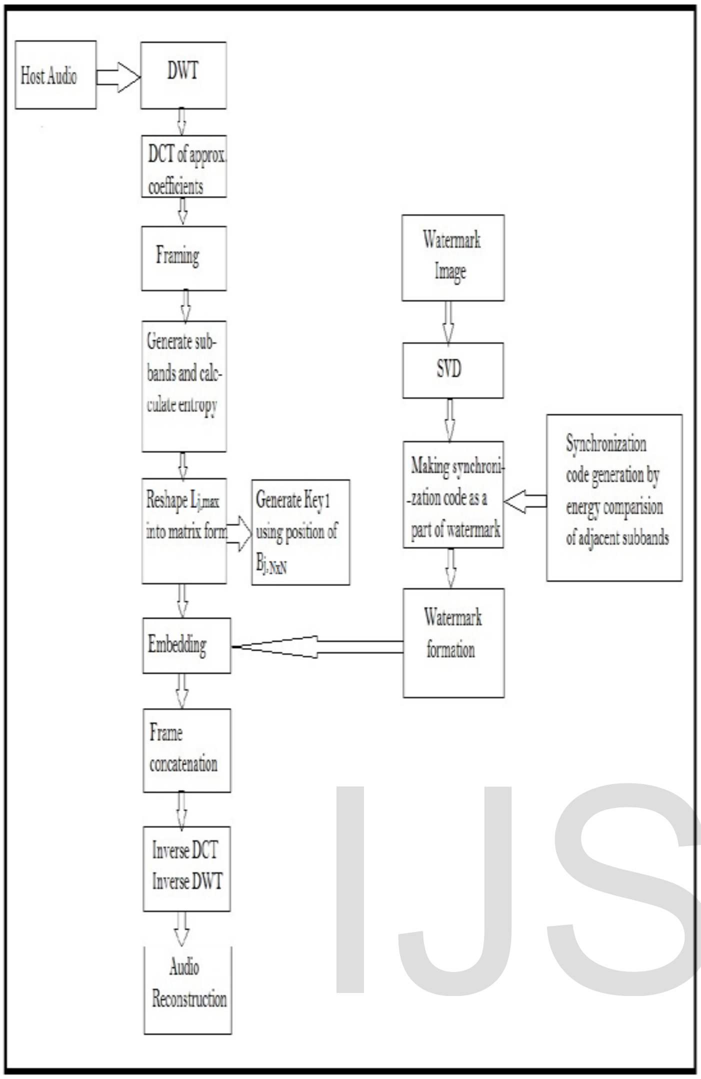


Fig2: Embedding algorithm

5 EXTRACTION

In this proposed scheme watermark is extracted in Non-blind way.

Steps for extraction of watermark:

- Perform one-level DWT to the watermarked audio signal.
- Perform DCT to the approximation coefficients obtained from DWT of the watermarked signal.
- Divide the signal into N frames F={f_1, f_2, f_3 .... ,f_N}. These frames are non-overlapping to each other.
- Select the former *L* low-frequency components of each frame.
- Divide the low-frequency components in M sub-bands, each containing N elements. Reshape them in N×M order matrix $L_{j,N\times M}$.
- Using the Key1, extract the highest entropy sub-bands of each frame.
- Subtract each highest entropy sub-band of host signal from the corresponding sub-band of watermarked signal.
- Extract the watermark signal:

$$\widehat{W}_{j,N\times N} = \frac{(\hat{B}_{j,N\times N} - B_{j,N\times N})}{\alpha},$$

Here, B and $\hat{B}$ are the highest entropy sub-band matrix of the host signal and watermarked signal respectively and $\widehat{w}$ is the extracted watermark.

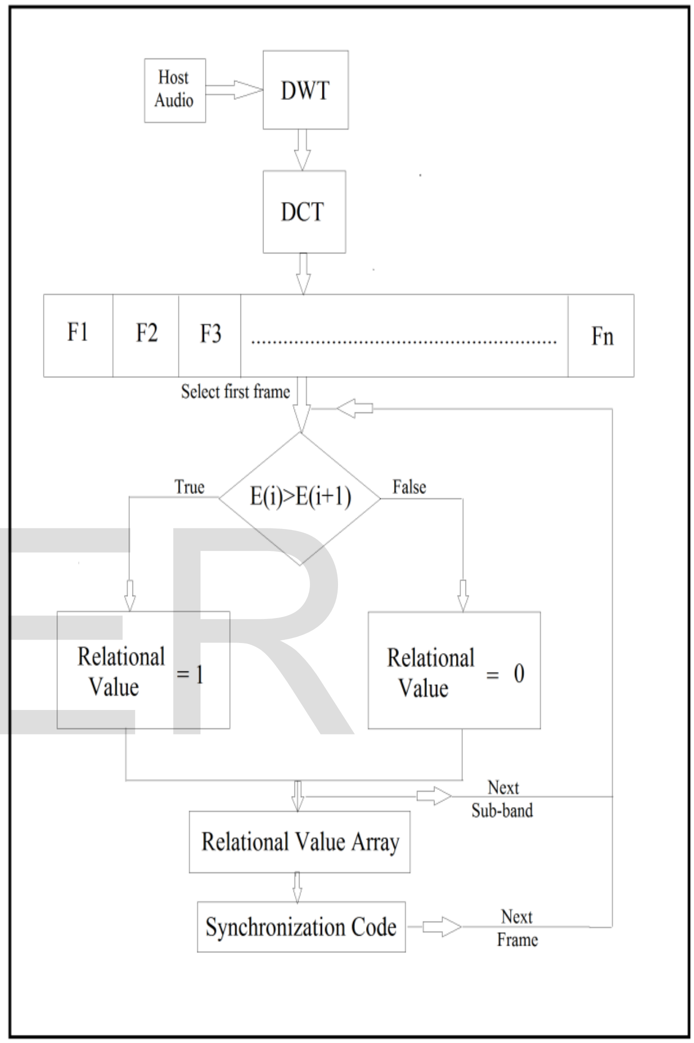


Fig3: Synchronization code generation

- First M bits are synchronization code bits and the rest (N-M) bits are the SVs $(\hat{S})$ of the watermark image.
- Perform inverse SVD to obtain the watermark image:

$$\hat{I} = U\hat{S}V^T$$

Where, $\hat{I}$ is the extracted watermark image. Fig4 shows extraction algorithm.

6 EXPERIMENTAL RESULTS

In order to examine the performance of the algorithm, several tests were performed on two audio files Audio1.wav and Audio2.wav of duration 12 seconds and 7 seconds respectively.

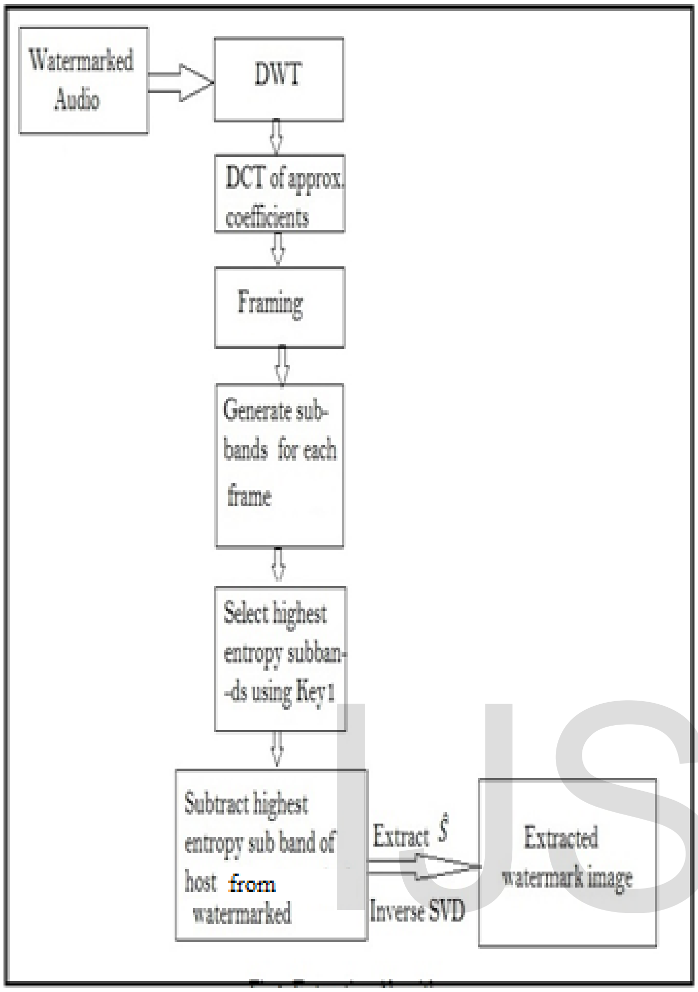


Fig4: Extraction Algorithm

Both the signals are sampled at 44100Hz frequency. The Audio1.wav file consists of 11,28,960 elements. It is divided into 30 frames, having 37,632 elements in each frame. 1-level DWT is performed and a total number of 5,64,480 approximation coefficients are obtained. Then the approximation coefficients obtained by DWT undergo DCT. From each frame, first 300 low frequency DCT coefficients are selected after this. From these 300 low frequency DCT coefficients a 30×10 matrix is formed in every frame i.e. 10 sub-bands are formed out of these coefficients and 30 elements are comprised in each sub-band.

Highest entropy sub-band is selected from every frame. A key named Key1 is stored containing the positions of highest entropy sub-bands from each frame. The energy values of each sub-band is compared with the energy of subsequent one.

For each frame, the Relational Value Array obtained by energy comparison is used as Synchronization Code. Thus a distinct 10 bit Synchronization Code is assigned to each frame. A 20×20 binary image is used as the watermark image. The watermark image undergoes SVD and singular values obtained from the SVD is concatenated with the Synchronization Code. In this way, the Synchronization Code is made a part of the watermark. The size of the watermark which is obtained finally is 30 bits. This watermark of 30 bits is embedded into the highest entropy sub-band of each frame. To obtain the optimum result after the attacks, the Embedding strength α is chosen to be 0.08 after a series of trials.

To check the performance of the algorithm, SNR (Signal to Noise Ratio) value of the watermarked signal is calculated [1] [4] [7].

SNR is the ratio of signal power to the noise power. It is usually measured in decibels (dB) and represents the strength of the signal. A higher value of SNR indicates higher imperceptibility.

$$SNR = 10\times\log\left(\frac{\sum_{k=0}^{N-1} x(i)^2}{\sum_{k=0}^{N-1}[x(i)^2 - x_w(i)^2]}\right)$$

The minimum allowable value of SNR for distortion less sound is 20 dB. In the proposed scheme SNR comes out to be 24.0821 dB and 23.3807 dB for Audio1.wav and Audio2.wav respectively.

The other criteria for measuring the performance are Normalized Correlation (NC) and Bit Error Rate (BER) [1] [3] [4] [7].

Normalized Correlation (NC) represents the similarity of two series. The value of it lies between 0 and 1. A higher value of NC indicates a better performance.

$$NC = \frac{\sum_{i=1}^{H}\sum_{j=1}^{H} W(i,j)\widehat{W}(i,j)}{\sqrt{(\sum_{i=1}^{L}\sum_{j=1}^{L} W(i,j)^2)}\times\sqrt{(\sum_{i=1}^{L}\sum_{j=1}^{L}\widehat{W}(i,j)^2)}}$$

Where $\hat{w}$ is the extracted watermark signal and w is the original watermark which is embedded.

Bit Error Rate (BER) indicates the number of error bits per unit signal bit.

$$BER = \frac{1}{H\times H}\sum_{i=1}^{H\times H}(\widehat{W}_i - W_i)$$

Where, H is equal to (N-M).

Attacks which are performed in order to examine the robustness of the algorithm are as follows [1] [3] [7] :

a) Noise addition: A 20dB White Gaussian Noise (AWGN) is added to the watermarked signal.
b) Cropping: The watermarked signal is cropped from three different ways-
   a) Front cropping-50% samples of the audio is removed from the end part.
   b) Middle cropping-50% samples from the middle part of the audio is retained and the rest part is removed. and

c) End cropping-50% samples from the front part is removed and the watermark is extracted from the rest part.

c) Re-quantization: The watermarked signal is quantized to 8bits/sample and then again stored back to 16bits/sample.
d) Re-sampling: The watermarked signal is first down-sampled to 22.05 kHz and then up-sampled to 44.1 kHz.
e) Mp3 Compression: 128 kbps MPEG layer III compression is performed.
f) Lowpass filtering: Watermarked audio file is passed through a 10^{th} order lowpass filter having cut-off frequency 11025 Hz.
g) Weiner Attack: Watermarked audio file is passed through a wiener filter with default argument value 3, considering the AWGN power to be noise [12] [14].

Table1 shows the experimental results of various attacks on the watermarked signal. Here extraction is non-blind.

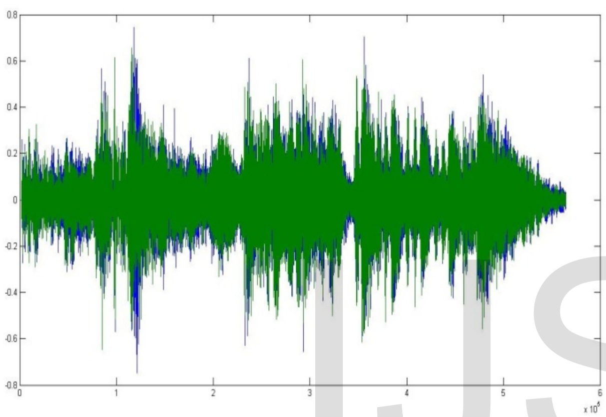


Fig 6: Host Audio(Audio1.wav)

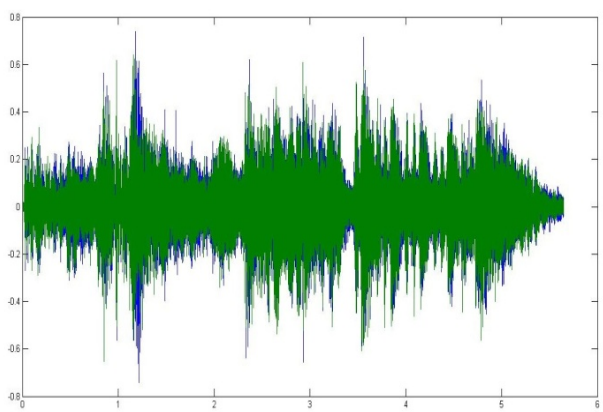
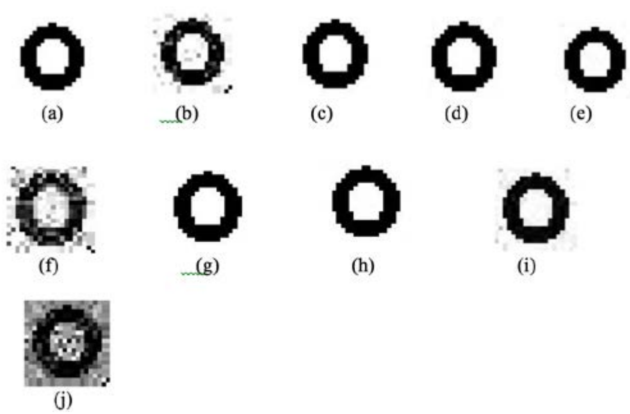


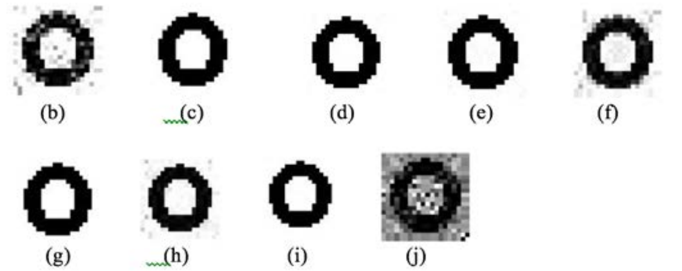
Fig 7: Watermarked Audio(Audio1.wav)



Audio1.wav

Audio2.wav

Table1: Experimental results



Attacks	Music1.wav		Music2.wav	
	BER	NC	BER	NC
1.No Attack	0	1	0	1
2.AWGN Attack	0.1716	0.9771	0.1490	0.9776
3.Re-Sampling	0.0012	1	0.0028	1
4.Re-Quantization	0.0052	1	0.0049	1
5.Mp3 Compression	0.0038	0.9998	0.0063	0.9997
6.Low pass filtering	0.1970	0.9500	0.0797	0.9921
7.Front Cropping	0	1	0	1
8.Middle Cropping	0.0044	1	0.0167	0.9996
9.End Cropping	0.0391	0.9980	0.0051	1
10.Wiener Attack	0.4032	0.8650	0.4034	0.8540

7 Conclusion

This paper presents a watermarking algorithm exploiting the masking property of Human Auditory System. Since this algorithm guarantees the imperceptibility of the embedded watermark, it comes really handy in case of copyright protection. The audio file undergoes DWT-DCT and watermark is embedded into highest entropy sub-bands in each frame of the audio. This schemes exploits the properties of DWT and DCT. Most of the energy content is being comprised in low frequency components, DCT is used as to get low frequency components of the audio file. The multi-resolution properties both in time and frequency domain are analyzed with the help of DWT.

This scheme is robust to almost all kind of attacks and guarantees high security against unauthorized extraction. Thus the original watermark which is here a binary image can be extracted even after the attacks. In this particular scheme the watermark is hidden in all the frames along with their respective synchronization codes rendering the embedding resistant to cropping attacks and other attacks like noise addition, re-sampling, low pass filtering etc. Thus the algorithm is much more robust than other schemes and is also imperceptible.

REFERENCES

[1] Mustapha Hemis and Bachir Boudraa, "Digital watermarking in audio for copyright protection", Advanced Computer Science and Information Systems (ICACSIS), 2014, International Conference on year: 2014, pages: 189-193.

[2] Prof. Sujit M. Deokar and Bhaveek Dhaigude, "Blind Audio Watermarking Based On Discrete Wavelet and Cosine Transform", 2015, International Conference on Industrial Instrumentation and Control (ICIC), College of Engineering Pune, India, May 28-30, 2015.

[3] Yang Yu, Lei Min, Cheng Mingzhi, Liu Bohuai, Lin Guoyuan, Xiao Da, "An Audio Zero-Watermark Scheme Based on Energy Comparing", China Communications, Year: 2014, Volume: 11, Pages: 110-116, IEEE Journals & Magazines.

[4] Vivekananda Bhat K, Indranil Sengupta and Abhijit Das, "An adaptive audio watermarking based on the singular Value decomposition in the wavelet domain", Digital Signal Processing, Volume 20, Issue 6, December 2010, Pages 1547-1558.

[5] Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik and Laurence Boney, "Robust audio watermarking using perceptual masking", Signal Processing, Volume 66, Issue 3, 28 May 1998, Pages 337-355.

[6] Fathi E. Abd El-Samie, "An efficient singular value decomposition algorithm for digital audio watermarking" , International Journal of Speech Technology, March 2009.

[7] Pranab Kumar Dhar and Tetsuya Shimamura, "Blind SVD-based audio watermarking using entropy and log-polar transformation", Journal of Information Security and Applications, Volume 20, February 2015, Pages 74-83.

[8] Ergun Ercelebi and Leyla Batakci, "Audio watermarking scheme based on embedding strategy in low frequency components with a binary image", Digital Signal Processing, Volume 19, Issue 2, March 2009, Pages 265 -277.

[9] Farooq Husain, Omar Farooq and Ekram Khan, "Digital Audio Watermarking", MIT International Journal of Electronics and Communication Engineering, Vol 1 No. 2, August 2011, pp 101-105.

[10] Shuo-Tsung Chen, Chih-Yu Hsu and Hunag-Nan Huang , "Wavelet-domain audio watermarking using optimal modification on low-frequency amplitude", IET Signal Processing, August 2014.

[11] Shogo Nakaya and Shigeo Wada, "Audio Watermarking Based on Vector Quantization Index Modulation Using Audio Fingerprint", Electronics and Communication in Japan, Vol. 98, No. 3, 2015.

[12] Jonathan K. Su and Bernd Girod, "Power-Spectrum Condition for Energy-Efficient Watermarking", IEEE Transactions on Multimedia, Vol. 4, No. 4, December 2002.

[13] Hwai-TsuHu and Ling-YuanHsu, "Robust, transparent and high-capacity audio watermarking in DCT domain", Signal Processing, Volume 109, April 2015, Pages 226-235.

[14] Jonathan K. Su and Bernd Girod, "Fundamental Performance Limits of Power-Spectrum Condition-Compliant Watermarks", Proc. SPIE, Vol. 3971 : Security and Watermarking of Multimedia Contents II San Jose, CA, USA, Jan. 2000.

[15] Shu-Min Tsai, "A Robust Zero-Watermarking Scheme for Digital Audio", International Journal of Information and Electronics Engineering, Vol. 5, No. 2, March 2015.

[16] Asha Rani, Amandeep K. Bhullar, Deepak Dangwal and Sanjeev Kumar, "A Zero-Watermarking Scheme using Discrete Wavelet Transform", 4th International Conference on Eco-friendly Computing and Communication System (ICECCS), Volume 70, 2015, Pages 603-609.