ID: 2016-ISFT-142

A Robust Audio Watermarking Technique Using Dual and Hybrid Watermark

Jeebananda Panda[1], Nitish Goel[2], Indu Kumari[3], Asok Bhattarchyya[4]

[1,2,3,4]Department of Electronics and Communication, Delhi Technological University, Shahbad Daultpur, Bawana Road, Delhi-110042
[1]ngoel0108@gmail.com

Abstract: *Digital watermarking has received a lot of consideration from the research community in the last few years with rapid advancements in internet services. Various multimedia applications such as images, audio and video signals are shared between different sources and hence the data should be secluded against illegitimate access. Watermarking techniques are used to provide copyright protection, owner's identification, tamper protection and message authentication. In this paper, a new robust watermarking technique has been proposed in which dual and hybrid watermark images, one binary image and other gray-scale image are embedded in audio samples transformed using FFT. The binary image is embedded in the real part of FFT coefficients using multiple bit plane method and the gray scale image is embedded in the imaginary part using a weighted coefficient after passing it through a 2-level DWT. Multiple bit plane scheme ensures robustness and resistance to attacks whereas DWT applied on the watermark prior to embedding certifies that only imperative feature of the image is embedded. The proposed algorithm provides high security of watermarks and better SNR of the audio signal. Also the proposed scheme is tamper proof as the owner's identity can be established even if any one of the watermarks are recovered. From the results, it can be concluded that the proposed audio watermarking scheme provides superior results then the existing methods.*

Keywords: *Digital audio watermarking, dual and hybrid watermark, frequency domain, multiple bit plane, DWT, similarity, correlation, SNR*

1. INTRODUCTION

Watermarking techniques are required to secure the digital data and to prevent the unauthorized access. Due to advancements in internet and digital technologies, the digital media such as images, audio and video signals are easily shared and accessible across the transmission channel. Since a large amount of data can be accessed from the communication media, it can be manipulated and tampered easily due to presence of a large number of digital signal and image processing tools. It is easier to misuse the valuable information by hacking. Hence, the copyright and ownership of various multimedia files is not protected against illicit sources. The technology needs to be developed to certain the security and authenticity of data being shared over a network. To counteract these malpractices, watermarking techniques are used. It is the process of embedding a watermark into the host digital signal. The original contents of the signal are modified by the secret watermark inserted as a proof of authentication and ownership. The host signal can either be an image [1] or an audio signal [2]. The watermarking technique applied must be robust and effectual so that the embedded watermark cannot be perceived by human eyes and also the quality of host signal is not degraded by insertion of a foreign signal. At the same time, it should be efficient in case of various attacks that degrade the quality of digital signal and high security of embedded watermark should be provided. Digital watermarking is of two types and can be categorized into time-domain and frequency domain watermarking. In time-domain or spatial domain watermarking techniques, the watermarks are inserted directly into the bit planes of the host message signal. The bit values of the host signal, which can be either an image or an audio signal, are set according to the value of the watermark signal vector. In frequency-domain, watermarks are inserted into the coefficients of the host signal obtained after applying the Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT). The method presented in this paper is Audio Watermarking [3] in which the host message signal is an audio file in which the watermark, used as a proof of novelty, is embedded. Many algorithms have been proposed for the same in time domain such as LSB Substitution method [4], multiple bit plane techniques and in frequency domain such as DCT and DWT based watermarking [5]. The method presented here is Fast Fourier Transform (FFT) method in concurrence with multiple bit plane technique and DWT. In this paper, two watermarks, a binary image and a gray scale image are embedded in the frequency transformed audio samples. The watermarked audio signal is then subjected to various attacks [6] before the watermark is extracted to check the authenticity of the message. The algorithm should be robust so that the embedded watermarks do not affect the original usability of the message signal and is invisible and inaudible to the human senses. The results of the proposed method are improved over existing techniques and have been summarized.

The following paper has been divided into 10 sections. Section 2 gives an overview of the previous research done in this field. Section 3 gives general framework of watermarking. Section 4 gives a basic idea of Fast Fourier Transform. Section 5 describes the Discrete Wavelet Transform. Section 6 briefs about multiple bit plane technique. Section 7 gives the proposed algorithm of watermarking. Section 8 briefs about various attacks. Section 9 includes experimental results followed by Section 10 that gives conclusions and future scope.

2. LITERATURE SURVEY

Deepshikha Chopra et. al [1] presented an invisible watermarking technique (lower LSB e.g. 1st) and a visible watermarking technique (higher LSB e.g. 7th) for image. Each technique is evaluated and studied by Peak signal to noise ratio (PSNR) and mean square error (MSE). The work presented in [2] by Prof. Jeebananda Panda et. al. is the a bit manipulation technique of audio watermarking in time-domain. The least significant bit of watermark is embedded in audio signal. Synchronization bits are added along with watermark before embedding for security for watermark signal in case of cropping. The watermarked signal is subjected to various attacks and by computing similarity between original and recovered watermark and SNR of audio signal, it is observed that algorithm is more robust corresponding to embedding of watermark on 3rd and 4th LSB. A watermarking method based on multiple bit plane and chaotic scrambling has been presented by Jeebananda Panda et. al. [3] The watermark is subjected to scrambling algorithm before it is embedded in LSB of audio samples in multiple bit plane. The proposed scheme is subjected to various attacks and is tested using various comparison measures. Ali Al-Haj et. al. [4] presented an algorithm for audio watermarking where the watermark image is embedded at LSB of the coefficients of transformed host audio signal that is obtained using Discrete Wavelet Transform (DWT). In [5] Mei Jiansheng introduced an algorithm of digital watermarking based on Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). According to the characters of human vision, in this algorithm, the information of digital watermarking which has been discrete cosine transformed, is put into the high frequency band of the image which has been wavelet transformed. Then distil the digital watermarking with the help of the original image and the watermarking image. This algorithm is invisible and has good robustness for some common image processing operations. The work presented in [6] by Jeebananda Panda et. al. is frequency domain energy efficient watermarking of audio signal i.e. in audio samples, an energy efficient watermark is embedded which satisfies the power spectrum condition. The watermarked audio is subjected to attacks such as noise, cropping, re-sampling. The feature of the proposed scheme is to resist minimum mean square error (MMSE) estimation attack. Md. Iqbal Hasan Sarker et. al. [7] did FFT based audio watermarking with a gray image for copyright protection. First, the original audio is transformed into FFT domain. The FFT coefficients are divided into a fixed number of subsections and the energy of each subsection is calculated. Next, the watermark is generated for gray image and are embedded into selected peaks of highest energy subsection. The performance is evaluated based on similarity measure and SNR. A high capacity robust audio watermarking scheme based on FFT and linear regression is used [8] by Mehdi Fallahpour et. al. Linear regression helps to minimise the alterations of FFT samples, which results in better transparency.

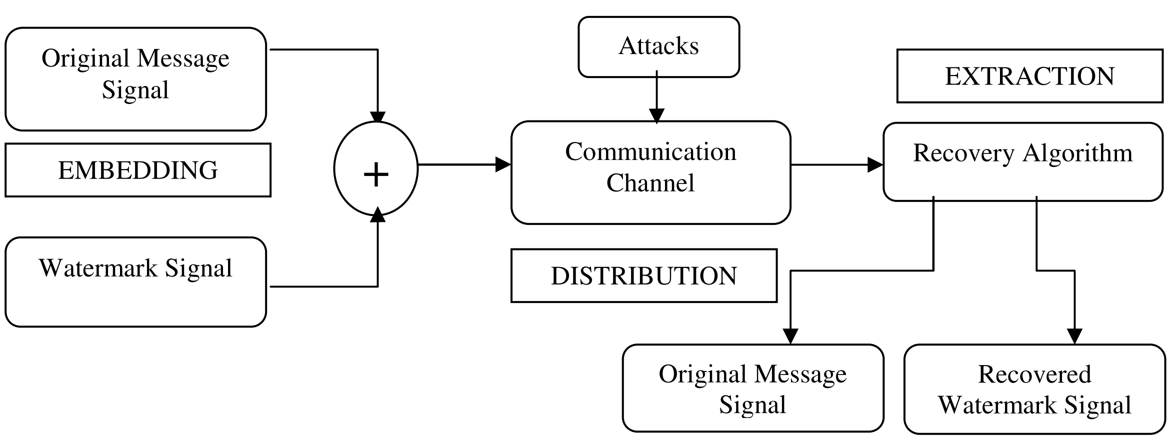


Fig. 1. Basic Model of Watermarking

3. GENERAL FRAMEWORK OF WATERMARKING

Digital Watermarking procedure is mainly classified into three steps:

1) Embedding

2) Distribution

3) Extraction or Detection

Embedding is the process in which the watermark vector used as a secret data is embedded into the host message signal. The bit values of host signal are set or reset according to the bit value of the watermark signal. Then this embedded data is transmitted from source to destination across a noisy channel where it is subjected to various attacks such as Gaussian noise, re-sampling, filtering, compression and cropping. This step is known as distribution. The third step is the extraction in which recovery algorithm is applied to get back the inserted watermark signal from the host signal.

4. FAST FOURIER TRANSFORM (FFT)

Fast Fourier Transform (FFT) is an efficient and fast algorithm for computing the Discrete Fourier Transform (DFT) and the inverse Discrete Fourier Transform (IDFT). DFT deals with a finite discrete-time signal and a finite or discrete number of frequencies. It decomposes a sequence of values into components of different frequencies. Given a sequence of N numbers x_0, x_1, x_2, $x_3 \ldots x_{N-1}$ the N point DFT is computed as :

$$X_k \stackrel{\text{def}}{=} \sum_{n=0}^{N-1} x_n \cdot e^{-2\pi i k n/N} \quad \text{where} \quad k = 0,1,2,3 \ldots N-1 \qquad (1)$$

X_k = DFT Coefficients

However computing DFT from above expression is too slow for a larger sequence of numbers. Hence, FFT is a swift way of computing DFT. Radix-2 FFT algorithm is the most common method used for computing the N-point DFT. It re-expresses the discrete Fourier transform (DFT) of an arbitrary composite size $N = N_1N_2$ in terms of smaller DFTs of sizes N_1 and N_2, recursively, in order to reduce the computation time to O(N log N). Computing DFT from equation (1) takes N^2 complex multiplications and N(N-1) complex additions while using Radix-2 FFT algorithm, the complex multiplications are reduced to ($N\log_2 N$) while the complex additions are reduced to ($N/2\log_2 N$). Due to lesser number of calculations, FFT is used in practical cases for computation of DFT of a sequence of numbers. In this paper, 64-point FFT is computed taking 8 x 8 matrix of audio samples at a time.

5. DISCRETE WAVELET TRANSFORM (DWT)

A Wavelet transform is a spatial domain analysis method with fixed window size and convertible form. It contains information that is similar to short-time Fourier transform but with additional special features of wavelets.

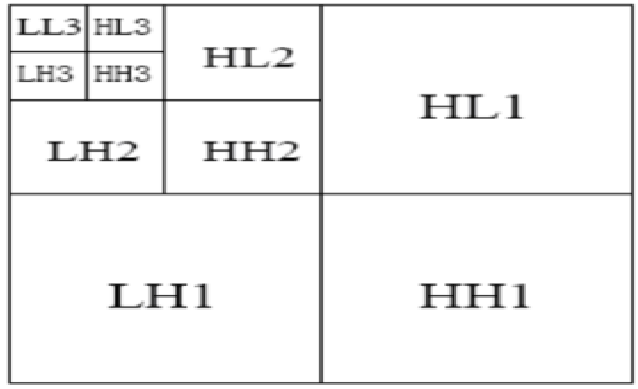


Fig. 2. Image decomposed by 3 level DWT

The basic principle of DWT when applied on an image is to multi-differentiated decomposes the image into sub-sections of different spatial domain and independent frequency components. This is known as single level DWT in which the original image is decomposed into 4 frequency components which is one low frequency component (LL) and three high frequency components (LH, HL, HH). The main information of the watermark image is contained in the low frequency component LL. In 2-level DWT, the LL obtained from 1-level DWT is transformed to obtain sub-level frequency district information. The low frequency component obtained from 2-level DWT can again be transformed to obtain a 3-level DWT. Hence, the original image can be decomposed into frequency regions LL1, LH1, HH1, HL1. The low-frequency component can be further decomposed into sub-level frequency regions LL2, LH2, HH2, HL2. The original image can be converted into n level wavelet transform by applying the DWT algorithm recursively on the low frequency component obtained after every stage of operation. Figure 2 gives the basic idea of 3-level DWT operation. The information stored in the low level component is closest to the original image and consists of most of the signal information. The frequency regions HL, LH and HH respectively represent the level detail, upright detail and diagonal detail of the original image. The watermark image used in this paper is a gray scale image of size 48 x 48 pixels that is decomposed using 2-Level DWT.

6. MULTIPLE BIT PLANE

In Multiple Bit Plane method, instead of modifying only the 1st LSB only of the all the audio samples, the watermark is inserted simultaneously in the 1st, 2nd, 3rd and 4th LSB of audio samples consecutively. If the audio samples are given by A such as:

$$A = \{ A(1,j) \; 0 \leq j \leq M \} \text{ where M=Number of samples} \qquad (2)$$

and watermark image is given by Img such as:

$$Img = \{ Img(1,i) \; 1 \leq i \leq M \times N \} \qquad (3)$$

where M, N are the dimensions of the watermark image and Img(1,i) is either 0 or 1 i.e. binary image, then 1st LSB of A(1,1) is set according to Img(1,1), 2nd LSB of A(1,2) is set according to Img(1,2), 3rd LSB of A(1,3) is set according to Img(1,3), 4th LSB of A(1,4) is set according to Img(1,4) and again repeating the same process till the entire length of audio samples.

7. THE PROPOSED SCHEME

In this section, the proposed watermark embedding and watermark extraction algorithms have been explained. The high channel capacity of audio signal permits a smaller watermark signal to be embedded in it multiple times. The algorithm proposed is the combination of frequency transform (FFT and DWT) and multiple bit plane method. For this, an audio signal S of extension .wav is used as the host signal and two watermarks, a binary image I_1 of size 50 pixels x 50 pixels and a gray scale image I_2 of size 48 pixels x 48 pixels are used as the watermark signals. The algorithm is written and implemented using MATLAB software.

A. Dual watermark embedding process

1) Firstly, the audio file is read and is sampled at the frequency of 44100 Hz. The duration of the audio signal taken is 16 seconds, hence total 705600 samples are used for watermark embedding.

$S = \{ S(1,j), 1 \leq j \leq K \}$ K= 705600 (4)

2) Convert the samples into 8-bit unsigned values prior to the process of embedding as the audio samples obtained, by default, are in double format.

3) Reshape the audio sample vector into a 2-D matrix $S_{2\text{-}D}$ of size M x N samples.

$S_{2\text{-}D} = \{ S(i,j), 1 \leq j \leq M, 1 \leq i \leq N \}$ M,N=840 (5)

4) Take the FFT of the audio sample matrix $S_{2\text{-}D}$ by taking 8 x 8 matrix at a time. Hence, at a time, 64 point FFT is being computed. Let the FFT Coefficient Matrix obtained be S_f. S_f obtained is 840 x 840 size matrix.

5) Separate the real part and imaginary part of the coefficient matrix S_f.

$S_f = R + jI$ where R and I are 2-D matrices of size 840 x 840 (6)

6) The binary watermark I_1 is embedded in the real part R of the coefficient matrix using multiple bit plane scheme and the gray scale watermark I_2 is embedded in the imaginary part I using a weighted coefficient α.

7) Gray scale watermark I_2 is passes though a 2-level DWT in which the LL1 of size 24 x 24 obtained after stage 1 is further decomposed into 4 frequency sections each of 12 x 12 pixel size. The LL2 obtained after stage 2 is used as the second watermark signal for embedding. The size of LL2 obtained is 12 x 12 pixels.

8) Both the watermarks are reshaped into column vector and a total of 128 replicas of both watermark signals are embedded in the transformed audio signal.

$C_1 = \{ C_1(i) = I_1(k,j), i = k \times j = 2500 \}$ (7)

$C_2 = \{ C_2(i) = LL2(k,j), i = k \times j = 144 \}$ (8)

9) After obtaining the reshaped vector, apply the Multiple bit plane method on real part R where R is the reshaped vector of size 705600 in which the 1st LSB of first sample R(1), 2nd LSB of second sample R(2), 3rd LSB of third sample R(3), 4th LSB of fourth sample R(4), then again 1st LSB of fifth sample R(5) till the entire length of audio signal is set according to the binary image vector. Zero value of the image array sets the LSB of audio file sample to 0 and non-zero value of image array sets the LSB of audio file sample to 1.

10) The low frequency vector C_2 is embedded in the imaginary part I where I is the reshaped vector of size 705600 using a weighted coefficient α as follows:

$I = I + \alpha C_2$ where $\alpha = 0.01$ (9)

11) The embedded real and imaginary parts are combined together to form the embedded coefficient matrix. The Inverse Fast Fourier Transform (IFFT) of the embedded coefficient matrix is computed to obtain the time domain watermarked signal.

B. Watermarks Extraction Algorithm

1) Take the FFT of the watermarked signal and separate the real and imaginary parts. Let the FFT coefficient matrix obtained be S'_f and the real and imaginary matrices be R' and I'.

2) Reshape the real part R' into a column vector of size 705600 samples and apply the reverse process of multiple bit plane in which the 1st LSB of first sample R'(1), 2nd LSB of second sample R'(2), 3rd LSB of third sample R'(3), 4th LSB of fourth sample R'(4) till all the binary watermark bits are extracted. The retrieved bits are stored in a different array I'_1.

3) Reshape the imaginary part I' into a column vector of size 705600 samples and extract the embedded 2-level DWT transformed watermark signal bits as follows:

$C'_2 = (I' - I)/\alpha$ where I' = Watermarked coefficient imaginary part (10)

I = Original coefficient imaginary part
α = Weighted coefficient = 0.01

4) The obtained C'_2 is transformed into a 2-D array of size 12 x 12 pixels and is passed through 2-level inverse DWT to obtain the original gray scale watermark image of size 48 x 48 pixels.

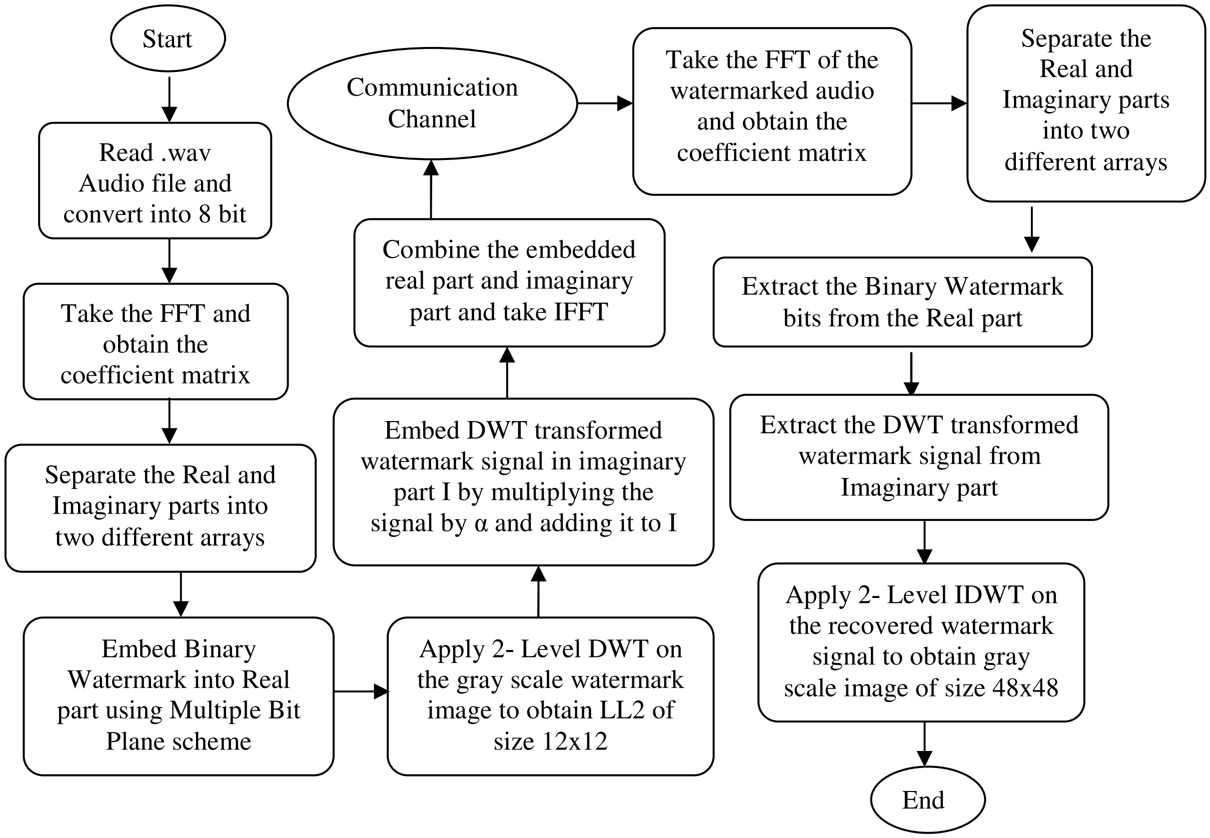


Fig. 3. Flow Chart for the proposed scheme

8. ATTACKS AND COMPARISON MEASURES

Watermarked signal may undergo various signal processing operations. In our research, it is subjected to various attacks such as Gaussian noise, cropping, compression, filtering. The degree of similarity between the original watermark and the watermark obtained from the distorted message signal evaluates the robustness and efficacy of the proposed method. Various types of attacks that can be added to the message in communication channel are as follows:

1) Noise: The signal to noise ratio of the watermarked message signal is changed to 100 dB by addition of Average White Gaussian Noise (AWGN).
2) Cropping: In cropping, some samples of the message signal are removed at random locations and are replaced with noise contaminated samples. In this paper, segments of 441 samples are removed and are replaced with Awgn (10 db) contaminated samples.
3) Compression: The message signal is compressed at 320 kbps and its converted to mp3 format.
4) Filtering: The signal is passed through a low pass filter (LPF) and a high pass filter (HPF) using Butterworth Filter.

The extracted watermark W' and the original watermark W are compared through a Similarity function (SIM). The mathematical expression is as follows:-

$$\mathrm{SIM}(W,W') = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(W(i,j) * W'(i,j))}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(W'(i,j))^2}}$$

where W = Original Watermark
W' = Extracted Watermark
m,n = Number of rows and columns

The correlation factor is given by:

$$\text{Correlation Factor} = \frac{\sum_{i=1}^{MN} f_i f'_i}{\sqrt{\sum_{i=1}^{MN} f_i^2}\sqrt{\sum_{i=1}^{MN} f'^2_i}}$$

where f = Original Watermark
f' = Extracted Watermark
M,N = Number of rows and columns

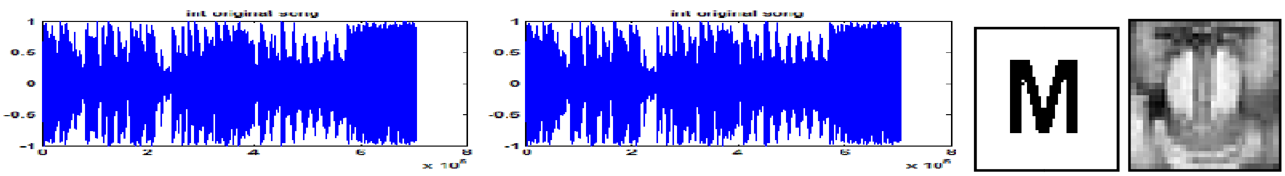
For watermarked audio signal quality, signal to noise ratio (SNR) is defined as follows:

$$SNR(Y,Y^*) = 10\ log_{10}\left(\frac{\sum_{i=0}^{length-1} y^2(i)}{\sum_{i=0}^{length-1}[y(i)-y^*(i)]^2}\right)$$

where Y = Original Audio Signal
Y*= Watermarked Audio Signal

9. EXPERIMENTAL RESULTS

The proposed algorithm has been tested extensively using MATLAB (2013b). In this paper, the host signal used is an audio signal of duration of 16 seconds sampled at frequency of 44100 Hertz with each sample being a 8-bit unsigned value. A binary image of size 50 pixels x 50 pixels and a watermark image of size 48 x 48 pixels are used as the watermark signal. Figure 4 shows the graphical representation of original and watermarked signals and the original watermarks embedded in the host signal. Figure 5 shows the extracted watermarks from the host signal in case of no attack and in case various attacks.



(a) (b) (c) (d)

(a) Original audio (b) Watermarked audio (c) Binary Watermark (d) Gray scale watermark

Fig. 4. Graphical representation of original and watermarked audio signal and the original watermarks

The recovered watermarks when there is no attack and when the host signal is subjected to various attacks, are summarized as follows:-

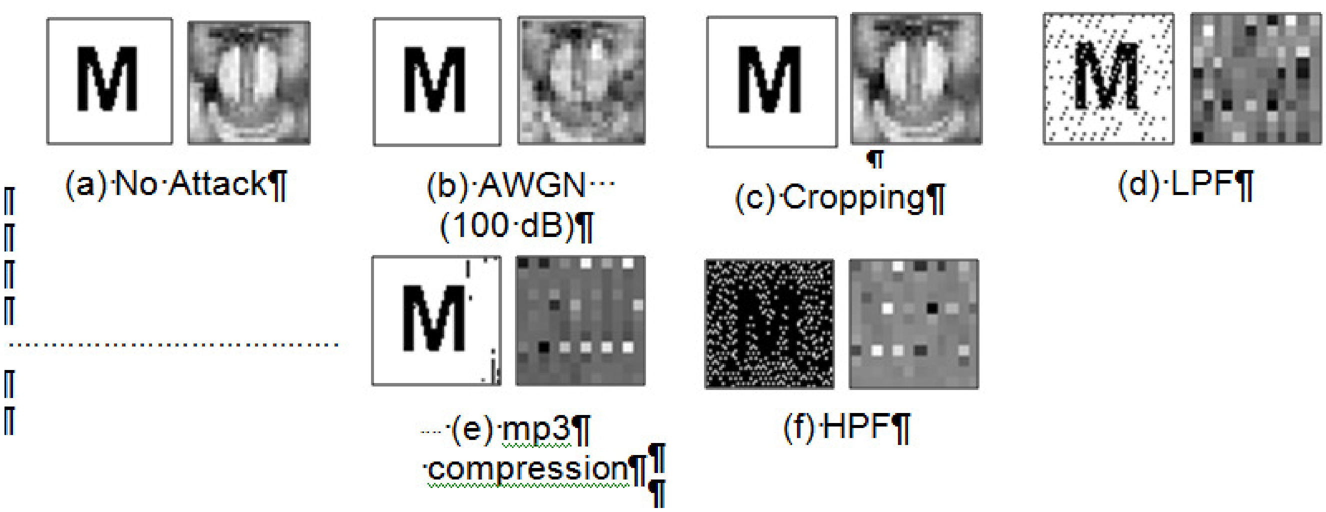


Fig. 5. Recovered Watermarks

It has been observed that at least one of the embedded watermarks is successfully extracted from the watermarked audio signal. The sole purpose of dual watermarking algorithm is that one of the watermarks can be recovered in case of attacks thereby establishing the identity of the owner. Hence, one watermarking method is robust in some cases and other watermarking method is robust in some other cases. The overall performance of the proposed scheme is evaluated under various attacks and either one of the watermarks can be efficiently extracted from the watermarked audio signal even when the audio quality is distorted. This ensures the effectiveness of the applied dual watermarking scheme. The following table summarize the results of various comparison measures.

Table I summarizes the results for Correlation Factor and Similarity Measure respectively between the two recovered watermarks and the original watermarks and Signal to noise ratio between the original audio signal and the attacked watermarked audio signal.

TABLE 1: Results for different comparison measures

Attack Type	CORRELATION FACTOR		SIMILARITY MEASURE (%)		SIGNAL TO NOISE RATIO (dB)
	Watermark 1	Watermark 2	Watermark 1	Watermark 2	
Original Watermark	1.0000	1.0000	100	100	Inf.
No attack	1.0000	1.0000	100	100	94.07
AWGN (100 dB)	1.0000	0.6316	100	63.16	94.73
Cropping	1.0000	0.6342	100	63.42	93.94
Mp3 compression	0.6423	0.1164	64.23	11.64	81.74
Low-Pass Filtering	0.9901	0.5149	99.01	51.49	94.72
High-Pass Filtering	0.6282	0.5103	62.82	51.03	90.82

The results tabulated above have high robustness and the watermarked audio signal has high imperceptibility. The embedded binary watermark 1 is extracted with correlation factor 1 in case of attacks due to AWGN and cropping whereas the gray scale watermark 2 is successfully extracted in case of AWGN, cropping, low pass and high pass filtering. Hence, at least one of the watermarks is extracted with an enhanced correlation factor in case of distortions produced due to various attacks.

10. CONCLUSIONS AND FUTURE SCOPE

This paper presents a robust technique of watermarking in which dual and hybrid watermarks are embedded in the frequency domain. Two watermarks increase the safety and the copyright protection as in case of attacks, either one of the watermark is efficiently recovered thereby establishing the identity of the owner. The gray scale watermark is passed through a 2 stage DWT prior to the embedding process to ensure only required information of the image is embedded. This reduces the payload and increases the SNR In previous works, the watermark is embedded in the time domain amplitude of the host audio signal. Hence, it can be easily distorted as most attacks affect the amplitude of the signal. In this paper, the watermark signal is embedded in the frequency domain and not in the time domain amplitude. Hence, the results are improved as compared to embedding in time domain. As a future scope, the watermark can be embedded in other features of audio signal such as frequency and pitch as they are least affected by distortions.

REFERENCES

[1] Chopra, D. et. al. LSB Based Digital Image Watermarking For Gray Scale Image. IOSR Journal of Computer Engineering (IOSRJCE, 2012, 6, 36-41.

[2] Panda, J.; Kumari, I.; Goel, N. Digital Watermarking of Audio in Time Domain Multiple Bit Plane based on Chaotic Scrambling, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 3, March 2015.

[3] Panda, J.; Kumari, I.; Goel, N. Digital Watermarking of Audio using LSB Substitution, National Conference on Advanced Computing Research, 2015.

[4] Al-Haj, A. et. al. DWT–Based Audio Watermarking. The International Arab Journal of Information Technology, 2011, 8.

[5] Jiansheng, M.; Sukang, L.; Xiaomei, T. A Digital Watermarking Algorithm Based On DCT and DWT. Proceedings of the 2009 International Symposium on Web Information Systems and Applications, 2009, 104-107.

[6] Panda, J. et al. Application of Energy Efficient Watermark on Audio Signal for Authentication, International Conference on Computational Intelligence and Communication Systems, 2011.

[7] Sarker, M.I.S.; Khan, M.I.; Deb, K. FFT-Based Audio Watermarking Method with a Gray Scale Image for Copyright Protection. International Journal of Advanced Science and Technology, 2012 47.

[8] Fallahpour, M.; Megias, D. High capacity robust audio watermarking scheme based on FFT and linear regression. Journal of Innovative Computing, Information and Control, 2012, 8.