Efficient and Robust Audio Watermarking for Content Authentication and Copyright Protection

Neethu V
PG Scholar, Dept. of ECE,
Coimbatore Institute of Technology,
Coimbatore, India.

R.Kalaivani
Assistant Professor, Dept. of ECE,
Coimbatore Institute of Technology,
Coimbatore, India.

Abstract—**Digital audio watermarking aims to embed digital information in the form of multimedia files such as text, image or audio into an original audio signal. The main requirement of audio watermarking is to prove ownership as well as copyright protection. This paper presents efficient audio watermarking based on Wavelet transform. The L level Haar Wavelet transform is performed on the audio signal and the obtained detail coefficients are divided into short frames and the magnitude of the samples are then replaced with the closest Fibonacci numbers. The security of the watermarking technique is further enhanced by adapting cryptographic methods on the embedded secret text. The suggested technique mathematically proves that the average error for each sample is 25%. The fidelity of the technique is also proved mathematically. The experimental outcomes suggest that the method is having high capacity (1kbps to 3 kbps), robustness against various signal processing attacks and no significant perceptual distortion (ODG is around -1).**

Keywords—**Audio Watermarking, fidelity, Fibonacci numbers, Golden Ratio, DWT.**

I. INTRODUCTION

With the rapid development of internet and various communication techniques leads to an increasing demand to protect the digital data from unauthorized access and piracy. The transfer and storage of multimedia data has become very common, so their illegal copy and distribution adversely affected the authors and publishers. Thus nowadays research works are focusing towards multimedia security [1]. Even though traditional encryption techniques offer limited solution, recently attention is focused on watermarking algorithms.

Digital audio watermarking technique is a process by which a watermark in the form of text, image or audio is hidden or embedded into an original audio signal. These embedded data can be later detected or extracted from the marked signal for various applications such as copyright protection, content authentication, finger printing and broadcast monitoring.

Watermarking an audio signal is quite difficult than watermarking an image or video sequence because of the wide range of human auditory system (HAS) as compared to the human visual system (HVS) [2]. The HAS perceives sounds over a range of power greater than 109:1 and a range of frequencies greater than 103:1. The sensitivity of the HAS to the additive white Gaussian noise (AWGN) [3] is high, thus noise in a sound file can be detected as low as 70 dB below the ambient level. On the other hand, HAS contains a fairly small differential range, i.e. loud sounds generally tend to mask out weaker sounds. Additionally, HAS is insensitive to a constant relative phase shift [4] in a stationary audio signal and some spectral distortions interprets as natural, perceptually non-annoying ones [2].

An audio watermarking technique can be classified into two categories: time domain and frequency domain technique. However frequency domain techniques are more effective [5]-[11] since the secret bits are added to the transformed coefficients of the host audio signal, thus maintaining robustness and inaudibility. In frequency domain methods, the Fourier transform is very popular. In Ref. [12] the FFT domain is selected to embed watermarks to take advantage of the translation-invariant property of the FFT coefficients which resist small distortions in the time domain. Better perceptual quality and less computational burden is achieved. Ref. [13] presents watermark in the Empirical Mode Decomposition domain to increase the number of binary data embedded to the audio signal. The challenges are the Constant embedding strength and alsoit cannot support various sampling rates as well as A/D and D/A conversion problems. Ref. [14] presents a time domain watermarking algorithm based on LSB Coding and S-box transformation. Even though it is a simple technique it provides least robustness against various signal processing attacks. Ref. [15] discusses watermarking an image in an audio as a whole or in a segment by segment basis. It enhances security by adapting chaotic encryption on singular value decomposition (SVD) transformed audio signal. [16] proposes watermarking based on Discrete Wavelet Transform (DWT) and singular value decomposition (SVD) but however has a reduced payload capacity.

This paper proposes an efficient audio watermarking that satisfies the requirement of inaudibility, robustness and security. After applying Haar Wavelet transformation, a part of DWT spectrum is selected which is then divided into

frames and each single bit of the encrypted secret message is embedded into each frame. All samples in a frame are modified based on the nearest Fibonacci numbers.

The rest of the paper is organized as follows: section II describes about the requirements of audio watermarking. Section III presents the significance of Fibonacci numbers and golden ratio. Section IV presents the proposed watermarking algorithm. Section V proves fidelity of the technique. In Section VI, the experimental results are shown. Finally, Section VII summarizes the conclusions of this research.

II. REQUIREMENTS OF AUDIO WATERMARKING

An audio watermarking system may have different properties but must satisfy the following basic requirements:

1. Perceptual Transparency: The main requirement of watermarking is perceptual transparency. The embedded watermark containing the owner's information should not degrade the quality of the host signal. The watermark should not be seen by human eye nor be heard by human ear [17].
2. Robustness: The embedded watermark should be unable to remove from the host audio signal even after the watermarked information is exposed to different types of attack. Robustness is one of the major design issues in all watermarking applications. The watermark should be robust against various signal processing attacks includes D/A & A/D conversion, linear & nonlinear filtering, compression and geometric transformation of host audio signal [17].
3. Security: The security of the watermarking system is dependent on the use of private or secret key. The watermark must be strongly resistant against unauthorized detection or from an unwanted agent who wanted to pirate the information [17].
4. Data Rate: The number of watermark that is embedded within a host signal without losing imperceptibility is termed as data payload. For audio, data payload refers to the number of watermark bits that is be reliably embedded within a host signal per unit time, measured in bits per second (bps) [17].

III. GOLDEN RATIO

The sequence 1,1, 2, 3,5,8,13,21,34… is known as the Fibonacci sequence and is named after the Italian mathematician Leonardo of Pisa, also known as Fibonacci. The book Liber Abaci introduced the sequence to western European mathematics. The Fibonacci numbers appear often in mathematics, hence an entire journal is dedicated to their study, the Fibonacci Quarterly. The application of Fibonacci numbers include computer algorithms such as the Fibonacci search technique and the Fibonacci heap data structure, and graphs and Fibonacci cubes used for interconnecting parallel and distributed systems. They are also significant in biological domain such as branching of trees, the arrangement of leaves on stem, the fruit spouts of pineapple, an uncurling of fern and so on.

The equation to produce Fibonacci numbers is shown below:

$$F_n = \begin{cases} 0, & if\, n < 0, \\ 1, & if\, n = 1, \\ F_{n-1} + F_{n-2}, & if\, n > 1. \end{cases} \quad (1)$$

One of the interesting feature of Fibonacci numbers is the ratio between two consecutive numbers [18].

$$F_n = F_{n-1} + F_{n-2} \quad (2)$$

$$\frac{F_n}{F_{n-1}} = \frac{F_{n-1} + F_{n-2}}{F_{n-1}}$$
$$= 1 + \frac{1}{\frac{F_{n-1}}{F_{n-2}}} \quad (3)$$

$$lim_{n\to\infty} \frac{F_n}{F_{n-1}} = \lim_{n\to\infty}\left(1 + \frac{1}{\frac{F_{n-1}}{F_{n-2}}}\right)^1 = 1 + \frac{1}{\lim_{n\to\infty}\left(\frac{F_{n-1}}{F_{n-2}}\right)^1}$$
$$(4)$$

$$If\ lim_{n\to\infty} \frac{F_n}{F_{n-1}} = \varphi\ ;\ then\ \varphi = 1 + \frac{1}{\varphi} \quad (5)$$

$$\varphi^2 - \varphi - 1 = 0\ ; \quad (6)$$

$$\varphi = \frac{1 \pm \sqrt{5}}{2} \quad (7)$$

Possible values of Ψ are 1.618 and -0.618, since Ψ is positive it is 1.618. This Ψ is referred to as Golden Ratio which is an irrational number. It obtained its name from the Golden Rectangle, whose sides are in the proportion of the golden ratio. Each Fibonacci number can be represented by the Golden Ratio as shown below:

$$F_n = \frac{\varphi^n - \bar{\varphi}^n}{\sqrt{5}} \quad (8)$$

Where $\bar{\varphi}$ is the negative solution.

IV. PROPOSED WATERMARKING ALGORITHM

In the proposed watermarking scheme, the following algorithm is used to embed the watermark which is in the form of secret bit stream into the selected DWT coefficients. The discrete wavelet transforms (DWT) presents time-frequency representation of the input signal. The inputted audio signals is decomposed using L level Haar DWT transformation which produces 2^L signals [19].Using DWT, the fine details of the signal can be separated and also reconstruction can be carried out with greater accuracy. The frequency band and frame size

are the two parameters which are used to adjust the properties of the watermarking system such as capacity, perceptual distortion and robustness.Increase in frequency band leads to increase in capacity and distortion but decrease in robustness while increase in frame size increase in robustness thus decreasing capacity.

A. Emedding

Before embedding the watermark, the two parameters need to be kept constant. Considering MP3 cutoff frequency which is higher than 16 kHz, the high frequency band is set to 16kHz or lower. The low frequency band is normally adjusted to set the frequency band, however whose default value is 12 kHz. The frame size is set to 5. The embedding steps are as follows:

1. Sample the original audio signal at a sampling rate of 44100 samples per second. If the size of the audio file is large, then it needs to be divided into blocks of shorter length, and the watermark need to be added to each block independently.
2. Perform a four level DWT transformation. This produces 5 multi-resolution sub bands: D1,D2, D3,D4 and A4. D represents detail sub bands and A4 represents the approximation sub bands.
3. The samples in the selected frequency band need to be taken and divided into frames of size d.
4. Input the watermark and perform encryption using a specific key which is known only to the content owner and the specified receiver. Cryptographic technique is mainly adapted to improve the security of the system. The embedded watermark is the XOR sum of the real watermark and the key.
5. The Fibonacci sequence used for embedding is
   F = { 1, 2, 3,5,8,13,21,34…}
   For each DWT coefficient find the closest two Fibonacci numbers. Let $\{F_{k,p}\}$ represents the corresponding k^{th} Fibonacci number which is lower than the magnitude of pth DWT sample.
6. Replace the DWT coefficients with that of the closest Fibonacci number based on the following condition

$$F_i' = \begin{cases} F_{k,p} \, , & if \ kmod2 = S_m \\ F_{k+1,p} \, , & if \ kmod2 \neq S_m \end{cases} (9)$$

S_m represents the mth secret bit embedded
$$m = \left[\frac{p}{d}\right] + 1,$$
Where$\left[\frac{p}{d}\right]$ represents the largest integer value lower than or equal to $\frac{p}{d}$.

7. Compute inverse DWT on the marked DWT coefficients to obtain the marked audio signal.

B. Extraction

The parameters, frame size and frequency band need to be known at the receiver end for extraction process. However original audio signal is not required hence the detector is known to be blind. Following steps are done at the receiver to detect the secret bit:

1. Compute the DWT coefficients of the marked audio signal.
2. Select the samples in the particular frequency band and divide it into frames of size d.
3. Find the closest Fibonacci number of the DWT sample. In case of two equally close number, select the lowest Fibonacci number.
4. The watermark bit can be extracted using the following equation.

$$E_n' = \begin{cases} 0, & if \ kmod2 = 0 \\ 1, & if \ kmod2 = 1 \end{cases} (10)$$

E_n'represents the bits extracted from each sample.
5. Divide the bits in terms of frames of size d. If the number of "1" is greater than "0" in a frame the watermarked bit is 1.
6. Extract the secret bit by XORing the watermarked bit with that of key.

The security of algorithm is based on the security of the knowledge of frame size and frequency band. If the attacker has guessed these values then security is enhanced by encrypting the secret text with the key. The use of DWT magnitudes results in more robustness against attacks.

V. DISCUSSION

Fibonacci numbers are used to keep the modification error in an acceptable range. Consider the original sample to be S, then the closest Fibonacci number is given by

$$F_k \leq S \leq F_{k+1} \quad (11)$$
The distance to each is given by

$$E_1 = S - F_k (12)$$
$$E_2 = F_{k+1} - S \quad (13)$$

The ratio between two Fibonacci numbers which is used to find the error ratio is given by,
$$R_k = \frac{F_{k+1}}{F_k} (14)$$

Where n = 1,2,3,……
$R_1 = 2, R_2 = 1.5, R_3 = 1.66, R_4 = 1.6, R_5 = 1.625$….

$$Maxerror = (F_{k+1} - F_k) = (R_k F_k - F_k) = (R_k - 1)F_k$$

$$(15)$$
Thus the maximum distortion introduced in the magnitude of an DWTsample lies between 0.38 and 0.61
Proof:
1. If S is converted into F_{k+1}

$$Maxerror = \frac{max \ error}{F_{k+1}} = \frac{(R_k - 1)F_k}{F_{k+1}} = \frac{(R_k - 1)F_k}{R_k F_k}$$

$$= \frac{R_{k-1}}{R_k} \qquad (16)$$

2. If S is converted into F_k

$$Maxerror = \frac{\max error}{F_k} = \frac{(R_k - 1)F_k}{F_k} = R_k - 1 \quad (17)$$

Assuming the value of R_k = 1.61, the maximum error rate lies between 0.38 and 0.61. thus the average error rate is proved to be 0.50. However, if the DWT samples have uniform distribution then the average error rate is only 0.25% [12].

VI. EXPERIMENTAL RESULTS

A music file of length 221184 samples (5 seconds) is sampled at 44.1 kHz with 16 bits per sample and two channels. The experiment is performed on each channel separately and the performance of the proposed algorithm is evaluated. The smaller level DWT influences the robustness of the watermark, however the larger ones results in calculation complexity, hence 4 level DWT is performed. The waveform of the original audio signal is shown in Figure 1 and the watermarked signal in figure 2.

- Fidelity

Fidelity refers to the closeness between the undistorted original audio signal and distorted watermarked audio signal. SNR metric is used for the subjective evaluation. The watermark audio signal should maintain more than 20dB SNR according to the recommendations of International Federation of the photographic industry (IFPI). An SNR value of 69 dB is obtained in the proposed algorithm.
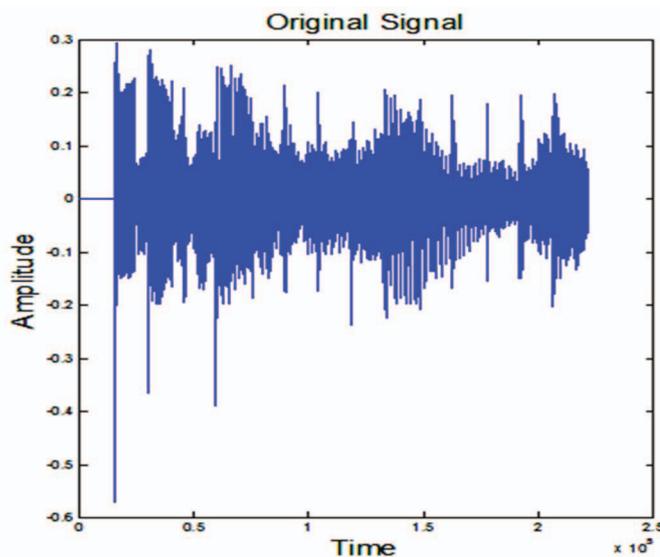

Fig. 1. Original Audio Signal.

Fig. 1 and fig.2 shows the original and watermarked audio signal which is visually indistinguishable. This proves the objective analysis of fidelity
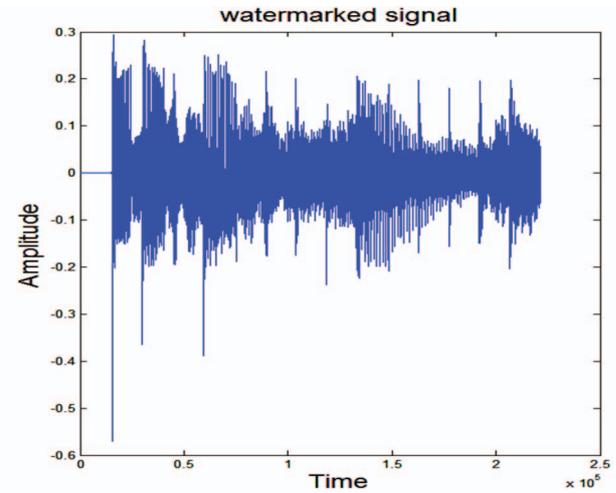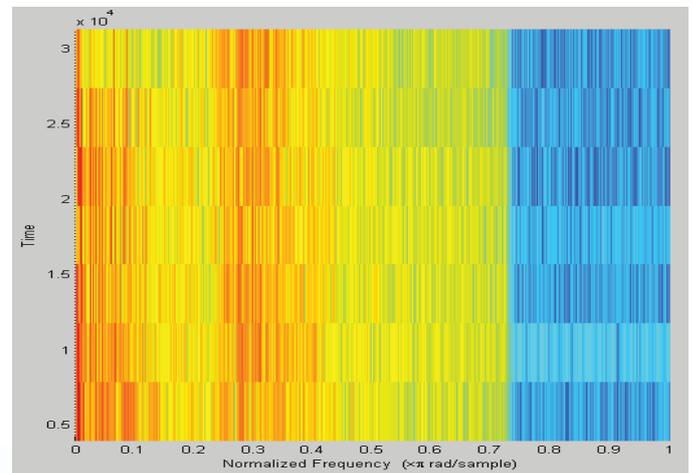

Fig. 2. Watermarked Audio Signal
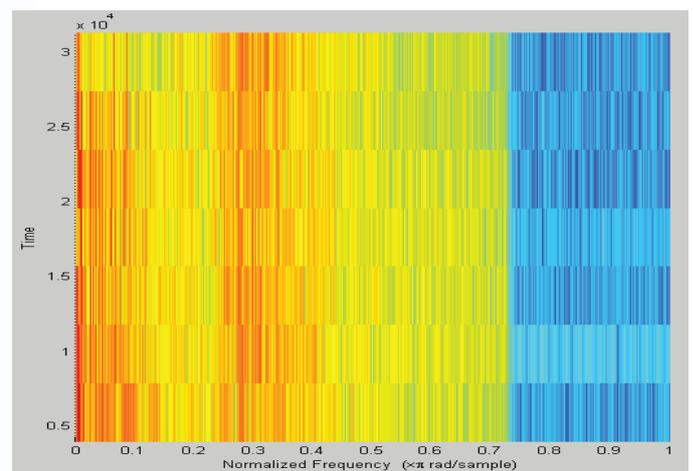

Fig.3. Spectrogram of original audio signal

Fig. 4. Spectrogram of watermarked signal

Fig.3 and Fig.4 represents the spectrogram of the original audio and watermarked signal. Spectrogram is the visual representation of the spectrum of frequencies in the audio signal with time. From the figure it is evident that the distributions of the frequencies in the original and watermarked audio signal have very close resemblance.

- Imperceptibility

Subjective listening test are performed for perceptual quality assessment. ODG is appropriate measurement of audio distortions since it is assumed to provide an accurate model of the subjective difference grade (SDG). ODG = 0 means no degradation and ODG = -4 means an annoying distortion. The ODG values of the watermarked signal are observed between -1 and 0 which reveals their good quality. Five participants were selected to hear the original and watermarked audio signal and were asked to report the dissimilarities between the two. The output of this test is an average of the quality ratings called Mean Opinion Score (MOS). Table1 shows the different MOS criterion and the imperceptibility for the watermarked output is 5.This result revealed that the output of the marked signal is good.

TABLE 1

MOS CRITERION

Score	Watermark Imperceptibility
5	Imperceptibility
4	Perceptibility but not annoying
3	Slightly annoying
2	Annoying
1	Very Annoying

TABLE II

SNR AND ODG BETWEEN ORIGINAL AND WATERMARK AUDIO

Audio file	SNR (dB)	ODG
Jazz	67.79	-0.95
Pop	64.11	-0.88
Classic	65.8	-0.91
Rock	62.34	-0.67

Table II shows the various SNR and ODG values for the different kinds of the audio signals including jazz, pop, classic and rock sampled at 44.1 kHz.

A. Robustness

To evaluate the performance of the proposed audio watermarking algorithm, it needs to be subjected to different kinds of attacks like add noise, amplify, echo, invert and other common attacks. Different attacks performed are:

- Noise: White Gaussian Noise (WGN) can be added to the watermarked signal.

- Compression:Watermark signal can be compressed and then decompressed using MP3.

- Filtering: Weiner filter can be used to filter the watermarked audio signal.

- Cropping: Samples can be removed from the watermarked audio signal and replaced with samples of audio signal with noise.

After performing each of the attacks BER needs to be calculated in each case. BER reflects the certainty of detection of the embedded watermark and is 0 for most of the attacks in the proposed algorithm.

B. Capacity

The proposed algorithm gives capacity ranging from 1kbps to 3kbps with the variation in frequency band and frame size. The payload can also be measured after subjecting to MP3 attack.
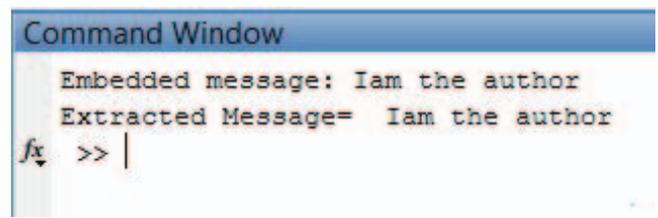

Fig. 5. Extracted Output

Fig. 5 shows the screenshot of the command window. Simulations are performed in MATLAB and from the figure it is evident that the embedded message is correctly extracted at the receiver end.

TABLE III
COMPARISION

Algorithm	Capacity(bps)	Imperceptibility in SNR(dB)	Imperceptibility Score
[12]	683-3k	35	5
[13]	46.9-50.3	26.38	5
[15]	Not reported	27.13	5
[16]	Not reported	28.55	5
[19]	Not reported	34.4935	5
Proposed	1k-3k	69	5

Table III provides the comparison of the existing five watermarking techniques with the proposed scheme. The parameters used for comparison is payload capacity measured in bits per second and fidelity measured in terms of SNR and MOS. From the results it is proved that the watermark does not affect the quality of the signal and also has a good capacity and highest SNR.

VII. CONCLUSION

In this paper, efficient audio watermarking has been presented which is based on transform domain approach. This technique is blind, since original audio signal is not required at the receiver end for the detection process. The suggested method guarantees that the maximum change of each DWT sample is less than 50% and average error in each sample is 25%. The security of the embedded information is enhanced by the encrypting the message with the key. The perceptual quality, capacity and robustness are the parameters which are measured by changing the frequency band and frame size. Analysis shows that the proposed algorithm is very efficient providing high capacity, significant perceptual distortion and provides robustness against common signal processing attack.

This work can be extended to watermark an image and also use other transform domain techniques individually or as a hybrid and measure the various performance parameters. The future watermarking techniques will be equipped with intelligence that reveals the content of audio file, the distribution channel, to whom it was distributed and so on.

REFERENCES

[1] H. J. Kim, "Audio watermarking techniques," in Proc. Pacific Rim Workshop Digital Steganogr., 2005, pp. 1–17.

[2] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," in Proc. SPIE, vol. 2420, San Jose, CA, Feb. 1995, p. 40.

[3] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Trans. Image Process. 6, 1673–1687,1997.

[4] I.J. Cox, M.L. Miller, A.L. Mckellips, Watermarking as communication with side information, Proc. IEEE 87,1127–1141,1991.

[5] Mehdi Fallahpur, David Megias, "Robust high capacity audio watermarking based on FFT amplitude modification", IEICE Trans. Inf. Syst., vol. E93-D,no.01,pp 87-93,Jan. 2010.

[6] Mehdi Fallahpur, David Megias, "DWT-based high-capacity audio watermarking based", IEICE Trans. Fundam.Electron., Commun. Comput. Sci., vol. E93-A,no.01,pp 331-335,Jan. 2010.

[7] Mehdi Fallahpur, David Megias, "High-capacity audio watermarkingusing the frequency band of the wavelet domain", in Mulitimedia Tools and Applications. New York,NY,USA : Springer, 2011, vol.52, pp. 485-498.

[8] Mehdi Fallahpur, David Megias, "Secure logarithmic audio watermarking based on the human auditory system",Mulitimedia Syst., 2013, DOI : 0.1007/s00530-013-0325-1,ISSN. O942-4962.

[9] Mehdi Fallahpur, David Megias, " High-capacity robust audio watermarking scheme based on FFT and linear regression", Int. J. Innovative Comput., Inf. Control, vol. 8, no. 4, pp 2477-2489,April 2012.

[10] S.T.Chen,G.D.Wu and H.N.Huang,"Wavelet domain audio watermarking scheme using optimization-based quantisation," IET Signal Process., vol. 4, no. 6, pp. 720-727,2010.

[11] S.T.Chen,G.D.Wu and H.N.Huang,"Energy proportion based scheme for audio watermarking," IET Signal Process., vol. 4, no. 5, pp. 576-587,2010.

[12] Mehdi Fallahpur, David Megias, "Audio Watermarking based on fibonacci numbers", IEEE Transactions on Audio, Speech and Language Processing, Vol.23, No. 8, pp.1273-1282, 2015.

[13] Kais Khaldi, Abdel- Ouahab Boudraa, "Audio Watermarking Via EMD",, IEEE Transactions on Audio, Speech and Language Processing , Vol.21, No. 3, pp. 675-680, 2013.

[14] Iqtadar Hussain, "A Novel approach of Audio Watermarking based on S-box transformation",, Elsevier Mathematical and Computer Modelling, Vol 57 , pp. 965-969, 2013.

[15] Waleed Al-Nuaimy, Mohsen A.M, "A SVD Audio Watermarking approach using chaotic encrypted images",, Elsevier Digital Signal Processing, Vol 21, pp. 764-779, 2011.

[16] Ali Al-Haj et al," Hybrid SVD-DWT Audio Watermarking", IEEE Transactions on Audio, Speech and Language Processing, Vol.62, No. 8, pp.525-529, 2010.

[17] L.Wei, Y.Yi-Qun, L.Xiao-Qiang, X.Xiang-Yang and L.Pei-Zhong,"Overview of digital audio watermarking",J.Commun.,vol.26,no.2,pp.100-111,2005.

[18] R. A. Dunlap, The golden ratio and fibonacci numbers. Hackensack,NJ,USA:WorldScientific,1997.

[19] N.V Lalitha, Ch.Srinivasa Rao, P.V.Y.JayaSree," DWT-Arnold Transform Based Audio Watermarking", IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics, pp 196-199, 2013