DWT BASED AUDIO WATERMARKING SCHEMES:A COMPARATIVE STUDY

N.V.Lalitha[1]Ch. Srinivasa Rao[2]and P.V.Y.JayaSree[3]

[1]Department of Electronics and Communication Engineering, GMR Institute and Technology, Rajam, A.P, India.
[2]Department of Electronics and Communication Engineering, JNTU-K University, Vizianagaram, A.P, India.
[3]Department of Electronics and Communication Engineering, GIT, GITAM University, Visakhapatnam, A.P, India.

ABSTRACT

The main problem encountered during multimedia transmission is its protection against illegal distribution and copying. One of the possible solutions for this is digital watermarking. Digital audio watermarking is the technique of embedding watermark content to the audio signal to protect the owner copyrights. In this paper, we used three wavelet transforms i.e. Discrete Wavelet Transform (DWT), Double Density DWT (DDDWT) and Dual Tree DWT (DTDWT) for audio watermarking and the performance analysis of each transform is presented. The key idea of the basic algorithm is to segment the audio signal into two parts, one is for synchronization code insertion and other one is for watermark embedding. Initially, binary watermark image is scrambled using chaotic technique to provide secrecy. By using QuantizationIndex Modulation (QIM), this method works as a blind technique. The comparative analysis of the three methods is made by conducting robustness and imperceptibility tests are conducted on five benchmark audio signals.

KEYWORDS

Discrete Wavelet Transform (DWT), Double Density DWT (DDDWT) and Dual Tree DWT (DTDWT), Quantization Index Modulation (QIM)

1. INTRODUCTION

The swift growth in multimedia technology and the usage of internet, the major problem facing by the owners is unauthorized copying, transmission and distribution of multimedia content.The most common solutionfor protection of copyright is digital watermarking [1, 2]. Watermarking is the process, in which watermark content is embedded into the digital content. Digital content may be audio, image or video. Developing audio watermarking algorithms are not that much easy [3,4] compared to image and video watermarking,. Firstly, Human Auditory System (HAS) is much sensitive than Human Visual System (HVS). Therefore, even small changes in audio are also recognized by the human ear. Secondly, video files are large compared to audio files in terms of size. Hence, data hidden in audio files is quietly large compared with the image or video and this high payload tends to degrade the audio quality. Therefore, trade-off exists between robustness and imperceptibility.

DOI: 10.5121/ijci.2016.5420

Recently, several audio watermarking algorithms are developed. Most of the algorithms are based on either time domain [5,6] or transform domain [7,8,9,10,11]. Watermarking in time domain is easier to implement and needs less computational resources thanwatermarking in transform domain [3,8] but, it is less robust against common signal processing attacks when compared to transform domain watermarking. Generally, Fast Fourier Transform (FFT)[11], Discrete Cosine Transform (DCT) [9], and Discrete Wavelet Transform (DWT)[10] are explored for transform domain audio watermarking.

Still, there is a need for robust and high secured audio watermarking algorithms. In this paper, the chaotic Gaussian map is used to encrypt the watermark image. The Logistic chaotic sequence is used to develop synchronization code. Then, the watermark is embedded in DWT/DDDWT/DTDWT coefficients of audio signal using QIM.

2. METHODS

2.1. Discrete Wavelet Transform (DWT)

The analysis filters (a1 and a2) decomposes the input signal x(n) into two sub-bands i.e., low-pass frequency band (c(n)) and high frequency band (d(n)) and each of which is then down-sampled by 2. The two sub-bands (c(n) and d(n)) are up-sampled by 2 and the synthesis filters (s1 and s2) combines the two sub-bands to acquire a single signal y(n)[12] shown in Figure 1.

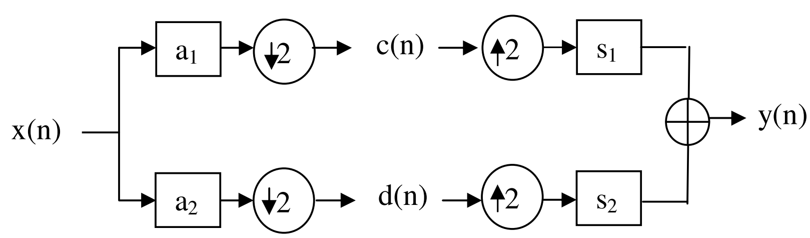

Figure 1. DWT decompose and combined process.

2.2. Double Density DWT (DDDWT)

Double –Density DWT [12] makes use of two distinct wavelets and a single scaling function. The analysis filters decomposes the x(n) signal into three bands, and every sub-band is down-sampled by 2. The filter bank for analysis consists of one low-pass filter (a1) and two high pass filters (a2 and a3). The synthesis filter bank consists of one low-pass filter (s1) and two high pass filters (s2 and s3). These3 sub-band coefficients pass through the system are up-sampled by two, synthesized and then combined to develop the signal y(n) shown in Figure 2.

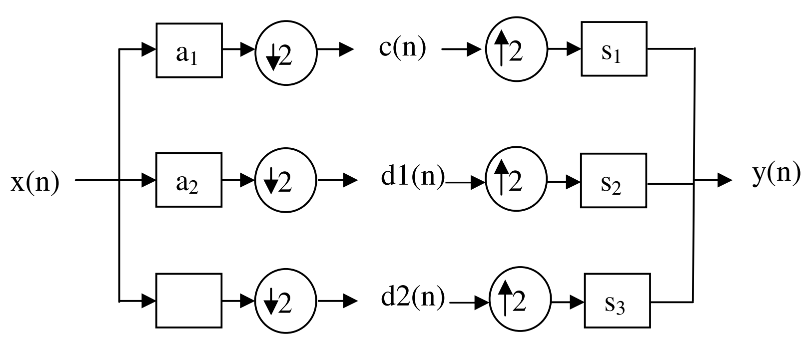

Figure 2. DDDWT decompose and combined process.

2.3. Dual Tree DWT (DTDWT)

The dual tree DWT of a signal x(n) is a parallel combination of two DWTs [13]. Therefore, it is 2-times expensive than DWT. The filters are chosen in a way that the upper DWT can be inferred as real part of the wavelet and lower DWT can be inferred as imaginary part of wavelet [14] and is shown in Figure 3.

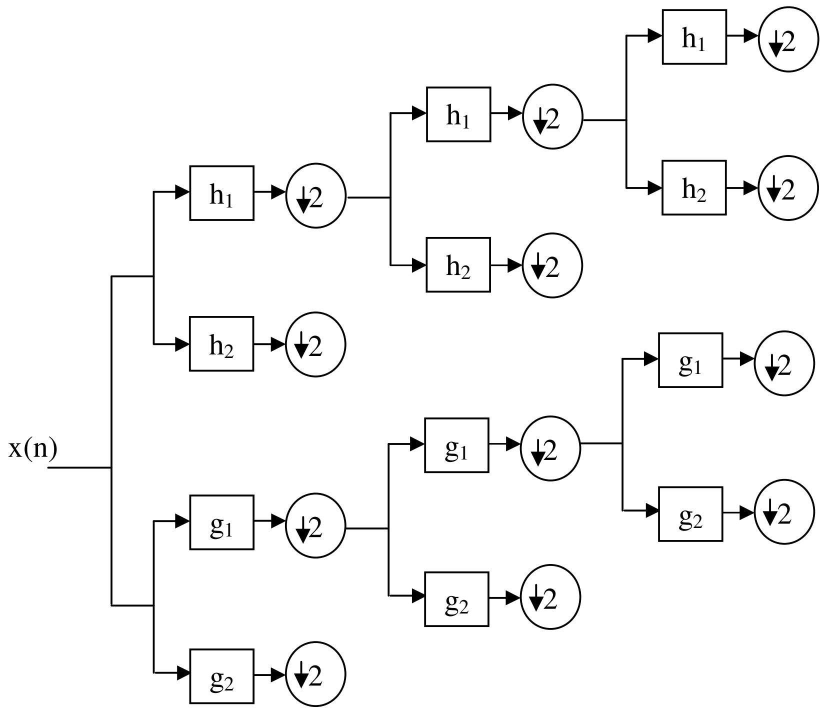

Figure 3. DTDWT decompose and combined process.

3. SYNCHRONIZATION CODE GENERATION AND INSERTION

The synchronization code [7,8,9] is used to resist the de-synchronization attacks. Desynchronization attack means the watermark cannot be recognized from the watermarked audio because of lack of synchronization. Desynchronization attacks are cropping, shifting and MP3 compression, they will change the audio signal length, which leads to unsuccessful extraction of the watermark.To overcome this problem, exact location of the watermark should be identified before the extraction process. For synchronization code generation, the logistic chaotic sequence is used, that is defined as:

$$y_{n+1} = \gamma y_n(1 - y_n) \qquad (1)$$

Where y_n is the initial value that is from 0 to 1, γ is the real parameter.

Synchronization code is generated using eq(1) based on the following condition.

$$S_n = \begin{cases} 1, & if y_n > 1/2 \\ 0, & otherwise \end{cases} \tag{2}$$

The host audio A is divided into two parts A_S and A_w. Synchronization code that is generated from the eq(2) is hosted into the first part of audio signal A_S with length LS is embedded as follows:

$$A'_S(n) = \begin{cases} round\left(\frac{A_S(n)}{\delta}\right) * \delta, & if S_n = 0 \\ (floor(\frac{A_S(n)}{\delta}) * \delta) + \frac{\delta}{2}, & if S_n = 1 \end{cases} \tag{3}$$

where δ is the embedding strength.

Embedded and attacked watermarked audio signal is also split into two parts. From first part of watermarked signal A''_S synchronization code will be detected with following condition.

$$S'_n = \begin{cases} 0, & if \delta/4 \leq mod(A''_S(n), \delta) < 3\delta/4 \\ 1, & otherwise \end{cases} \tag{4}$$

4. WATERMARK EMBEDDING AND EXTRACTION

4.1. Pre-processing of a Watermark

To improve the security and robustness, watermark image must be pre-processed by using chaotic scrambling technique. Gaussian map [11] is one of the chaotic encryption methods. Gaussian map chaotic encryption technique is defined as:

$$z_{n+1} = e^{(-\alpha(z_n)^2)} + \beta \tag{5}$$

Where z1 is the initial value that ranges from 0 to 1. α and β are the real parameters.

$$v_n = \begin{cases} 1, & if z_n > Th \\ 0, & otherwise \end{cases} \tag{6}$$

WhereTh is the predefined threshold. Two dimensional binary watermark is converted into a vector w_n of size M X M. This w_n is encrypted by v_n using following condition:

$$G_n = XOR(w_n, v_n) \tag{7}$$

4.2. Watermark Concealing Procedure

The watermark concealing procedure is given in Figure 4 . In this procedure, total audio signal is segmented into two parts. The synchronization code is insert in audio signal first part to overcome the de-synchronization attacks. The audio signal second part is used to host the pre-processed watermark image.

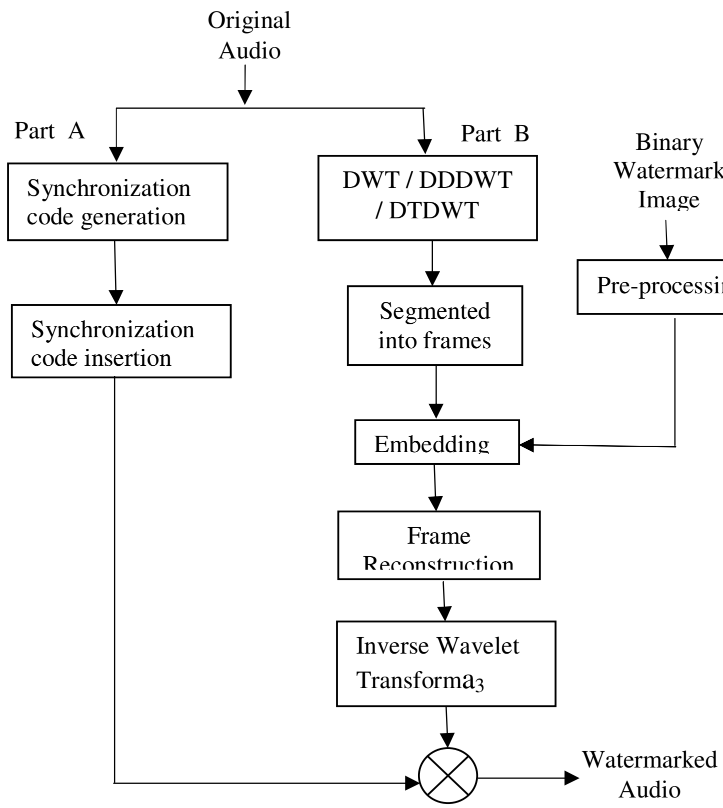

Figure 4. Flowchart of watermark embedding process.

The concealing procedure is detailed as follows:

Step 1: Apply DWT/DDDWT/DTDWT on second part of audio signal.
Step 2: Wavelet coefficients are segmented into frames, and number of frames must be greater than the watermark size.
Step 3: The pre-processed watermark is embedded into each frame using the following rule.

$$F_w^{'}(n) = \begin{cases} round\left(\frac{F_i(n)}{Q}\right) * Q, & if G_n = 0 \\ (floor(\frac{F_i(n)}{Q}) * Q) + \frac{Q}{2}, & if G_n = 1 \end{cases} \quad (8)$$

where Q is the embedding strength.
Step 4: Reconstruct the modified frames.
Step 5: Apply inverse wavelet transform on watermarked audio.

4.3. Extraction Algorithm

The process of extraction is the exact reverse process of concealing process and the algorithm is given below:

Step1: Apply DWT/DDDWT/DTDWT on the second part of attacked watermarked audio signal.
Step2: Wavelet coefficients are segmented into frames.
Step3: Binary encrypted watermark vector is extracted from each frame by using following equation.

$$g_n^{'} = \begin{cases} 0, & if Q/4 \leq mod(F_w^{''}(n), Q) < 3Q/4 \\ 1, & otherwise \end{cases} \tag{9}$$

Step4: The decryption process is same as encryption to determine the binary watermark sequence.
Step5: Finally, convert the one dimensional extracted and decrypted binary sequence into two dimensional watermark image of size M X M.

5. SIMULATION RESULTS

The experimental results give the comparative analysis of the three methods. The performance of the three methods iscompared in terms of robustness, imperceptibility and payload. The experiment is carried on 5 different types of 16-bit audio signals in the .WAV format with the sampling rate 44.1 kHz. Each audio is of 10sec duration.

Binary image of 64 X 64 size is used as a watermark. For increasing the security of the watermark, a Gaussian map chaotic encryption technique is used. Figure 5 illustrates Original and encrypted watermark images.

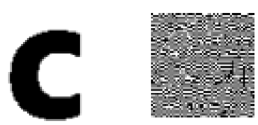

Figure 5. Original watermark and its encrypted watermark images.

5.1. Imperceptibility Test

The audio signal quality should not be degraded upon embedding. The two approaches to perform the perceptual audio quality evaluation [15]. i) Objective test by perceptual evaluation of audio signal ii) Subjective listening test based on HAS.

i) Objective evaluation test:

To evaluate the objective quality, SNR metric is used. International Federation of the Phonographic Industry (IFPI) quotes that watermarked audio should have SNR more than 20dB [8]. SNR Vs Quantization step for three methods are shown in Figure 6.

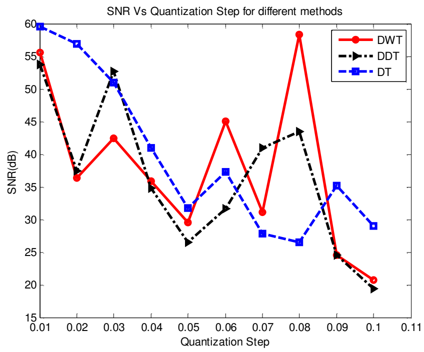

Figure 6. SNR Vs Quantization step

	DWT	DDDWT	DTDWT
Audio-1	31.1205	41.0349	27.7986
Audio-2	42.311	30.6061	27.2856
Audio-3	41.2256	27.0774	53.433
Audio-4	58.0209	41.3026	48.2897
Audio-5	29.8392	36.1878	36.0735
Average	40.5034	35.2417	38.5760

Table 1. SNR in dB for benchmark audio

Table 1 shows the SNR values and their average SNRs for different classes of benchmark audio signals at Q=0.07 are above 20dB and hence meets IFPI requirement.

ii) Subjective Listening Test:

The SNR measure is not sufficient to measure imperceptibilty [8]. Therefore, subjective listening test is also important to evaluate the imperceptibility. Subjective Difference Grade (SDG) is a popular method to evaluate the watermarked audio quality [11]. Table 2 shows the SDG ranges, which is from 5.0 to 1.0. This listening test is performed with ten listeners. Subjects are listened original and watermarked audio signals and they report if any variation is identified between two signals using SDG. The average SDG values are also called as Mean Opinion Score (MOS). The MOS values for DWT,DDDWT and DTDWT is 4.5, 4.8 and 4.7 respectively at Q=0.07.

Table 2. SDG Ranges

Report by subject	**Quality**	**Grade**
Imperceptible	Excellent	5
Perceptible, but not annoying	Good	4
Slightly annoying	Fair	3
Annoying	Poor	2
Very annoying	Bad	1

5.2. Robustness Test

Robustness of this scheme is evaluated with the below attacks on watermarked audio.

i) Resampling: The watermarked audio is resampled to 22.05 kHz, 11 kHz and 8 kHz and sampled back to 44.1 kHz.
ii) Re-quantization: Quantized down to 8-bit and re-quantized back to 16-bit.
iii) Noise: Added with random noise of 30dB signal.
iv) Low-pass Filtering: Cut-off frequency of 20 kHz is applied.
v) Echo addition: 10 ms and 1% decay of echo signal is added.
vi) MP3 Compression: 128 kbps and 256 kbps MPEG compression is applied to the watermarked audio signal and then decoded back to the .WAV format.

vii) Additive Noise: Additive Gaussian Noise with 50 dB and 60 dB.
viii) Cropping: 1000 samples of the watermarked audio signal are made zero at beginning, middle and ending parts.
ix) Signal Addition: Beginning samples are added with original audio samples.
x) Signal Subtraction: Watermarked audio signal beginning samples are subtracted with original audio samples.

For comparison of original watermark and extracted watermark, Bit Error Rate (BER) and Normalized Correlation (NC) are used.

$$\text{BER} = \frac{\text{Number of error bits}}{\text{Number of total bits}} \quad (10)$$

$$NC = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\sum_m \sum_n (A_{mn} - \bar{A})^2 \sum_m \sum_n (B_{mn} - \bar{B})^2}} \quad (11)$$

Table 3 shows BER and NC for all mentioned signal processing attacks for three methods at Q=0.07.

Table 3. BER and NC values for signal processing attacks.

Method	**DWT**		**DDDWT**		**DTDWT**	
Signal Processing Attack	**BER**	**NC**	**BER**	**NC**	**BER**	**NC**
Without attack	0	1	0	1	0.0002	0.9994
Resampling(22.05kHz)	0.0007	0.9982	0	1	0.1182	0.7316
Resampling(11kHz)	0.1741	0.6096	0.1528	0.6508	0.3726	0.2303
Resampling(8kHz)	0	1	0	1	0.0012	0.9971
Re-quantization	0	1	0	1	0.0447	0.8954
Noise	0	1	0	1	0.0059	0.9861
Filtering	0	1	0.0002	0.9994	0.0269	0.9363
Echo addition	0	1	0.0002	0.9994	0.0203	0.952
MP3 Compression (256)	0	1	0	1	0.0063	0.9848
MP3 Compression (128)	0.0004	0.9988	0.0012	0.9971	0.0354	0.9167
Additive Noise (50dB)	0	1	0	1	0.0591	0.863
Additive Noise (60)	0	1	0	1	0.0146	0.9651
Cropping (middle)	0	1	0	1	0.0002	0.9994
Cropping (end)	0	1	0	1	0.0002	0.9994
Cropping (front)	0.0022	0.9948	0.0022	0.9948	0.0024	0.9942
Signal Addition	0.002	0.9953	0.0022	0.9948	0.0022	0.9948
Signal Subtraction	0.002	0.9953	0.0022	0.9948	0.0024	0.9942

6. CONCLUSIONS

The performance of DWT based audio watermarking schemes viz., DWT, DDDWT and DTDWT is analyzed. SNR is above 20 dB for all the three schemes. The watermarked signal is tested against various signal processing attacks for different classes of audio signals and the performance parameters BER and NC are obtained. The parameters shows that DDDWT outperforms DTDWT for different values of quantization step. Also, DDDWT performance is almost nearer to DWT scheme.

REFERENCES

[1] I.Cox, M.Miller, J.Bloom, J.Fridrich, T.Kalke, “Digital Watermarking and steganography”, Second Edition, Morgan Kaufmann, Burlington, MA, 2007.

[2] M.Swanson, M.Kobayashi, A.Tewfik, “Multimedia data embedding and watermarking techniques”, Proc.IEEE 86 (1998) 1064-1087.

[3] E.Ercelebi, L.Batakci, “Audio watermarking scheme based on embedding strategy in low frequency components with a binary image”, Digital Signal Processing, Vol.19, pp 265-277, 2009.

[4] M.D.Swanson, B.Zhu, A.H.Tewfik, L.Boney, “Robust audio watermarking using perceptual masking”, Signal processing, 66(3), pp.337-355, 1998.

[5] Lie W.N, Chang L.C, “Robust high quality time domain audio watermarking based on low frequency amplitude modification”, IEEE Transactions on Multimedia, 8(1), pp.46-59, 2006.

[6] P.Basia, I.Pitas, N.Nikolaidis, “Robust audio watermarking in the time domain”, IEEE Transactions on Multimedia, 3(2), pp. 232-241, 2001.

[7] X.Y.Wang, H.Zhao, “A novel synchronization invariant audio watermarking scheme based on DWT and DCT”, IEEE Transactions on Signal Processing, 54(12) pp. 4835-4840,2006.

[8] Lei, I.Y.Soon, F.Zhou, Z.Li, H.Lei, “A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition”, Signal Processing, Vo. 92, pp. 1985-2001, 2012.

[9] B.Y.Lei, I.Y.Soon, Z.Li, “Blind and robust audio watermarking scheme based on SVD-DCT”, Signal Processing, Vo. 91, pp. 1973-1984, 2011.

[10] V.Bhat K, I. Sengupta, A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain", Digital Signal Processing 20 (2010) 1547-1558.

[11] P.K.Dhar, T.Shimamura, “Audio watermarking in transform domain based on singular value decomposition and Cartesian-Polar Transformation”, International Journal of Speech Technology, Vol.17, pp.133-144, 2014.

[12] Ivan W.Selesnick,“The Double Density DWT”, http://eeweb.poly.edu/iselesni/double/double.pdf

[13] N.Kingsbury, “Image Processing with complex wavelets”, Phil.Trans. R. Soc. London A, 1997.

[14] N.G. Kingsbury, “The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters”, in proceedings of Eighth IEEE DSP Workshop, Salt Lake City, UT, Aug,9-12, 1998.

[15] Y.Q.Lin, W.H.Abdulla, “Perceptual Evaluation of Audio Watermarking using Objective Quality Measures”, in proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 1745-1748.

AUTHORS

N. V. Lalitha is presently pursuing PhD at GIT, GITAM University. She obtained her M.Tech from Jawaharlal Nehru Technological University, Kakinada and B.Tech from Jawaharlal Nehru Technological University. Presently, she is working as Assistant professor in the Department of Electronics and Communication Engineering at GMR Institute of Technology, Rajam, Srikakulam District. Her research interests are Audio and Image Processing. She is a Life Member of IETE.

Srinivasa Rao Ch is currently working as Professor in the Department of ECE, JNTUK University College of Engineering, Vizianagaram, AP, India. He obtained his PhD in Digital Image Processing area from University College of Engineering, JNTUK, Kakinada, AP, India. He received his M. Tech degree from the same institute. He published 40 research papers in international journals and conferences. His research interests are DigitalSpeech/Image and Video Processing, Communication Engineering and Evolutionary Algorithms. He is a Member of CSI. Dr Rao is a Fellow of IETE.

P. V. Y. Jayasree is currently working as Associate Professor in the Department of ECE, GIT, GITAM University. She obtained her PhD from University College of Engineering, JNTUK, Kakinada, AP, India. She received M.E. from Andhra University. She published more than 50 research papers in international journals and conferences. Her research interests are Signal Processing, EMI/EMC, RF & Microwaves.