



THE STATISTICAL ANALYSIS OF AUDIO WATERMARKING USING THE DISCRETE WAVELETS TRANSFORM AND SINGULAR VALUE DECOMPOSITION

Mr. Jaykumar. S. Dhage

*Assistant Professor, Department of Computer Science & Engineering
MIT, Aurangabad, Maharashtra, India*

Mr. Rahul.B.Mapari

*Assistant Professor, Department of Computer Science & Engineering
MIT, Aurangabad, Maharashtra, India*

Ms. Dipa D. Dharmadhikari

*Assistant Professor, Department of Computer Science & Engineering
MIT, Aurangabad, Maharashtra, India*

Abstract- For the music industry, the piracy is very serious issue. So effective solution to avoid further financial losses and intellectual property violations is necessary. Audio watermarking technology embeds copyright information into audio files as a proof of their ownership. In this paper watermarking algorithm has been brought by virtue of applying a cascade of two powerful mathematical Transforms; the discrete wavelets transform (DWT) and



the singular value decomposition (SVD). When these two algorithms result is compared with the existing algorithms then we can easily get an idea of its effectiveness for the audio watermarking.

Keywords- *Audio watermarking, An inaudible watermarking, Copyright protection Discrete wavelets transform, Singular value decomposition.*

I. INTRODUCTION

Simple data protection techniques like encryption for protecting the music industry's intellectual properties are inefficient. Digital watermarking technology is now attracting attention as a new method of protecting against unauthorized copying of digital multimedia files that includes image, audio and video components. Digital watermarking aims at embedding a watermark in the media file without introducing perceptual degradation. The embedded watermarks may be generated to refer to originators, receivers, unique serial numbers, or time stamps. These watermarks assure the integrity and origin source authentication of the multimedia file without degrading its overall quality. Inaudibility and watermark robustness to removal or degradation, are two necessary requirements for any effective audio data-hiding algorithm..

However, inaudibility must be given special attention since, if the quality of the original audio cannot be preserved, neither users nor owners will accept the audio watermarking technology. Research in audio watermarking is not as mature, compared to research in image and video watermarking (Arnold, M. (2003)) techniques employ human perceptual properties and frequency masking characteristics of the human auditory system for watermarking. These techniques usually use DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), or DWT (Discrete Wavelets Transform) to transform the audio signal to locate appropriate embedding location. But computational complexity and synchronization overhead may be unacceptably high. However, these techniques do not make use of a psychoacoustic model. In



International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists



<http://www.ijccr.com>
VOLUME 2 ISSUE 1 JANUARY 2012

In addition, architectures of standard audio compression engines need to be modified, to incorporate the watermarking modules. In this paper, we propose an audio watermarking algorithm that satisfies the requirements of effective audio watermarking; inaudibility and watermark robustness to removal or degradation.

II Basics Transforms

Discrete Wavelets Transform

The discrete wavelets transform (DWT) is a novel discipline capable of giving a time-frequency representation of any given signal. Starting from the original audio signal S , DWT produces two sets of coefficients as shown in Figure 1. The approximated coefficients A (low frequencies) are produced by passing the signal S through a low pass filter y . The details coefficients D (high frequencies) are produced by passing the signal S through a low pass filter g .

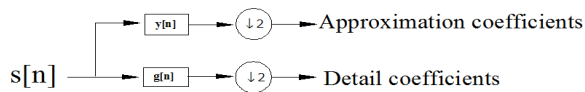


Figure 1. One-level DWT decomposition.

Depending on the application and the length of the signal, the low frequencies part might be further decomposed into two parts of high and low frequencies. Figure 2 shows a 3-level DWT decomposition of signal S . The original signal S can be reconstructed using the inverse DWT process.

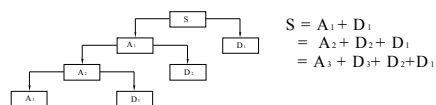




Figure 2. Three-level DWT decomposition

Due to its excellent spatio-frequency localization properties, the DWT is very suitable to identify areas in an audio signal where a watermark can be embedded effectively. Many DWT-based audio watermarking techniques can be found in literature.

The Singular Value Decomposition Transform

The traditional frequency transforms; FFT, DCT and DWT transforms attempt to decompose an image in terms of a standard basis set. This need not necessarily be the optimal representation for a given image. On the other hand, the singular value decomposition (SVD) is a numerical technique for diagonalizing matrices in which the transformed domain consists of basis states that is optimal in some sense (Andrews & Patterson, 1976). The SVD of an $N \times N$ matrix A is defined by the operation $A = U S V^T$ as shown in Figure 3.

$$\begin{pmatrix} V_{1,1} & \dots & V_{1,n} \\ V_{2,1} & \dots & V_{2,n} \\ \dots & \dots & \dots \\ V_{n,1} & \dots & V_{n,n} \end{pmatrix} \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \sigma_{nn} \end{pmatrix} \begin{pmatrix} U_{1,1} & \dots & U_{1,n} \\ U_{2,1} & \dots & U_{2,n} \\ \dots & \dots & \dots \\ U_{n,1} & \dots & U_{n,n} \end{pmatrix}$$

Figure 3: The SVD operation $SVD(A) = U S V^T$.

The diagonal entries of S are called the singular values of A and are assumed to be arranged in decreasing order $\sigma_i > \sigma_{i+1}$. The columns of the U matrix are called the left singular vectors while the columns of the V matrix are called the right singular vectors of A .



III Method

The proposed algorithm employs a cascade of two transforms; the discrete wavelet transform and the singular value decomposition transform. *Watermark Embedding Procedure*

The procedure is illustrated in the block diagram shown in Figure 4, and described in details in the steps which follow.

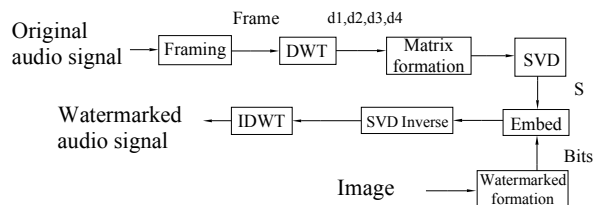


Figure 4- Watermark Embedding Procedure

Step # 1: Convert the binary-image watermark into a one-dimensional vector W of length $m \times n$.

$$W_i = \{[0, 1], 1 \leq i \leq (m \times n)\} \quad (1)$$

Step # 2: Sample the original audio signal at a sampling rate of 44100 samples per second. Then, partition the sampled file into frames each having 50,000 samples. The summation of N frames makes up the overall sampled audio signal as illustrated in the following equation:

$$A = \sum_{i=1}^Y A_i \quad (2)$$

Step # 3: Perform a four-level DWT transformation on each frame A_i . This operation produces five multi-resolution sub-bands: D_1 , D_2 , D_3 , D_4 and A_4 . The D_s represent the details sub-bands and A_4 represents the approximation sub-band.



<http://www.ijccr.com>
VOLUME 2 ISSUE 1 JANUARY 2012

Step # 4: Arrange the four details sub-bands D1, D2, D3, and D4 in a matrix form as shown in Figure 5 below. The matrix, named DC thereafter, has the size $4 \times (L/2)$, where L is the length of each frame.

D1							
D2				D2			
D3		D3		D3		D3	
D4	D4	D4	D4	D4	D4	D4	D4

Figure 5- Matrix formulation of the details D sub-bands.

Step #5: Decompose the DC matrix using the SVD operator. This operation produces the three orthogonal matrices S, U and V^T as follows:

$$DC = U \times S \times V^T \quad (3)$$

Where the S is the following 4×4 diagonal matrix:

$$\begin{pmatrix} S_{11} & 0 & 0 & 0 \\ 0 & S_{22} & 0 & 0 \\ 0 & 0 & S_{33} & 0 \\ 0 & 0 & 0 & S_{44} \end{pmatrix}$$



International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists



<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

S =

(4)

The diagonal s_{ii} entries are the non-zero singular values of the DC matrix. The s_{11} value is used for embedding as will be shown later, and therefore it needs to be stored for later use in the watermark extraction procedure.

Step # 6: Embed the binary-image watermark bits into the DWT-SVD-transformed audio signal according to the following formula:

$$S_{11W} = S_{11} \times (1 + \alpha \times w(n)) \quad (5)$$

Where $w(n)$ is the watermark bit: 0 or 1, α is the watermark intensity, s_{11} is the top left value in the S-matrix, and s_{11w} is the watermarked s_{11} . If α was set to 0.2, then s_{11w} will equal $(1.2 \times s_{11})$ when $w(n)$ is 1, and to (s_{11}) when $w(n)$ is 0.

Step # 7: Produce the final watermarked audio signal as follows:

Apply the inverse SVD operation using the U and V^T matrices, which were unchanged, and the S matrix, which has been modified according to Equation (5). The CDW matrix is the watermarked DC matrix of Equation (3).

$$CD_w = U \times SW \times V^T \quad (6)$$

Apply the inverse DWT operation on the CDW matrix to obtain each watermarked audio frame A_{iw} . The overall watermarked audio signal AW is obtained by summing all watermarked frames.



$$Aw = \sum_i Aiw \quad (7)$$

B. Watermark Extraction Procedure

The watermark extraction procedure requires the watermarked audio signal and the singular values of each frame of the original audio signal. The procedure is illustrated in the block diagram shown in Figure 6, and described in details in the steps which follow.

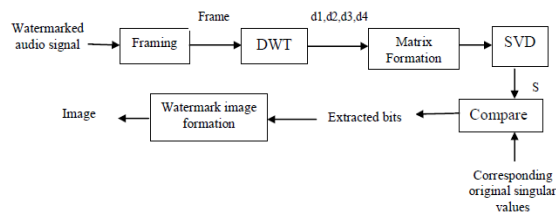


Figure 6- The watermark extraction procedure.

IV Results and Discussion

Pop music and speech audio clips are used to evaluate performance of the proposed algorithm. The two audio types have different perceptual properties, characteristics and energy distribution, and thus their performances may vary from one type to another. The watermark used in our experiments is the binary image shown in Figure 7. The image has a size of 6 x 4 pixels, with each pixel is either a 0 (black) or 255 (white). Below is the description of the metrics used to evaluate performance of the algorithm.



<http://www.ijccr.com>
VOLUME 2 ISSUE 1 JANUARY 2012



Figure 7- The binary watermark image.

C. Performance Evaluation Metrics

Performance of audio watermarking algorithms is usually evaluated with respect to fidelity, imperceptibility (inaudibility), and robustness. In what follows, we give a brief description of each metric. Imperceptibility is related to the perceptual quality of the embedded watermark data within the original audio signal. It ensures that the quality of the signal is not perceivably distorted and the watermark is imperceptible to a listener.

Signal to Noise Ratio (SNR) is a statistical difference metric which is used to measure the similitude between the undistorted original audio signal and the distorted watermarked audio signal. The SNR computation is done according to Equation (8), where A corresponds to the original pop signal, and A' corresponds to the watermarked pop signal.

$$\text{SNR}(\text{db}) = 10 \log_{10} \frac{\sum_n A_n^2}{\sum_n (A_n - A'_n)^2} \quad (8)$$

Therefore, in our experiments the PAQM scores will be mapped to the grading scale of MOS which is shown in Table 1.

TABLE 1: MOS GRADING SCALE.

MOS Grade	Description



**International Journal of Computing
and
Corporate Research**

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists



<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very Annoying

A listening (hearing) test is to be performed with five listeners to estimate the subjective MOS grade of the watermarked signals. Each listener will be presented with the pairs of original signal and the watermarked signal and will be asked to report whether any difference could be detected between the two signals. The average grade for of each pair from all listeners corresponds to the final grade for the pair.

D. Robustness

Watermarked audio signals may undergo common signal processing operations such as linear filtering, lossy compression, among many others (Voloshynovskiy et al., 2001; Arnold, 2003). Although these operations may not affect the perceived quality of the host signal, they may corrupt the watermark image embedded within the signal.

$$BER = \frac{100}{l} \sum_{n=0}^{l-1} \begin{cases} 1 & \text{if } W_n = W'_n \\ 0 & \text{if } W_n \neq W'_n \end{cases} \quad (9)$$

Where l is the watermark length, W_n corresponds to the n th bit of the embedded watermark and W'_n corresponds to the n th bit of the extracted watermark.

Pop Music Watermarking Results

The proposed algorithm first evaluated using a .WAV pop music file of length 600,000 samples (13 seconds). The .WAV music signal is a stereo-type having left and right channels, and therefore the watermark was embedded into both channels.



International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists



<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

The watermarks extracted after application of the various attacks on the watermarked signal are shown in given in Table 2. The BER values obtained from two published algorithms are given below.

Table.2 : BER Values (%) Obtained From Two Published Algorithms for The Audio Signal

Stir Mark Attack	Ozer(STFT-SVD)	Cox(DCT)
Sinus	0	0.77
Echo	27	23.43
Noise	0	1.56

Inaudibility

The watermarked pop signal is as shown in Figures 8. We conducted listening test to find out the SNR and MOS grade values for the same signal.

Figure 8: Original and watermarked pop signal ($\alpha = 0.2$).

Table 3. SNR and MOS values for different watermarking intensities.

Intensity	SNR	MOS
0.20	28.55	5.00



International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists



<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

0.30	25.03	5.00
------	-------	------

E. Speech Signal Watermarking Results

This proposed algorithm is evaluated using a .AU speech signal of length 1,200,000 samples. Unlike the .WAV stereo-type signals which have two channels, speech signals are of the mono-type (have only one channel). Performance results will be calculated as below.

Inaudibility

The watermarked speech signal is shown in Figures 9. Table lists the corresponding SNR values, and MOS grades obtained by conducting the listening test. The waveform in the figure, and the SNR & MOS values verify imperceptibility of the algorithm.

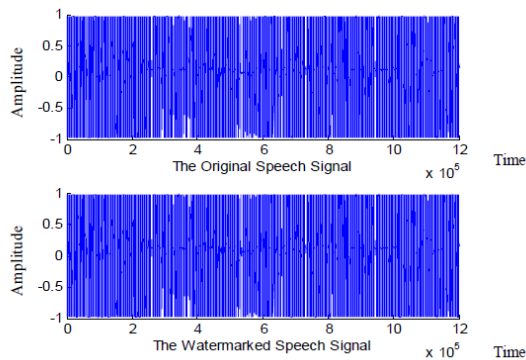


Figure 8: Original and watermarked speech signal ($\alpha = 0.2$).

F. Robustness

We will get data about the watermarks extracted after application of the various attacks on the watermarked speech signal. We should not be apply the Extra Stereo attack since the speech signals are of the mono-type and not the stereo-type.



International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists








<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

Again, lastly performance of the proposed algorithm will be compared with the previous algorithms.

Table 4. BER values (%) for the speech audio signal.

Stir Mark Attack	Extracted Watermark	Proposed(DWT -SVD)	Ozer(2005) STFT-SVD	Cox(1997) DCT
Amplify		0	0	49.6
AddNoise		0	0	0
Invert		0	0	48.75
Zerocross		0	6	0
Echo		10	0	48.96

The below table shows the statistics related to different attacks on an image given in the referred paper. The extracted watermark is feuded to different attacks. These attacks are shown below.

Table 5 : BER Values (%) for The Audio Signal

StirMark	Extrac	DWT-	Ozer(2005)	Cox(1997)(D
AddBrumm		0	0	1.25
AddSinus		0	0	0.77
AddNoise		0	0	1.56
LSB Zero		0	0	0




International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists



<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

Amplify		0	0.75	52.32
Smooth		0	0	0
Stat		45	0	0
Invert		0	0	52.42
ZeroCross		0	0	0
Extra Stereo		0	0	0
Cut		12	0	100
Zero		0	0	100
Exchange		0	0	0
MP3		0	N/A	N/A
Equalization		0	N/A	N/A
Echo		27	0	23.43

In the Cut Samples attack, the audio signal loses samples resulting in a reduction in its length. Therefore, it is possible to lose information in terms of the number of frames in extraction (which have to be an integer number). The watermarks extracted after application of the various attacks on the watermarked speech signal are shown. This is due to the nature of speech signals which have larger sample values. Below table shows the details of signal to noise ratio as well as the mean opinion score with different intensity of watermark.

Table 6 SNR And MOS Values For Different Watermarking Intensities

Intensity	SNR	MOS
0.20	28.55	5.00



**International Journal of Computing
and
Corporate Research**

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists

<http://www.ijccr.com>

VOLUME 2 ISSUE 1 JANUARY 2012

0.30	25.03	5.00
------	-------	------

Pop music and speech audio clips were used to evaluate performance of the proposed algorithm. The two audio types have different perceptual properties, characteristics and energy distribution, and thus their performances may vary from one type to another.

V Conclusion

Many digital audio watermarking have been developed, and claims about their performance are made public. However, many of such algorithms are not evaluated with respect to imperceptibility (SNR, MOS) and robustness (BER), as we have done in this paper. We have studied evaluation metrics of several algorithms . In this paper, we proposed an imperceptible (inaudible) and robust audio watermarking technique based on cascading two powerful mathematical transforms; the Discrete Wavelet Transform (DWT) and the Singular Value Decomposition (SVD). The watermark bits are not be embedded directly on the wavelet coefficients, but rather on the elements of singular values of the DWT sub-bands of the audio frames. By virtue of cascading the two transforms, inaudibility and different levels of robustness are achieved, as we have to demonstrated using pop music and speech audio signals. The simulation results which are obtained verify the effectiveness of audio watermarking as a reliable solution to the copyright protection problem which is facing the music industry.

REFERENCES

[1] Mohammad, A., Al-Haj, A., & Shaltaf, S. (2008). An improved SVD-based watermarking scheme for protecting rightful ownership. *Signal Processing Journal*; 88(9): 2158-2180.



[2] Arnold, M. (2003). Attacks on digital audio watermarks and countermeasures. Proceedings of the IEEE International Conference on WEB Delivering of Music, 1-8.

[3] Bassia, P., & Pitas, I. (2001). Robust audio watermarking in the time domain. IEEE Transaction on Multimedia ; 3(2): 232-241.

[4] Andrews, H., & Patterson, C. (1976). Singular Value Decomposition (SVD) Image Coding. IEEE Transactions on Communications; 42(4): 425-432.

[5] Arnold, M. (2000). Audio watermarking: Features, applications and algorithms. Proceeding of the IEEE International Conference on Multimedia and Expo, 1013–1016.

[6] Arnold, M., Wolthusen, S., & Schmucker, M. (2003). Techniques and applications of digital watermarking and content protection. Artech House. Zwicker E, Fastl H. Psychoacoustics: Facts and models. Springer-Verlag.

[7] Bao, P., & Ma, X. (2004). MP3-Resistant Music Steganography based on Dynamic range transform. Proceedings of the International Conference on Intelligent Signal Processing and Communication Systems, 266–271.

[8] Arnold, M. (2000). Audio watermarking: Features, applications and algorithms. Proceeding of the IEEE International Conference on Multimedia and Expo, 1013–1016.