An Efficient Approach of an Audio Watermarking: Cascading of the Discrete Wavelets Transform and Singular Value Decomposition

J. S. Dhage A. K Gulve

Govt. College of Engineering, Aurangabad, Mahrastra - 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, discrete wavelets transform, singular value decomposition, an inaudible watermarking, copyright protection.

1. 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. The relative importance given to each of these requirements in a watermarking implementation depends on the desired application of the system. Therefore, in practice there exists a fundamental trade-off between the two requirements in such a way to ensure effective audio watermarking.

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)) This is due to the fact that, the human auditory system is much more sensitive than the human visual system, and that inaudibility is much more difficult to achieve than invisibility for images.. In LSB (Bassia, P., & Pitas, I. 2001), watermark information bits are embedded into the least significant bits of the audio signal and in echo hiding as delayed attenuated versions of the original signal. Frequency-domain audio watermarking 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. Time-domain audio watermarking is relatively easy to implement and requires less computing resources, as compared with the frequency-domain watermarking, however, it is weaker against signal processing attacks such as audio compression and filtering, among many others. Similarly, pure frequency-domain approaches may show unsatisfactory robustness in signals with very few transform domain components. But computational complexity and synchronization overhead may be unacceptably high. Compressed audio watermarking operates on compressed, rather than uncompressed, audio for watermarking. In such techniques, the watermark is embedded directly into the already compressed audio bit stream to prevent the watermark from being removed by compression. Therefore, time-consuming decoding, watermarking embedding, and re-encoding are not necessary to embed the watermark, resulting in low computational cost. However, these techniques do not make use of a psychoacoustic model. 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. The requirements were met by the proposed algorithm by exploiting the attractive properties of two powerful mathematical transforms; the Discrete Wavelet Transform (DWT) (Mallat, 1989; Strang & Nguyen, 1996), and the Singular Value Decomposition (SVD) (Andrews & Patterson, 1976). In the proposed algorithm, watermark bits are not embedded directly into DWT's coefficients, but rather on the elements of singular values of the DWT sub-bands of the audio frames. The proposed audio DWT-SVD watermarking algorithm is described below.

2. TRANSFORMS BASICS

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.

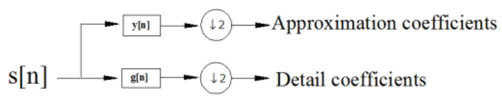
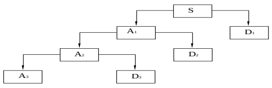


Figure 1: One-level DWT decomposition.

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.



$$S = A_1 + D_1 = A_2 + D_2 + D_1 = A_3 + D_3 + D_2 + D_1$$

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 diagonali zing matrices in which the transformed domain consists of basis states that is optimal in some sense (Andrews & Patterson, 1976). The SVD of an N x N matrix A is defined by the operation A = U S V T as shown in Figure 3.

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

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. By virtue of the fact that slight variations in the elements of matrix S does not affect visual perception of the quality of the cover object, SVD-based audio watermarking algorithms add the watermark information to the singular values of the diagonal matrix S in such a way to meet the imperceptibility(inaudibility) and robustness requirements of effective digital audio watermarking algorithms.

3. THE PROPOSED ALGORITHM

The proposed algorithm employs a cascade of two transforms; the discrete wavelet transform and the singular value decomposition transform. The algorithm is described in this section by outlining the major steps in its two procedures; the watermark embedding procedure and the watermark extraction procedure.

Watermark Embedding Procedure

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

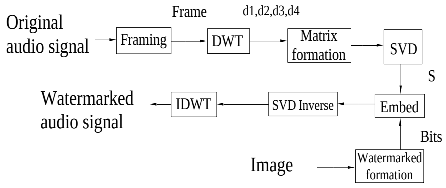


Figure 4: Watermark Embedding Procedure

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

$$Wi = \{[0,1], 1 \leq t \geq (m \times n)\} \qquad (1)$$

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:

$$A = \sum_{i=1}^{Y} Ai \qquad (2)$$

Step # 3: Perform a four-level DWT transformation on each frame Ai. This operation produces five multi-resolution sub-bands: D1, D2, D3, D4 and A4. The Ds represent the details sub-bands and A4 represents the approximation sub-band.

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 x (L/2), where L is the length of each frame.

<table>
<tr><td colspan="8">D1</td></tr>
<tr><td colspan="4">D2</td><td colspan="4">D2</td></tr>
<tr><td colspan="2">D3</td><td colspan="2">D3</td><td colspan="2">D3</td><td colspan="2">D3</td></tr>
<tr><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td><td>D4</td></tr>
</table>

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 VT as follows:

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

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

$$S = \begin{bmatrix} S_{11} & 0 & 0 & 0 \\ 0 & S_{22} & 0 & 0 \\ 0 & 0 & S_{33} & 0 \\ 0 & 0 & 0 & S_{44} \end{bmatrix} \qquad (4)$$

The diagonal sii entries are the non-zero singular values of the DC matrix. The s11 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:

$$S11W = S11 \times (1+ \alpha \times w(n)) \quad (5)$$

where w(n) is the watermark bit: 0 or 1, α is the watermark intensity, s11 is the top left value in the S-matrix, and s11w is the watermarked s11. If α was set to 0.2, then s11w will equal (1.2 s11) when w(n) is 1, and to (s11) when w(n) is 0.

Step # 7: Produce the final watermarked audio signal as follows:
• Apply the inverse SVD operation using the U and VT 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 Aiw. The overall watermarked audio signal AW is obtained by summing all watermarked frames.

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

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.

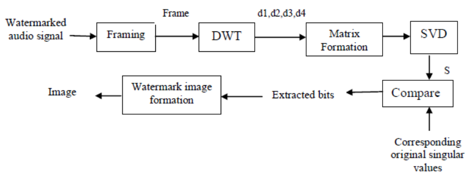


Figure 6: The watermark extraction procedure.

Step # 1: Perform steps 1 through 5 of the embedding procedure until the S matrix is obtained for all frames of the watermarked audio signal.
Step # 2: If the original audio signal has been used instead of the original singular values of each frame, then perform steps1 through 5 of the embedding procedure until the S matrix is obtained for all frames of the original signal.
Step # 3: Determine the watermark bit w(n) by recalling that for embedding a bit into a frame, s11 was modified according to Equation (5). Therefore, the watermark bit w(n) will have a value of 1 if s11w / s11 is 1.2, and 0 if s11w / s11 is 1.0.
Step # 4: Assemble the extracted bits from the individual frames and construct the original binary image watermark.

4. PERFORMANCE EVALUATION

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 with pixel is either a 0 (black) or 255 (white). Below is the description of the metrics used to evaluate performance of the algorithm.

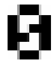


Figure 7: The binary watermark image.

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. To measure imperceptibility, we will use signal-to-noise ratio (SNR) as an objective measure, and a listening test as a subjective measure.
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.

$$SNR(db) = 10\log10 \quad \sum_n An' / \sum_n (An - An')^2 \quad (8)$$

Although SNR is a simple way to measure the noise introduced by the embedded watermark and can give a general idea of imperceptibility, it does not take into account the specific characteristics of the human auditory system. Therefore, we employ the Perceptual Audio Quality Measure (PAQM). PAQM derives an estimate of the signals on the cochlea and compares the representation of the reference signal with that of the signal under test. Its has been shown in that the correlation between PAQM and the mean opinion score (MOS) is 0.98. 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
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.

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. To evaluate robustness of the proposed algorithm, we have to implement a set of attacks that commonly affect audio signals like compression of watermarked (.WAV) audio signals at 128 kbps. The extracted watermarks, robustness against attacks are measured using the BER (Bit Error Rate). BER is defined as the ratio of incorrect extracted bits to the total amount of embedded bits, as expressed in Equation (9) below. The use of BER has become common recently (Grody & Brutun, 2000), as it allows for a more detailed scale of values.

$$BER = 100/t \sum_{n=0}^{i-1} \text{ is 1 if } W'n = Wn \text{ \& is 0 if } W'n \neq Wn \qquad (9)$$

Where l is the watermark length, Wn corresponds to the nth bit of the embedded watermark and W'n corresponds to the nth 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.

Inaudibility

We conducted listening test to find out the SNR and MOS grade values for the same signal.

Table 2: SNR and MOS values for different watermarking intensities.

Intensity	SNR	MOS
0.20	28.55	5.00
0.30	25.03	5.00

Speech Signal Watermarking Results

This proposed algorithm is evaluated using a an .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.

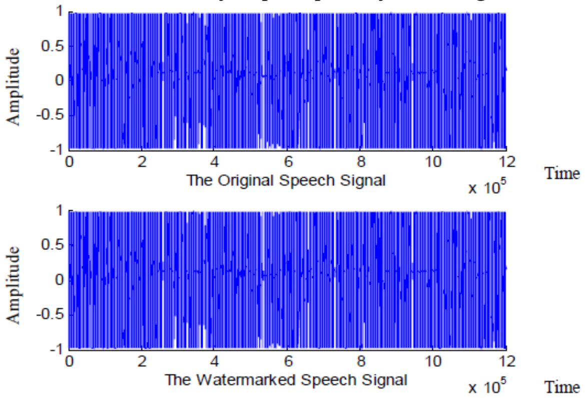


Figure 9: Original and watermarked speech signal (α = 0. 2).

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. We are expecting the better results with speech watermarking compared with the pop music watermarking, especially for the Add/remove category of attacks. The expectation is due to the nature of speech signals which have larger sample values.

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

Table 3: 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

5. 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. Also we analyze SNR and MOS values of traditional techniques.

International Federation of Photographic Industry (IFPI 2009; Wu et al., 2005). IFPI states that the watermark should not degrade perception of audio, the algorithm should offer more than 20 dB SNR, the watermark should be able to resist most common audio processing operations and attacks, and the watermark should prevent unauthorized removal unless the quality of audio becomes very poor. Finally we have to fulfill performance of the proposed algorithm with the desired IFPI required performance.

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 demonstrate 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.

6. 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.