

RESEARCH

Open Access

An imperceptible and robust audio watermarking algorithm

Ali Al-Haj

Abstract

In this paper, we propose a semi-blind, imperceptible, and robust digital audio watermarking algorithm. The proposed algorithm is based on cascading two well-known transforms: the discrete wavelet transform and the singular value decomposition. The two transforms provide different, but complementary, levels of robustness against watermarking attacks. The uniqueness of the proposed algorithm is twofold: the distributed formation of the wavelet coefficient matrix and the selection of the off-diagonal positions of the singular value matrix for embedding watermark bits. Imperceptibility, robustness, and high data payload of the proposed algorithm are demonstrated using different musical clips.

Keywords: Audio watermarking; Copyright protection; Discrete wavelet transform; Singular value decomposition; Imperceptibility; Robustness; Data payload

1 Introduction

The recent advancements of digital audio technology have increased the ease with which audio files are stored, transmitted, and reproduced. However, along with such conveniences come new risks such as copyright violation. Conventional encryption algorithms permit only authorized users to access encrypted digital data; however, once decrypted, there is no way to prohibit illegal copying and distribution of the data [1]. A promising solution to the copyright violation problem is to apply audio watermarking in which audio files are marked with secret, robust, and imperceptible watermarks to achieve copyright protection [2-5]. Indeed, a digital watermark is a good deterrent to illicit copying and dissemination of copyrighted audio since it can provide evidence of copyright infringements after the copyright violation has occurred.

Audio watermarking techniques which are used for copyright protection of digital audio signals must satisfy two main requirements: imperceptibility and robustness [6]. Imperceptibility refers to the condition that the embedded watermark should not produce audible distortion to the sound quality of the original audio. That is, the watermarked version of the audio signal must be

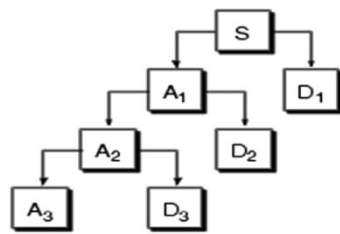
indistinguishable from the original audio signal. On the other hand, robustness ensures the resistance of the watermark against removal or degradation. The watermark should survive malicious attacks such as random cropping and noise adding. Some watermarking applications may demand additional requirements such as high data payload and low computational time of the watermarking algorithm [3]. In practice, there exists a fundamental trade-off between the different watermarking requirements.

Audio watermarking can be carried out in the time domain or the transform domain of the audio signal. Time-domain techniques based on least significant bit substitution and echo hiding are found extensively in literature [7-12]. In general, time-domain audio watermarking techniques are relatively easy to implement and require few computing resources. However, they are less robust than transform-domain techniques which employ the human perceptual properties and frequency masking characteristics of the human auditory system [13]. Popular transforms that have been widely used in digital watermarking include the discrete Fourier transform (DFT), the discrete cosine transform (DCT), the discrete wavelet transform (DWT), and the singular value decomposition (SVD) [14-20].

It has been reported recently that imperceptible and robust audio watermarking can be achieved by applying

Correspondence: ali@psut.edu.jo

Department of Computer Engineering, King Abdullah II Faculty of Engineering, Princess Sumaya University for Technology, Al-Jubeiha, PO Box 1438, Amman 11941, Jordan



$$\begin{aligned}
 S &= A_1 + D_1 \\
 &= A_2 + D_2 + D_1 \\
 &= A_3 + D_3 + D_2 + D_1
 \end{aligned}$$

Figure 1 Three-level DWT decomposition of signal S .

a cascade of two different transforms on the original audio signal. Being different, the cascaded transforms may provide different, but complementary, levels of robustness against the same attack. Many audio watermarking techniques based on hybrid transforms have been proposed in literature. These techniques include but are not limited to DWT-DCT [21], DWT-SVD [22], and SVD-STFT [23].

Several hybrid algorithms based on the SVD transform have been recently proposed in literature. In the algorithm proposed by [23], the audio signal is first converted into a matrix form using the short-time Fourier transform (STFT), the SVD transform is then applied on the matrix, and finally embedding is carried out by adaptively modifying the SVD coefficients with watermark bits. In the hybrid algorithm proposed by [24], the audio signal is partitioned into blocks, and the watermark bits are embedded using dither modulation quantization of the singular values of the blocks. In [23], an audio watermarking algorithm is proposed in which watermark embedding and extraction procedures are based on the quantization of the norms of the singular values of audio blocks. The same authors proposed in [25] a hybrid algorithm in which watermark bits are embedded by applying quantization index modulation (QIM) on the singular values of wavelet-domain blocks. All of the abovementioned SVD-based hybrid algorithms employ some sort of quantization to embed watermark bits. Although quantization is simple, an acceptable level of robustness against noise and filtering attack may not always be achieved.

In this paper, we propose a semi-blind hybrid audio watermarking algorithm based on the DWT and SVD transforms. In the proposed algorithm, the audio signal is sampled, partitioned into short audio segments called frames, and a four-level DWT decomposition is applied on each frame. A matrix is then formed by arranging the wavelet coefficients of all detail sub-bands in a unique

distributed pattern which scatters the watermark bits throughout the transformed frame to provide a high degree of robustness. The SVD operator is then applied on the matrix, and the watermark bits are embedded onto the off-diagonal zero elements of the S matrix produced by the SVD transform. Unlike the other SVD-based algorithms, the proposed algorithm leaves the non-zero singular values of the S matrix unchanged to ensure high watermarking imperceptibility.

The rest of the paper is organized as follows. In the next section, the DWT and SVD transforms are described, and their unique utilization in the proposed algorithm is outlined. The proposed audio DWT-SVD watermarking algorithm is described in detail in Section 3, and evaluated with respect to imperceptibility, robustness, and data payload in Section 4. Concluding remarks are given in Section 5.

2 Related work and contribution

The proposed algorithm is based on cascading the two transforms: DWT and SVD. The uniqueness of the proposed algorithm is twofold: the distributed formation of the DWT coefficient matrix and the selection of the off-diagonal positions of SVD's singular value matrix for embedding watermark bits. Description of the two transforms and their exact utilization in the proposed algorithm is given in this section.

2.1 DWT-based audio watermarking

DWT is a frequency transform capable of giving a time-frequency representation of any given signal [26]. Starting from an audio signal S , DWT produces two sets of coefficients: the approximation coefficients A_1 produced by passing S through a low-pass filter and the detail coefficients D_1 produced by passing S through a high-pass filter. Depending on the application and the length of S , A_1 can be further decomposed into more levels. Figure 1

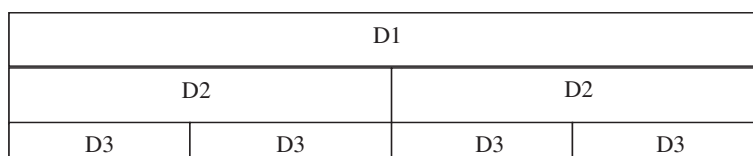


Figure 2 Matrix formation of the of detailed coefficient sub-bands.

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

Figure 3 The SVD operation $SVD(A) = U \Sigma V^T$.

illustrates a three-level DWT decomposition of the audio signal S .

Many DWT-based audio watermarking algorithms can be found in literature. Many variations among the different algorithms exist; however, the main variation is in the sub-band chosen for embedding the watermark bits. In [27-29], the approximation sub-band is used for embedding the watermark bits, while in most algorithms, only one detail sub-band is used to embed the watermark bits [30-36]. Claims of good imperceptibility and robustness have been reported using the two embedding approaches.

In this paper, watermark bits are not embedded in one sub-band only, rather the bits are distributed among all multi-resolution detail sub-bands. For a three-level DWT decomposition, this is done by forming a matrix of the detail sub-bands (D_1 , D_2 , and D_3) as shown in Figure 2. This matrix formation allows for better scattering of the watermark bits throughout the sub-bands, leading to a higher degree of robustness. The resultant DWT matrix is processed by the SVD transform to embed the watermark bits, as will be explained in the next subsection.

2.2 SVD-based audio watermarking

The SVD of matrix A is defined by the operation $A = U \Sigma V^T$, as shown in Figure 3. The non-zero diagonal entries of Σ 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 .

The SVD transform has been used in several audio watermarking algorithms [22-25,37-39]. The algorithms varied in the way the singular values were used in the watermarking process. For example, in [37], the single largest singular value, σ_{11} , was quantized and used to embed the watermark, whereas in [38], the encrypted watermark signal was added to all singular values of matrix Σ . In [22,24,25], the norms of all singular values were quantized and used in the watermark embedding process.

In our proposed algorithm, matrix A represents the detail sub-bands matrix shown in Figure 2, which is produced after applying DWT on the original audio signal. After applying the SVD operator on the DWT matrix, watermark bits are embedded onto the off-diagonal zero elements of the S matrix, while the diagonal singular values of the matrix remain unchanged. This embedding procedure will eliminate the possibility of any distortion caused to the singular values which may affect imperceptibility and watermarking quality. Related preliminary works have been published by the author and others in [40,41]. The algorithms reported in those papers have low capacity as they embed the watermark bits in the single largest singular value, σ_{11} , and not in the off-diagonal zero elements of the Σ matrix, as it is the case in the proposed algorithm.

3 Proposed DWT-SVD audio watermarking algorithm

In this section, we describe the proposed DWT-SVD algorithm. The algorithm consists of two procedures: watermark embedding and watermark extraction procedures.

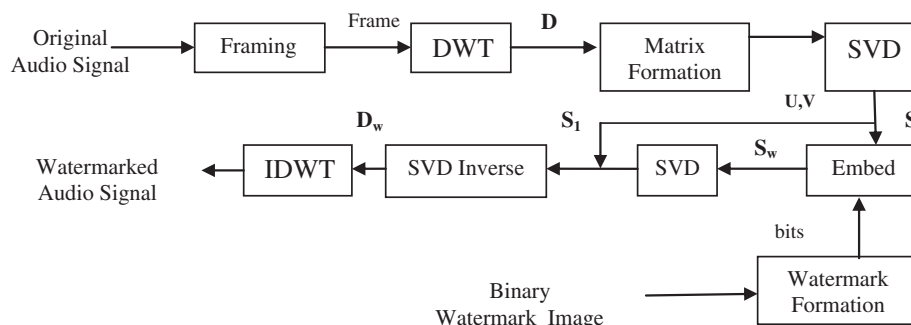


Figure 4 The watermark embedding procedure.

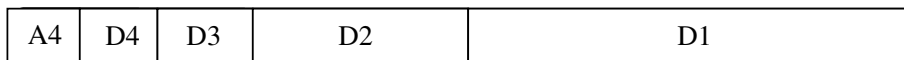


Figure 5 A vector representing the five DWT multi-resolution sub-bands.

3.1 Watermark embedding procedure

The watermark embedding procedure transforms the audio signal using DWT and SVD, embeds the bits of a binary image watermark in appropriate locations in the transformed signal, and finally produces a watermarked audio signal by performing inverse SVD and DWT operations. The procedure is illustrated in the block diagram shown in Figure 4 and described thereafter.

Step 1: Convert the binary image watermark into a one-dimensional vector \mathbf{b} of length $M \times N$. A watermark bit b_i may take one of two values: 0 or 1.

$$b_i = \{[0, 1], 1 \leq i \leq (M \times N)\} \quad (1)$$

Step 2: Sample the original audio signal at a sampling rate of 44,100 samples per second and partition the sampled file into N frames. The optimal frame length will be determined experimentally in such a way to increase data payload.

Step 3: Perform a four-level DWT transformation on each frame. This operation produces five multi-resolution sub-bands: D_1, D_2, D_3, D_4 , and A_4 . The D sub-bands are called ‘detail sub-bands’ and the A_4 sub-band is called ‘approximation sub-band’. The five sub-bands are arranged in the vector shown in Figure 5.

Step 4: Arrange the four detail sub-bands D_1, D_2, D_3 , and D_4 in a matrix D as shown in Figure 6. The matrix formation is done this way to distribute the watermark bits throughout the multi-resolution sub-bands D_1, D_2, D_3 , and D_4 . Forming the matrix with the D s, rather than using A alone, is done to allow for matrix formation and subsequent application of the matrix-based SVD operator. The size of matrix D is $4 \times (L/2)$, where L refers to the length of the frame.

Step 5: Decompose matrix D using the SVD operator. This operation produces the three orthonormal matrices Σ, U , and V^T as follows:

$$D = U * \Sigma * V^T \quad (2)$$

where the diagonal matrix Σ has the same size of the D matrix. The diagonal σ_{ii} entries correspond to the singular values of the D matrix. However, for embedding purposes, only a 4×4 subset of matrix Σ , assigned the name S hereafter, is used as shown below. This is a trade-off between imperceptibility (inaudibility) and payload (embedding capacity). That is, using the whole Σ matrix for embedding will increase embedding capacity but will lead to severe distortion in imperceptibility (inaudibility) of the watermarked audio signal.

$$S = \begin{bmatrix} S_{11} & 0 & 0 & 0 \\ 0 & S_{22} & 0 & 0 \\ 0 & 0 & S_{33} & 0 \\ 0 & 0 & 0 & S_{44} \end{bmatrix} \quad (3)$$

Step 6: Arrange 12 bits of the original watermark bit vector \mathbf{b} into a scaled 4×4 watermark matrix W . The watermark bits must be located in the non-diagonal positions within the matrix, as shown below.

$$W = \begin{bmatrix} 0 & \text{bit1} & \text{bit2} & \text{bit3} \\ \text{bit4} & 0 & \text{bit5} & \text{bit6} \\ \text{bit7} & \text{bit8} & 0 & \text{bit9} \\ \text{bit10} & \text{bit11} & \text{bit12} & 0 \end{bmatrix} \quad (4)$$

As an example, the watermark 12-bit watermark pattern 1010 0011 0101 must be converted to the following matrix form before the actual embedding is carried out.

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad (5)$$

Step 7: Embed watermark matrix W bits into matrix S according to the following ‘additive-embedding’ formula:

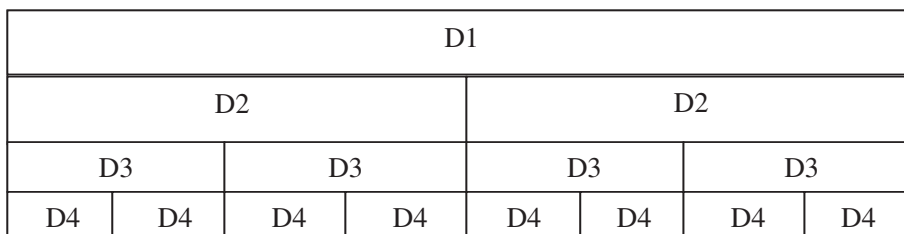


Figure 6 Matrix formation of the detail coefficient sub-bands (D matrix).

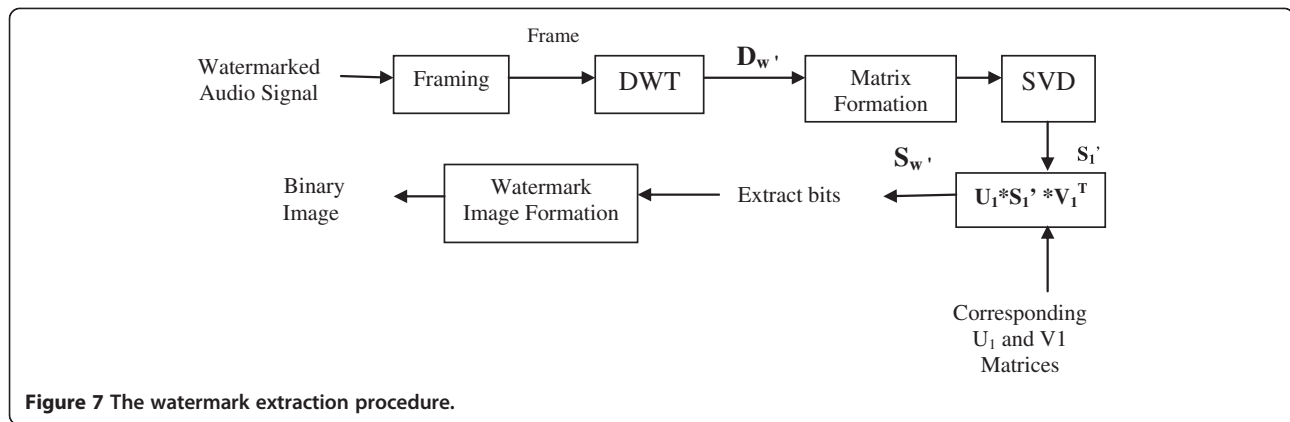


Figure 7 The watermark extraction procedure.

$$S_w = S + \alpha * W \tag{6}$$

where S_w is the watermarked S matrix, and α is the watermark intensity which should be chosen to tune the trade-off between robustness and imperceptibility. With this type of embedding, the singular values of D remain unchanged, and thus, audible distortion caused by modifying the singular values is avoided.

Step 8: Decompose the new watermarked matrix S_w using the SVD operator. This operation produces three new orthonormal matrices as follows:

$$S_w = U_1 * S_1 * V_1^T \tag{7}$$

The matrices U_1 and V_1^T are stored for later use in the extraction process. This makes the proposed watermarking algorithm semi-blind, as the whole original audio frame is not required in the extraction process.

Step 9: Apply the inverse SVD operation using the U and V^T matrices, which were unchanged, and the S_1

matrix, which has been modified according to Equation (6). The D_w matrix given below is the watermarked D matrix given in Equation (2).

$$D_w = U * \Sigma' * V^T \tag{8}$$

where matrix Σ' is the original Σ matrix with the S sub-matrix replaced by the S_1 sub-matrix.

Step 10: Apply the inverse DWT operation on the D_w matrix to obtain the watermarked audio frame.

Step 11: Repeat all previous steps on each frame. The overall watermarked audio signal is obtained by concatenating the watermarked frames obtained in the previous steps.

3.2 Watermark extraction procedure

Given the watermarked audio signal and the corresponding U_1 and V_1 matrices that were computed in Equation (7) and stored for each frame, the embedded watermark can be extracted according to the procedure outlined in Figure 7 and described in detail in the following steps:

Step 1: Obtain the matrix S_1' from each frame of the watermarked audio signal following the general steps presented in Figure 7.

Step 2: Multiply matrix S_1' by U_1 and V_1 which were computed in the watermark embedding procedure and stored for use in the extraction process. This results in the following matrix.

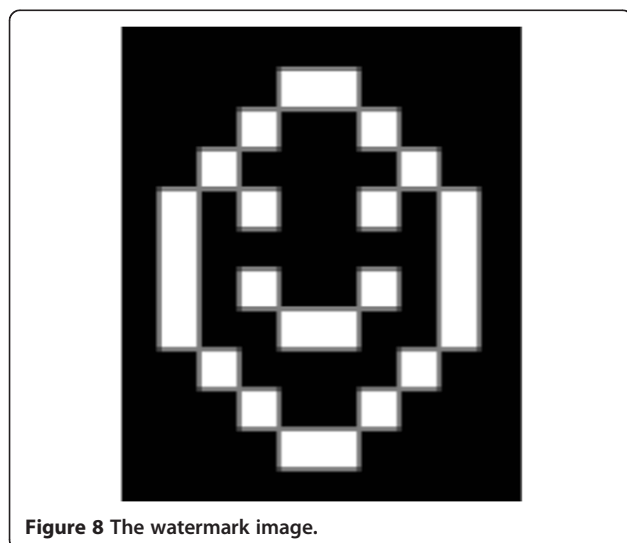


Figure 8 The watermark image.

Table 1 SNR values for different audio signals

Audio type	SNR _{dB}
Pop audio	38.75
Instrumental audio	39.02
Speech audio	37.50
Average	38.17

Table 2 Subjective and objective grades for audio quality measurement

Audio quality	Subjective difference grade (SDG)	Objective difference grade (ODG)
Imperceptible	5	0
Perceptible, but not annoying	4	-1.0
Slightly annoying	3	-2.0
Annoying	2	-3.0
Very annoying	1	-4.0

$$S_w' = U_1 * S_1' * V_1^T \quad (9)$$

Step 3: Extract the 12 watermark bits from each frame by examining the non-diagonal values of matrix S_w' . It has been experimentally noticed that there are two groups of non-diagonal values that are extremely distinct. The values at the positions where a 0 bit has been embedded tend to be much smaller than those values at the positions where a 1 bit has been embedded. Thus, to determine the watermark bit $W(n)$, the average of non-diagonal values is first computed, name it avg, then for each non-diagonal value $S_w'_{ij}$, $W(n)$ is extracted according to the following formula:

$$W(n) = \begin{cases} 0 & S_w'_{ij} \leq \text{avg} \\ 1 & \text{otherwise} \end{cases} \quad (10)$$

Step 4: Construct the original watermark image by assembling the bits extracted from all frames.

4 Experimental results

Different types of audio signals have different perceptual properties, and therefore, watermarking performance may vary from type to another. Accordingly, we evaluated the performance of the proposed algorithm using three mono audio signals representing pop music, instrumental music, and speech. Each signal has a duration of 11 s and was sampled at 44.1 kHz and quantized to 16 bits per sample. The watermark used for experimentation is the 12×10 binary image shown in Figure 8. The watermark is embedded repeatedly throughout the sampled signal, such that one

Table 3 SDG and ODG values for different audio signals

Audio type	SDG	ODG
Pop audio	4.67	-0.67
Instrumental audio	4.72	-0.71
Speech audio	4.81	-0.91
Average	4.73	-0.76

Table 4 Imperceptibility results for different algorithms

Reference	Algorithm	SNR (average)	SDG (average)	ODG (average)
Wang and Zhao [21]	DWT-DCT based	43.11	-	-
Xiang [27]	DWT based	23.98	-	-1.98
Fallahpour and Megias [30]	DWT based	30.65	-	-0.7
Bhat et al. [24]	SVD-DM based	-	4.64	-0.73
Bhat et al. [25]	SVD-DWT based	24.37	4.46	-
Proposed algorithm	SVD-DWT based	38.17	4.73	-0.76

single watermark image is embedded in a sequence of ten frames.

Four-level DWT decomposition is applied on each frame using the Daubechies wavelet (*db1*). Using other wavelet types has a little effect on the performance, as it was observed experimentally. Values ranging from 1 to 5 were used for the watermark intensity α . However, the results reported in this paper were obtained when the intensity value was set to 3. In what follows, we present performance results of the proposed algorithm with respect

Table 5 BER values for common signal processing operations

Attack type	Pop audio	Instrumental audio	Speech audio	Average BER
Gaussian noise (15 _{dB})	0	0	0	0
Gaussian noise (20 _{dB})	0	0	0	0
Re-sampling 22.05	0.0021	0.000	0.0363	0.0128
Re-sampling 11.025	0.0061	0.0011	0.0448	0.0173
Re-sampling 6 kHz	0.0901	0.0330	0.0543	0.0591
Re-quantization 24	0	0	0	0
Re-quantization 8	0	0	0	0
MP3 compression 128 kbps	0	0	0	0
MP3 compression 96 kbps	0.0301	0.0541	0.0721	0.0430
MP3 compression 64 kbps	0.0521	0.0841	0.0820	0.0727
MP3 compression 32 kbps	0.0810	0.1410	0.2901	0.1707
Echo (delay 100 ms, decay 50%)	1.1264	1.5932	1.878	1.5325
Echo (delay 100 ms, decay 40%)	1.0536	1.5641	1.7330	1.4500
Low-pass filtering 8 kHz	0.0972	0.1540	0.3168	0.1893
High-pass filtering 50 Hz	0.2701	0.2810	0.5231	0.3580
Band-pass filtering (100 to 4,000 Hz)	0.1080	0.132	0.2130	0.1510

Table 6 Comparison between BER values of different transform-based algorithms

	Algorithm					
	DWT based [29]	DWT-DCT based [21]	DWT based [50]	SVD-QIM based [22]	SVD-DWT based [25]	SVD-DWT based Proposed algorithm
Gaussian noise (20 _{dB})	7.525	0.0115	0	0	0	0
Re-quantization 8	0	0	0	0	0	0
Re-sampling 22.05	0	0	0	0	2	0.0128
MP3 compression 64 kbps	4.34	0	0.08	0.5615	0	0.0727
MP3 compression 32 kbps	17.22	0.03525	0.67	2.2094	1	0.1707
Echo (delay 100 ms, decay 40%)	-	-	5.83	3.955 (98, 41)	2 (98, 41)	1.450
Low-pass filtering 8 kHz	-	-	0.97	0.3540 (11,025 Hz)	0 (11,025 Hz)	0.1893

to three metrics: imperceptibility, robustness, and data payload [42,43].

4.1 Imperceptibility results

Imperceptibility ensures that the quality of the signal is not perceptibly distorted and the watermark is imperceptible to listeners. To measure imperceptibility, different authors use different metrics; however, the most commonly used metrics are signal-to-noise ratio (SNR) and listening tests.

4.1.1 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 (11), where A corresponds to the original signal, and A' corresponds to the watermarked signal.

$$\text{SNR}(dB) = 10\log_{10} \frac{\sum_n A_n^2}{\sum_n (A_n - A'_n)^2} \quad (11)$$

We obtained the SNR_{dB} values given in Table 1. As shown in the table, the values are much higher than the 20_{dB} minimum requirement set by the International Federation of Phonographic Industry [13]. Although SNR is a simple metric 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.

4.1.2 Listening tests

For better evaluation of imperceptibility, subjective and objective listening tests are used. Subjective difference grade (SDG) listening tests are implemented by human listeners, and objective difference grade (ODG) listening tests are implemented by software packages incorporating the human auditory system. The two listening tests use the 5-grade scale shown in Table 2.

We employed a blind subjective listening test to estimate the audio quality of the watermarked signals. The listening test was performed repeatedly with five adults in a listening room equipped with audio testing and recording devices. A computer system running a special software was also used for computer-controlled presentation of the watermarked signals to the listeners and for recording their responses. Each person was presented with ten pairs of signals (original and watermarked) and then asked to give performance scores using the 5-grade impairment scale given in Table 1. The five persons listened to each pair of signals ten times and gave an average SDG value for each pair. The average grade for each pair submitted by all persons is considered the final grade for that particular pair of signals. The SDG averages obtained for the subjective listening tests are 4.67, 4.72, and 4.81 for the pop, instrumental, and speech signals, respectively. These values clearly indicate that imperceptibility has been achieved by the proposed audio watermarking algorithm.





The ODG scores were also computed using the Perceptual Evaluation of Audio Quality (PEAQ) standard. The standard is specified in ITU-R BS.1387 [44] and implemented by the software tool EAQUAL [45]. The ODG values we obtained are -0.67, -0.71, and -0.91 for the pop, instrumental, and speech signals, respectively. These results confirm with those obtained by subjective listening tests. The measured SDG and ODG values are given in Table 3.

Comparing imperceptibility results with results achieved by other algorithms is not straightforward, since different authors use different evaluation metrics. Moreover, subjective evaluation is relative and may differ from one listener to another. This may explain why imperceptibility results are hardly compared in literature. Nonetheless, and for the sake of completion, we present in Table 4 some imperceptibility results achieved by recently proposed algorithms. It is important to note that the values in table are average values taken over different audio types.

Table 7 BER values due to Stirmark® attacks

Stirmark attack	Extracted watermark (pop)	Pop audio	Instrumental audio	Speech audio	Average BER
AddBrumm (55 Hz Sinus)		0	0	0	0
AddSinus (3000 Hz sinus)		0	0	0	0
AddNoise (20 dB level)		0	0	0	0
Stat1 (statistical distortion)		0	0	0	0
Stat2 (statistical distortion)		0	0	0	0
Smooth1 (simple smoothing)		0.80	1.40	0.36	0.853
Smooth2 (simple smoothing)		0.65	1.34	0.29	0.760
Amplify (increases amplitude)		0	0	0	0
Invert (phase shift 180°)		0	0	0	0
Exchange (swaps samples)		5.01	5.54	3.68	4.74

Table 7 BER values due to Stirmark® attacks (Continued)

CutSamples (7 samples per 1,000)		2.41	3.11	1.08	2.20
LSBZero (reset LSBs)		0	0	0	0
ZeroCross (reset samples)		0	0	0	0
ZeroRemove (removes 0 samples)		0	0	0	0

4.2 Robustness results

Watermarked audio signals may undergo signal processing operations such as linear filtering, lossy compression, among many other operations [46,47]. Although these operations may not affect the perceived quality of the host signal, they may corrupt the watermark embedded within the signal. Two sets of attacks were performed to test the robustness of our proposed algorithm. The first set includes the following set of common signal processing operations: Gaussian noise

addition, re-quantization, re-sampling, MP3 compression, low-pass filtering, and echo addition. The other set is the Stirmark® audio watermarking benchmark which includes a whole set of add, modify, and filter attacks [48,49].

Robustness is measured using the bit error rate (BER) metric since the watermark used in the simulation is a binary image. BER is defined as the ratio of incorrect extracted bits to the total amount of embedded bits, as expressed in Equation (12).

Table 8 Comparison between BER values due to Stirmark® attacks

	Algorithm					
	DCT based [51]	SVD-STFT based [23]	DWT based [27]	DWT based [30]	SVD-QIM based [22]	SVD-DWT based Proposed algorithm
AddBrumm (55 Hz Sinus)	1.25	0	15.79	0	-	0
AddSinus (3,000 Hz sinus)	0.77	0	0	0	-	0
AddNoise (20 dB level)	0.78	0	5.875	0	0	0
Stat1 (statistical distortion)	0	0	0	9	0.1831	0
Stat2 (statistical distortion)	0	0	-	-	0.7324	0
Smooth1 (simple smoothing)	0	0	0	14	2.0874	0.853
Smooth2 (simple smoothing)	0	0	0	-	1.0986	0.760
Amplify (increases amplitude by 50%)	0	0.375	0	0	-	0
Invert (phase shift 180°)	52.42	0	0	0	0	0
Exchange (swaps samples)	0	0	0	-	0	4.74
CutSamples (7 samples per 1,000)	100	1.5	0	-	-	2.20
LSBZero (sets LSBs to 0)	0	0	-	0	-	0
ZeroCross (reset samples)	0	0	0	-	0	0
ZeroRemove (removes 0 samples)	100	0	0	-	-	0

Table 9 Effect of frame length on data payload

	Frame length (samples)							
	512	1,024	2,048	4,096	8,192	16,384	32,768	65,536
Data payload (bps)	1,032	516	258	129	64	32	16	8

$$BER = \frac{100}{l} \sum_{n=0}^{i-1} \left\{ \begin{array}{l} 1, \quad W'_n = W_n \\ 0, \quad W'_n \neq W_n \end{array} \right\} \quad (12)$$

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

4.2.1 Common signal processing operations

The following common signal processing attacks were applied to test the robustness of the proposed algorithm:

1. *Additive white Gaussian noise*: White Gaussian noise is added to corrupt the watermarked signal to SNR levels of 15_{dB} and 20_{dB}.
2. *Re-quantization*: The 16-bit watermarked audio signal is re-quantized to 8 bits per sample and 24 bits per sample.
3. *Re-sampling*: The watermarked signal, originally sampled at 44.1 kHz, is down-sampled to 22.05, 11.025, and 6 kHz.
4. *MP3 compression*: The watermarked audio signal is compressed at different bit rates: 128, 96, 64, and 32 kbps.
5. *Low, high, and band-pass filtering*: Filtering at different cutoff frequencies is applied to the watermarked signal.
6. *Echo addition*: An echo signal with a delay of 100 ms and different decay rates are added the watermarked signal.

The BER values we obtained after applying the common signal processing operations are listed in Table 5. As shown in the table, the BER values, which have been computed over the whole period of the test signals, are very small in magnitude and thus reflect the robustness of the proposed algorithm against common signal operations. Maximum robustness has been achieved against the Gaussian noise attacks, re-quantization, and MP3 compression at 128 kbps. BER values due to re-sampling increased as the watermarked signal was down-sampled to lower frequencies.

The same observation is also seen for the MP3 compression attack, where higher BER values were obtained as the compression rate of the watermarked signal was increased. The watermarked signal is also robust against filtering operations as shown in the corresponding small BER values. The least robustness is seen against the echo addition operation as indicated by the relatively higher BER values.

Finally, we compared the robustness of the proposed algorithm with the robustness of recently published transform-based algorithms. Its clear from Table 6 that the proposed algorithm performs better when compared with the other algorithms. It is important to note that the values in Table 6 represent average values taken over different audio types.

4.2.2 Stirmark® attacks

To evaluate robustness of the proposed algorithm furthermore, we implemented a set of attacks defined by *Stirmark® benchmark for audio* [48,49]. The attacks are comprehensive as they include add, filter, and modification attacks. The results are recorded in Table 7 alongside with snapshots of extracted watermarks from the watermarked signals. It is noted in Table 7 that BER values due to most of the attacks are zero. It is also noted that the proposed algorithm performs comparably well with regard to the three audio signal types.

The *Stirmark®* attacks have been used by several transform-based algorithms. Table 8 compares the BER results we obtained and the BER results reported in four relevant references. As shown in the table, the results are comparable among the different transform-based references with regard to most of the *Stirmark®* attacks. It is instructive to note here that *Stirmark®* package can be used to simulate composite attacks, where two or more attacks are tested in one run. Such composite attacks may give better comparison between the different algorithms; however, they are rarely reported in literature.

4.3 Data payload results

Data payload is defined as the data embedding capacity of the algorithm and is measured as the number of bits

Table 10 Data payload results for different algorithms

	Algorithm					
	DWT based [29]	SVD-STFT based [23]	DWT based [50]	DWT based [27]	SVD-DWT based [25]	SVD-DWT based Proposed
Payload (bits per second)	172	32	25	28.71	45.9	258

embedded within one second of the audio signal (bps). In the proposed algorithm, the audio signal is segmented into frames, with each frame having a fixed embedding capacity of 12 watermark bits, as shown in matrix W given in (5). Therefore, the payload is computed by multiplying number of frames per second by the bit capacity of the frame. The number of frames per second depends on the frame length and is computed by dividing the 44.1 KHz sampling rate by the frame length. Table 9 shows the data payload as a function of the frame length.

As shown in the table, the payload increases as the frame length decreases. However, short-length frames degrade performance and result in unacceptable imperceptibility and robustness results. A frame length of 2,048 samples has been fixed and used to evaluate imperceptibility and robustness of the proposed algorithm.

The data payload we obtained is higher than payload rates obtained by other recently proposed algorithms. Table 10 lists the payload of different transform-based audio watermarking algorithms.

5 Conclusions

In this paper, we proposed an imperceptible and a robust audio watermarking technique based on cascading two well-known transforms: the discrete wavelet transform and the singular value decomposition. The two transforms were used in a unique way that scatters the watermark bits throughout the transformed frame in order to achieve high degrees of imperceptibility and robustness. High data payloads were also achieved. The simulation results obtained were in total agreement with the requirements set by IFPI for audio watermarking, thus proving the effectiveness of the proposed algorithm.

Future research will focus on enhancing the proposed algorithm to resist de-synchronization attacks such as random cropping, pitch shifting, amplitude variation, time-scale modification, and jittering. Methods proposed in the literature that counter de-synchronization attacks include the all-list-search method, the combination of spread spectrum and spread spectrum code method, the self-synchronization strategy method, and the synchronization code method. Our approach will be based on embedding synchronization codes with the watermark bits so that the hidden data have the self-synchronization capability.

Competing interests

The author has no competing interests.

Received: 17 April 2014 Accepted: 18 September 2014

Published online: 09 October 2014

References

1. B Furht, D Kirovski, *Encryption and Authentications: Techniques and Applications* (Auerbach, USA, 2006)
2. M Arnold, S Wolthusen, M Schmucker, *Techniques and applications of digital watermarking and content protection*. Artech House, in

- Psychoacoustics: Facts and Models*, ed. by E Zwicker, H Fastl (Springer-Verlag, Massachusetts, USA, 2003)
3. A Acevedo, *Digital Watermarking for Audio Data in Techniques and Applications of Digital Watermarking and Content Protection* (Artech House, USA, 2003), pp. 75–114
 4. C Xu, J Wu, Q Sun, K Xin, Applications of watermarking technology in audio signals. *J. Audio Eng. Soc.* **47**(10), 805–812 (1999)
 5. M Swanson, B Zhu, A Tewfic, L Boney, Current state of the art, challenges and future direction for audio watermarking, in *Proceeding of the IEEE International Conference on Multimedia Computing and Systems* (1999), pp. 19–24
 6. M Arnold, Audio watermarking: Features, applications and algorithms, in *Proceeding of the IEEE International Conference on Multimedia and Expo* (2000), pp. 1013–1016
 7. P Bassia, I Pitas, Robust audio watermarking in the time domain. *IEEE Trans. Multimed.* **3**(2), 232–241 (2001)
 8. WN Lie, LC Chang, Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. *IEEE Trans. Multimed.* **8**(1), 46–59 (2006)
 9. S Dumitrescu, W Wu, Z Wang, Detection of LSB steganography via sample pair analysis. *IEEE Trans. Signal Process.* **51**(7), 1995–2007 (2003)
 10. O Chen, W Wu, Highly robust, secure, and perceptual-quality echo hiding scheme. *IEEE Trans. Speech Audio Process.* **16**(3), 629–638 (2008)
 11. BS Ko, R Nishimura, Y Suzuki, Time-spread echo method for digital audio watermarking. *IEEE Trans. Multimed.* **7**(2), 212–221 (2005)
 12. H Kim, Y Choi, A novel echo-hiding scheme with backward and forward kernels. *IEEE Trans. Circ. Syst. Video Tech.* **13**(8), 885–889 (2003)
 13. S Katzenbeisser, F Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking* (Artech House, USA, 2000)
 14. M Fallahpour, D Perez-Megias, High capacity audio watermarking using FFT amplitude interpolation. *IEICE Electron. Express* **6**(14), 1057–1063 (2009)
 15. M Fan, H Wang, Chaos-based discrete fractional sine transform domain audio watermarking scheme. *Comput. Electr. Eng.* **35**(3), 506–516 (2009)
 16. I Yeo, H Kim, Modified patchwork algorithm: a novel audio watermarking scheme. *IEEE Trans. Speech Audio Process.* **11**(4), 381–386 (2003)
 17. M Hsieh, D Tseng, Y Huang, Hiding digital watermarks using multiresolution wavelet transform. *IEEE Trans. Ind. Electron.* **48**(5), 875–882 (2001)
 18. C Chang, W Shen, H Wang, Using counter-propagation neural network for robust digital audio watermarking in DWT domain. *Proc. IEEE Int. Conf. Syst. Man. Cybern.* **2**, 1214–1219 (2006)
 19. R Liu, T Tan, An SVD-based watermarking scheme for protecting rightful ownership. *IEEE Trans. Multimed.* **4**(1), 121–128 (2002)
 20. A Mohammad, A Al-Haj, S Shaltaf, An improved SVD-based watermarking scheme for protecting rightful ownership. *Signal Process. J.* **88**(9), 2158–2180 (2008)
 21. X Wang, H Zhao, A novel synchronization invariant audio watermarking scheme based on DWT and DCT. *IEEE Trans. Signal Process.* **54**(12), 4835–4840 (2006)
 22. KV Bhat, I Sengupta, A Das, A new audio watermarking scheme based on singular value decomposition and quantization. *Circ. Syst. Signal Process.* **30**, 915–927 (2011)
 23. H Ozer, B Sankur, N Memon, An SVD-based audio watermarking technique, in *ACM Workshop on Multimedia and Security* (2005), pp. 51–56
 24. KV Bhat, I Sengupta, A Das, An audio watermarking scheme using singular value decomposition and dither-modulation quantization. *Multimed. Tool. Appl.* **52**, 369–383 (2011)
 25. K Bhat, I Sengupta, A Das, An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. *Digit. Signal Process.* **20**, 1547–1558 (2010)
 26. G Strang, T Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Wellesley, MA, 1996)
 27. S Xiang, Audio watermarking robust against D/A and A/D conversions. *EURASIP J Adv. Signal Process.* **3**, 1–14 (2011)
 28. H Peng, J Wang, Z Zhang, Audio watermarking scheme robust against desynchronization attacks based on kernel clustering. *Multimed. Tool. Appl.* **3**, 1–14 (2011)
 29. S Wu, J Huang, D Huang, Y Shi, Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Trans. Broadcast.* **51**(1), 69–76 (2005)
 30. M Fallahpour, D Megias, High capacity audio watermarking using the high frequency band of the wavelet domain. *Multimed. Tool. Appl.* **52**, 485–498 (2011)

31. M Swanson, B Zhu, A Tewfic, L Boney, Robust audio watermarking using perceptual masking. *Signal Process.* **66**(3), 337–355 (1998)
32. X Li, M Zhang, L Sun, Adaptive audio watermarking algorithm based on SNR in wavelet domain, in *International Conference on Natural Language Processing and Knowledge Engineering* (2003), pp. 287–292
33. Y Wu, S Shimamoto, A study on DWT-based digital audio watermarking for mobile ad hoc networks, in *International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, 2006. 5–7 June
34. E Erelebi, L Bataki, Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. *Digit. Signal Process.* **19**(2), 265–277 (2009)
35. L Wei, X Xue, An audio watermarking technique that is robust against random cropping. *Comput. Music. J.* **27**(4), 58–68 (2003)
36. X Li, H Yu, Transparent and robust audio data hiding in sub-band domain, in *Proceedings of the International Conference on Information Technology: Coding and Computing* (2000), pp. 74–79
37. C Chang, P Tsai, C Lin, SVD-based digital image watermarking scheme. *Pattern Recogn. Lett.* **26**(10), 1577–1586 (2005)
38. F Abd El-Samie, An efficient singular value decomposition algorithm for digital audio watermarking. *Int. J. Speech Tech.* **21**, 27–45 (2009)
39. A Basso, F Bergadano, D Cavagnino, V Pomponiu, A Vernone, A novel block-based watermarking scheme using the SVD transform. *Algorithms* **2**(1), 46–75 (2009)
40. A Al-Haj, A Mohammad, Digital audio watermarking based on the discrete wavelets transform and singular value decomposition. *Eur. J. Sci. Res.* **39**(1), 6–21 (2010)
41. A Al-Haj, C Twal, A Mohammad, Hybrid DWT-SVD audio watermarking, in *Proceedings of the International Conference on Digital Information Management* (2010), pp. 525–529
42. M Sehirli, F Gurgun, S Ikizoglu, Performance evaluation of digital audio watermarking techniques designed in time, frequency and cepstrum domains, in *International Conference on Advances in Information Systems* (2004), pp. 430–440
43. J Grody, L Brutun, Performance evaluation of digital audio watermarking algorithms, in *The 43rd IEEE Midwest Symposium on Circuits and Systems* (2000), pp. 456–459
44. T Thielde, WC Treurniet, R Bitto, C Schmidmer, T Sporer, JG Beerends, C Colomes, M Keyhl, G Stoll, K Brandenburg, B Feiten, PEAQ – the ITU standard for objective measurement of perceived audio quality. *J. Audio Eng.* **48**(1/2/3), 3–29 (2000)
45. A Lerch, Zplane development, EAQUAL- Evaluate Audio QUALity, version:0.1.3alpha, 2002. <http://www.mp3-tech.org/programmer/misc.html>
46. S Voloshynovskiy, S Pereira, T Pun, Attacks on digital watermarks: classification, estimation-based attacks, and benchmarks. *Comm. Mag.* **39**(8), 118–126 (2001)
47. M Arnold, Attacks on digital audio watermarks and countermeasures, in *Proceedings of the IEEE International Conference on WEB Delivering of Music* (2003), pp. 1–8
48. M Steinebach, F Petitcolas, F Raynal, J Dittmann, C Fontaine, S Seibel, N Fates, LC Ferri, Stirmark benchmark: audio watermarking attacks, in *Proceedings of the International Conference on Information Technology: Coding and Computing* (2001), pp. 49–54
49. A Lang, Stirmark benchmark for audio (SMBA): evaluation of watermarking schemes for audio [Online], 2006. Available: http://omen.cs.uni-magdeburg.de/alang/smba.php#smba_LA
50. N Kalantari, M Akhaee, S Ahadi, S Feizi, H Amindavar, Robust multiplicative patchwork method for audio watermarking. *IEEE Trans. Audio Speech Lang. Process.* **17**(6), 1133–1141 (2009)
51. I Cox, J Kilian, T Leighton, T Shamoan, Secure spread spectrum watermarking for multimedia. *IEEE Trans. Image Process.* **6**(12), 1673–1687 (1997)

doi:10.1186/s13636-014-0037-2

Cite this article as: Al-Haj: An imperceptible and robust audio watermarking algorithm. *EURASIP Journal on Audio, Speech, and Music Processing* 2014 **2014**:37.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
