Blind Audio Watermarking in Transform Domain Based on Singular Value Decomposition and Exponential-Log Operations

Pranab Kumar DHAR[1], Tetsuya SHIMAMURA[2]

[1]Dept. of Computer Science and Engineering, Chittagong University of Engineering and Technology (CUET), Chittagong-4349, Bangladesh
[2]Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan

pranabdhar81@gmail.com, shima@sie.ics.saitama-u.ac.jp

Abstract. *Digital watermarking has drawn extensive attention for copyright protection of multimedia data. This paper introduces a blind audio watermarking scheme in discrete cosine transform (DCT) domain based on singular value decomposition (SVD), exponential operation (EO), and logarithm operation (LO). In our proposed scheme, initially the original audio is segmented into non-overlapping frames and DCT is applied to each frame. Low frequency DCT coefficients are divided into sub-bands and power of each sub band is calculated. EO is performed on the sub-band with highest power of the DCT coefficients of each frame. SVD is applied to the exponential coefficients of each sub bands with highest power represented in matrix form. Watermark information is embedded into the largest singular value by using a quantization function. Simulation results indicate that the proposed watermarking scheme is highly robust against different attacks. In addition, it has high data payload and shows low error probability rates. Moreover, it provides good performance in terms of imperceptibility, robustness, and data payload compared with some recent state-of-the-art watermarking methods.*

Keywords

Copyright protection, Discrete Cosine Transform, Singular Value Decomposition (SVD), Exponential Operation (EO), Logarithm Operation (LO)

1. Introduction

The recent development in computational world and the wide availability of internet have facilitated the transmission and distribution of multimedia data. As a result, the protection of intellectual property rights of digital data has been the key problem. Digital watermarking has drawn an extensive attention and has been a focus in data security. It is a process of embedding watermark into original data objects such as audio, video, and image. It is widely used for several purposes including copyright protection, broadcast monitoring, fingerprinting, data authentication, and medical safety.

In general, the digital watermarking can be broadly classified into robust and fragile (or semi-fragile) watermarking. For robust watermarking schemes, two important issues need to be addressed: one is to show a trustworthy evidence to protect rightful ownership and the other is to provide a good trade-off among perceptual transparency, robustness, and data payload. A comprehensive survey on audio watermarking can be found in [1], [2]. Most audio watermarking methods utilize either a time domain [3], [4] or a transform domain such as fast Fourier transform (FFT) [5], discrete cosine transform (DCT) [6], discrete wavelet transform (DWT) [7], [10], and lifting wavelet transform (LWT). Bassia et al. [3] proposed a watermarking scheme in which watermark bits are embedded by modifying the audio samples directly. Lie and Chang [4] introduced a method in which group amplitudes are modified to achieve high robustness. However, both methods have low data payload. Megías et al. [5] suggested a watermarking method that embeds watermark in FFT domain, but it has low data payload. Zeng and Qiu [6] introduced a watermarking method based on DCT which embeds binary image as watermark, but the subjective evaluation of watermarked audio signal has not been assessed. In [7], authors presented a watermarking algorithm based on energy-proportion scheme which is robust against attacks; however, the SNR results of this algorithm are not satisfactory. Chen et al. [8] proposed an adaptive method based on wavelet based entropy, but robustness to resampling and low-pass filtering attacks are quite low. In [9], authors introduced a robust watermarking scheme in DWT domain which embeds watermark in the low frequency DWT coefficients using optimization-based quantization technique but the subjective evaluation of watermarked audio signals has not been done in this scheme. Wang et al. [10] proposed a method using DWT in which watermark bits are embedded into low-middle frequency wavelet coefficients and linear predictive coding (LPC) is utilized in detection process. But the SNR result is

just above 20 dB and the robustness to re-sampling and low-pass filtering attacks are quite low. Erçelebi and Batakçı [11] presented a watermarking method in LWT domain in which a binary image is embedded as watermark; however, from the reported result, robustness against attacks is quite low. Khaldi et al. [12] proposed a watermarking scheme based on empirical mode decomposition (EMD) in which watermark bits are embedded into the intrinsic mode functions (IMFs). But the subjective evaluation of the proposed scheme was not conducted and the data payload is quite low. Recently, singular value decomposition (SVD) has been utilized as an effective technique in audio watermarking [13–16]. In [13], [14], authors introduced a domain adaptive audio watermarking algorithm using SVD. In addition, the proposed segment-by-segment approach enhanced the delectability compared to the simple approach utilizing the whole original audio signal directly. However, the signal-to-noise ratio (SNR) result is not very good and the detection scheme is non-blind. The most recent SVD based watermarking methods proposed by Bhat et al. [15] and Lei et al. [16] provide high robustness; however the data payload of both methods is quite low. Moreover, some other techniques such as audio histogram technique [17], [18] and time spread (TS) echo method [19], [20] are becoming popular in audio watermarking field. Dhar et al. [21] introduced an audio watermarking method based on DCT, SVD, and Log-polar transform (LPT). But the robustness against MP3 compression is little low. The main limitation of the existing audio watermarking techniques is the difficulty to obtain a favorable trade-off among the imperceptibility, robustness, and data payload. To overcome these limitations, in this paper, we propose an audio watermarking scheme in discrete cosine transform (DCT) domain based on SVD, exponential operation (EO), logarithm operation (LO), and quantization. The main features of the proposed scheme include (*i*) it utilizes DCT, SVD, EO, LO, and quantization jointly, (*ii*) it uses a Gaussian map which contains the chaotic characteristic to enhance the confidentiality of the proposed scheme, (*iii*) watermark is embedded into the largest singular value of the exponential coefficients obtained from the DCT sub band with highest power, (*iv*) watermark extraction process is blind, and (*v*) it achieves a good trade-off among imperceptibility, robustness, and data payload. Simulation results indicate that the proposed watermarking scheme is highly robust against various attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression. Moreover, it shows good performance in terms of imperceptibility, robustness, and data payload compared with some state-of-the-art methods [7–9], [11], [12], [14–19], [21]. The SNR of the proposed scheme range from 27 to 36 dB, in contrast to the state-of-the-art methods whose SNR's range from only 12 to 28 dB. In addition, the bit error rate (BER) of the proposed scheme ranges from 0 to 3.1250 whereas the BER of the state-of-the-art methods range from 0 to 17.50. Furthermore, the data payload of the proposed scheme is

172.39 bps which is relatively higher than the state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 briefly describes the background information including DCT and SVD. Section 3 introduces our proposed watermarking scheme including watermark embedding and detection processes. Section 4 compares the performance of our proposed scheme with the state-of-the-art methods in terms of imperceptibility, robustness, and data payload. In addition, it also provides the error analysis of the proposed scheme. Lastly, the conclusion of the paper is presented in Sec. 5.

2. Background Information

2.1 Discrete Cosine Transform

The DCT has been widely used in signal and image processing, especially for lossy data compression [16]. It represents the signal in the form of a series of coefficients obtained from a sum of cosine functions oscillating at different frequencies and amplitudes. DCT can be written as

$$X(k) = c(k) \sum_{n=1}^{N} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right), \quad k = 1,......,N \quad (1)$$

where $x(n)$ is the audio signal with length of N samples and

$$c(k) = \begin{cases} \dfrac{1}{\sqrt{N}}, & k = 1, \\[2mm] \sqrt{\dfrac{2}{N}}, & 2 \le k \le N. \end{cases} \quad (2)$$

The DCT has the ability to compress the energy of a signal in a few samples, leaving the other samples very small in magnitude. This property can be utilized in audio watermarking to reduce the deterioration of watermarked signal.

2.2 Singular Value Decomposition

SVD is a mathematical tool mainly used to analyze matrices. In SVD transformation, a given matrix A is decomposed into three matrices. Let A be a $p \times p$ matrix with SVD of the form $A = USV^{T}$ where U and V are orthogonal $p \times p$ matrices and S is a $p \times p$ diagonal matrix with nonnegative elements. The diagonal entries of S are called the singular values (SVs) of A where $S = \text{diag}(\sigma_1, \sigma_2,...,\sigma_n)$, the columns of U are called the left singular vectors of A, and the columns of V are called the right singular vectors of A. The SVD has some interesting properties: (*i*) the sizes of the matrices from SVD operation are not fixed, and the matrices need not be square; (*ii*) changing SVs slightly does not affect the quality of the signal much; (*iii*) the SVs are invariant under common signal processing operations; (*iv*) the SVs satisfy intrinsic algebraic properties.

3. Proposed Watermarking Scheme

Let $X = \{x(n), 1 \leq n \leq L\}$ be an original audio signal with L samples, $W = \{w(k, l), 1 \leq k \leq M, 1 \leq l \leq M\}$ be a binary logo image to be embedded into the original audio signal, and $w(k, l) \in \{0, 1\}$ be the pixel value at point (k, l).

3.1 Watermark Preprocessing

Watermark should be preprocessed first in order to improve the robustness and enhance the confidentiality. This paper uses a Gaussian map that contains the chaotic characteristics to encrypt the binary watermark image for enhancing the confidentiality of the proposed method. It can be defined as follows:

$$y(i+1) = \exp\left(-a(y(i))^2\right) + b \qquad (3)$$

where $y(1) \in (0,1)$, a and b are real parameters (map's initial condition). Then a binary sequence $z(i)$ is calculated by using the following equation:

$$z(i) = \begin{cases} 1 & \text{if } y(i) > T \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

where T is a predefined threshold. The binary watermark image W is converted into a one dimensional sequence q, where $q = \{q(i), i = 1, 2, 3,\ldots, M \times M\}$. Finally $q(i)$ is encrypted using $z(i)$ by the following rule:

$$u(i) = z(i) \oplus q(i), \quad 1 \leq i \leq M \times M \qquad (5)$$

where $\oplus$ is the exclusive-or (XOR) operation. After this random chaotic encryption, the original watermark is permuted and cannot be found by random search. In this study, the value of $y(1)$, a, b, and the encrypted watermark sequence $u(i)$ are used as secret key K.

3.2 Watermark Embedding Process

The proposed watermark embedding process is shown in Fig. 1. The embedding process is implemented in the following steps:

1) The original audio signal X is first segmented into non-overlapping frames $F = \{F_1, F_2, F_3,\ldots,F_{M \times M}\}$.

2) Each audio frame F_i is transformed into DCT domain to calculate the DCT coefficients D_i, where i indicates the frame number.

3) The DCT coefficients of each frame F_i are divided into m numbers of sub bands $B = \{B_j, 1 \leq j \leq m\}$ with r numbers of coefficients in each sub band B_j, where j indicates the sub band number.

4) The power $G = \{G_j, 1 \leq j \leq m\}$ of the sub bands $B = \{B_j, 1 \leq j \leq m\}$ of the DCT coefficients of each frame F_i is calculated using the following equation:

$$G_j = \frac{1}{r}\sum_{k=1}^{r}(V_k)^2 \qquad (6)$$
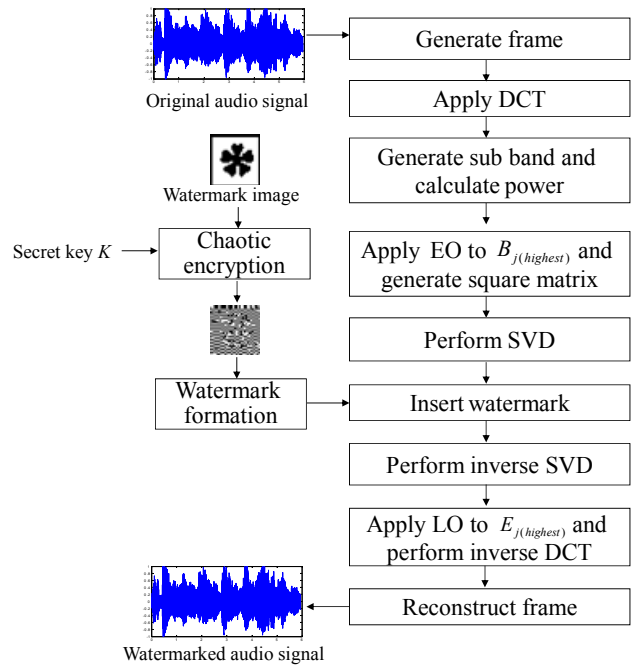

Fig. 1. Watermark embedding process.

where G_i represents the power of each sub band B_i, V_k and k are the amplitude of the DCT coefficients and the index of the DCT coefficients, respectively in each sub band B_j.

5) Find $G_{max} = \max\{G_1, G_2, G_3,\ldots,G_m\}$ from the sub band $\{B_1, B_2, B_3,\ldots,B_m\}$ of each frame F_i, where max operation returns the largest element from $\{G_1, G_2, G_3,\ldots, G_m\}$. Thus, G_{max} represents the largest power in $\{G_1, G_2, G_3,\ldots,G_m\}$.

6) Select G_{max} and find the corresponding sub band from $\{B_1, B_2, B_3,\ldots,B_m\}$ i.e., the sub band with the highest power (denoted by $B_{j(\text{Highest})}$) of the DCT coefficients of each frame F_i. This sub band is selected because it is the significant band to embed watermark data.

7) EO is applied to each element of $B_{j(\text{Highest})}$) to calculate the corresponding exponential coefficients which are denoted by $E_{j(\text{Highest})}$). For each element, the EO is represented by the following equation:

$$E_{j(\text{Highest})} = \exp\left(B_{j(\text{Highest})}\right). \qquad (7)$$

The exponential coefficients $E_{j(\text{Highest})}$ of each frame F_i are rearranged into an $N \times N$ square matrix $\boldsymbol{R}_i$. This is done by dividing the coefficient set into N segments with N coefficients.

8) SVD is performed to decompose each matrix $\boldsymbol{R}_i$ into three matrices: $\boldsymbol{U}_i$, $\boldsymbol{S}_i$, and $\boldsymbol{V}_i$. The SVD operation is represented as follows:

$$\boldsymbol{R}_i = \boldsymbol{U}_i \times \boldsymbol{S}_i \times \boldsymbol{V}_i^{\mathrm{T}}. \qquad (8)$$

9) The binary watermark image is encrypted using the Gaussian map which contains the chaotic characteristics.

10) The largest singular value $S_i(1,1)$ of each matrix $\boldsymbol{S}_i$ is selected. Watermark information should be embed-

ded into the most significant perceptual components of the audio signal in order to guarantee the robustness and imperceptibility of the proposed method. As $S_i(1, 1)$ contains the most power of the signal, therefore, it represents the significant perceptual component of the audio signal. The proposed method embeds watermark bits into the largest singular value $S_i(1, 1)$ of each matrix $\boldsymbol{S}_i$ by using a quantization function. Let $Y_i = \text{Round}(S_i(1, 1)/Q)$, where Q is a predefined quantization coefficient. The embedding equation is given as follows:

$$S_i'(1,1) = \begin{cases} Y_i + C - (Y_i \bmod M), & \text{if } u(i) = 1 \\ Y_i + C - ((Y_i + C) \bmod M), & \text{if } u(i) = 0 \end{cases} \quad (9)$$

where $M = 2C$, C is an integer, mod is the modulo operation, and $S_i'(1, 1)$ is the modified largest value.

11) Re-insert each modified largest singular value $S_i'(1,1)$ into matrix $\boldsymbol{S}_i$ and inverse SVD is applied to obtain the modified matrix $\boldsymbol{R}_i'$ which is given by

$$\boldsymbol{R}_i' = \boldsymbol{U}_i \times \boldsymbol{S}_i' \times \boldsymbol{V}_i^{\mathrm{T}}. \quad (10)$$

Each matrix $\boldsymbol{R}_i'$ is then reshaped to create the modified exponential coefficients $E'_{j(\text{Highest})}$ of each frame F_i by performing the inverse operation of step 5.

12) LO which is the inverse operation of EO is performed on each $E'_{j(\text{Highest})}$ to calculate the coefficients of modified sub band $B'_{j(\text{Highest})}$ of each frame F_i. The LO is represented by the following equation:

$$B'_{j(\text{Highest})} = \log(E'_{j(\text{Highest})}). \quad (11)$$

13) After substituting the modified sub band $B'_{j(\text{Highest})}$ for $B_{j(\text{Highest})}$, an inverse DCT is performed on D_i' to obtain the watermarked audio frame F_i'.

14) Finally, all watermarked frames are concatenated to calculate the watermarked audio signal X'.

3.3 Watermark Detection Process

The proposed watermark detection process does not need the original audio signal to extract the watermark, which is shown in Fig. 2. The detection process is implemented in the following steps:

1) The DCT is performed on each frame F_i^* of the attacked watermarked audio signal.

2) Calculate the power of each sub band B_j^* and EO is performed on $B_{j(\text{Highest})}^*$ of each frame F_i^* to get $E_{j(\text{Highest})}^*$.

3) Rearrange $E_{j(\text{Highest})}^*$ of each frame F_i^* to get $\boldsymbol{R}_i^*$ and SVD is applied to $\boldsymbol{R}_i^*$ of each matrix $\boldsymbol{S}_i^*$ of the attacked watermarked audio frame.

4) Calculate Y_i^* of each $S_i^*(1,1)$.

5) Watermark sequence is extracted as follows:

$$u^*(i) = \begin{cases} 1 & \text{if } (Y_i^* \bmod 2) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$
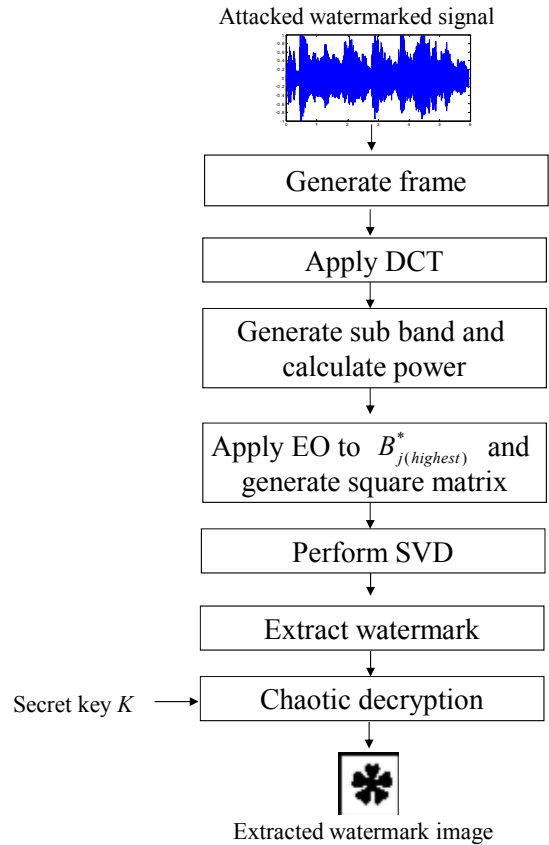
Attacked watermarked signal

Generate frame

Apply DCT

Generate sub band and calculate power

Apply EO to $B_{j(highest)}^*$ and generate square matrix

Perform SVD

Extract watermark

Secret key K → Chaotic decryption

Extracted watermark image

Fig. 2. Watermark detection process.

6) Perform chaotic decryption by using the secret key K_1 to find the hidden binary sequence with the following rule:

$$q^*(i) = z(i) \oplus u^*(i). \quad (13)$$

7) Finally, watermark image is obtained by rearranging the binary sequence $q^*(i)$ into a square matrix $\boldsymbol{W}^*$ of size $M \times M$.

4. Simulation Results and Discussion

In this section, several experiments were conducted to demonstrate the performance of the proposed watermarking scheme. The performance of the proposed scheme is assessed in terms of imperceptibility, robustness, and data payload. In this study, we used 40 audio clips belonging to four different audio groups as original audio signals, which are as follows:

Group 1: 10 clips containing Pop music;
Group 2: 10 clips containing Jazz music;
Group 3: 10 clips containing Classical music;
Group 4: 10 clips containing male and female speeches;

All audio files are mono-channel and contain 262,144 samples (duration 5.94 sec). They are sampled at the rate of 44.1 kHz, quantized with 16 bits. By using a frame size of 256 samples, we have 1024 non-overlapping frames for each audio sample. In each frame of audio signal, we embed one bit binary watermark information. Thus, the length
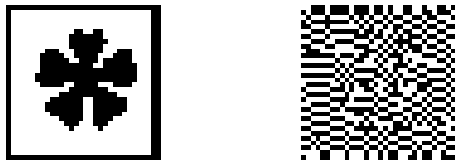
Fig. 3. (a) Binary watermark image. (b) Encrypted watermark image.

SDG	ODG	Description	Quality
5.0	0	Imperceptible	Excellent
4.0	−1	Perceptible, but not annoying	Good
3.0	−2	Slightly annoying	Fair
2.0	−3	Annoying	Poor
1.0	−4	Very annoying	Bad

Tab. 1. Subjective and objective difference grades.

Types of Signal	Subjective evaluation		Objective evaluation	
	MOS	Correct Detection	SNR	ODG
Pop	4.8	46%	31.93	−1.05
Jazz	5.0	48%	36.52	−1.43
Classical	4.9	42%	29.68	−0.49
Speech	4.8	52%	35.76	−1.21
Average	**4.88**	**47%**	**33.47**	**−1.05**

Tab. 2. Subjective and objective evaluation of different watermarked sounds.

of the watermark sequence is 1024. A binary logo image and the corresponding encrypted image by chaotic encryption of size $M \times M = 32 \times 32 = 1024$ are shown in Fig. 3. In this study, the selected value for $y(1)$, *a, b,* and *T* are 0.6, 5.90, –0.39, and 0.25, respectively. For convenience, the selected value for the sub band number *m* and the coefficients of each sub band *r* are both 16. These parameters have been selected in order to achieve a good trade-off among the conflicting requirements of imperceptibility, robustness, and data payload.

4.1 Imperceptibility Test

Audio watermarking intends to embed an unperceivable and secure watermark into the original audio signal. Therefore, the watermarking scheme should be perceptually transparent. The imperceptibility of the watermarked audio is evaluated by using two ways: (1) subjective listening test, (2) objective test.

1) Subjective Listening Test

Subjective listening tests are essential for perceptual quality assessment. In the subjective listening test, ten participants were given the original and watermarked audio signals and were asked to report the dissimilarities between two signals, using a five point subjective difference grade (SDG) ranges from 5.0 to 1.0 (imperceptible to very annoying) as shown in Tab. 1. The average SG (i.e., mean opinion) scores for different watermarked sounds using the proposed scheme are shown in Tab. 2. From the test results, it is seen that the average mean opinion score (MOS) results are within 4.8 to 5.0 for all watermarked sound using the proposed scheme, indicating that watermarked audio signals are perceptually similar to original audio signals.

Subjective test was also done using ABX method. The subjects were ten male and female persons whose hearing ability is normal. Initially, each subject listened the original audio signal (A) and the watermarked audio signal (B) and that subject listened a third audio signal (X) in random order, which can be either A or B. The subjects were asked to verify which of A or B was the same as X. One time of identification was considered as one trial and five times of trials was done by each subject. A detection percentage of 50% indicates that the difference between the original and watermarked sounds was imperceptible. The evaluation results of all subjects were summarized in terms of percentage of correct detection and are shown in Tab. 2. We observed that correct detection scores range from 42 % to 52 %, indicating the high imperceptibility of the watermarked sound.

2) Objective Test

SNR is widely used to measure the objective quality of watermarked audio signal which is formulated as:

$$SNR = 10\log_{10} \frac{\sum_{i=1}^{L} x^2(n)}{\sum_{i=1}^{L} \left[x(n) - x^*(n) \right]^2} \qquad (14)$$

where $x(n)$ and $x^*(n)$ are the original and watermarked audio signals in time domain, respectively. After embedding watermark information, the SNRs of the watermarked audio signals using the proposed scheme are above 20 dB, shown in Tab. 2. According to the International Federation of the Phonographic Industry (IFPI) standard [15], audio watermarking should be imperceptible when SNR is over 20 dB. From Tab. 2, we observed that the proposed scheme satisfied the IFPI standard.

Objective test was also done using the objective difference grade (ODG). The perceptual evaluation of audio quality (PEAQ) measurement technique which is specified in ITU-R BS.1387 (International Telecommunication Union-Radio-communication Sector) standard [21] incorporates a psychoacoustic model. The output of the PEAQ algorithm is the ODG which corresponds to the subjective grade used to measure the differences between the original and watermarked audio signals. The ODG ranges from 0.0 to –4.0 (imperceptible to very annoying) which is shown in Tab. 1. The objective quality of the watermarked audio signals in terms of ODG is shown in Tab. 3. We observed that the ODG values range from –0.49 to –1.43, indicating that original and watermarked audio signals are perceptually similar.

Table 3 shows a comparison of SNR and MOS results between the proposed scheme and the several recent methods which are based on the reported results in the references [7–9], [12], [14–16], [19], [21]. From the comparison results, it is seen that our proposed scheme shows better performance than the recent watermarking methods in terms of SNR and MOS, except the method reported in

Reference	Algorithm	SNR	MOS
[7]	DWT-based energy proportion	17.95	4.15
[8]	Wavelet-based entropy	22.46	4.38
[19]	TS echo hiding	22.70	4.70
[12]	EMD	24.12	--
[15]	DWT-SVD	24.37	4.46
[14]	SVD	27.13	--
[9]	Optimisation-based quantisation	29.50	--
[16]	DCT-SVD	32.53	4.71
[21]	DCT-SVD-LPT	37.20	4.85
Proposed	**DCT-SVD-EO-LO**	**33.47**	**4.88**

Tab. 3. SNR and MOS comparison between the proposed scheme and several recent methods
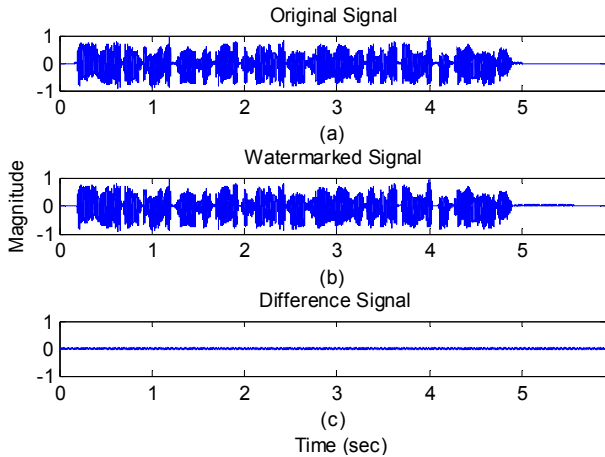

Fig. 4. Imperceptibility of watermarked audio using the proposed scheme: (a) original audio signal 'Speech', (b) watermarked audio signal 'Speech', (c) difference between original and watermarked audio signals.

[21] for SNR. In other word, subjective and objective evaluations prove a high transparency of the proposed scheme with higher MOS and SNR results.

Figure 4 shows the time domain representation of the original audio signal with a watermarked audio signal in which the watermark is imperceptible using the proposed scheme for the signal 'Speech'.

4.2 Robustness Test

Normalized correlation (*NC*) coefficient is used to compare the similarities between the original watermark *W* and the extracted watermark W^*, which is calculated as:

$$NC\,(W,W^*) = \frac{\sum_{k=1}^{M}\sum_{l=1}^{M} w(k,l)\cdot w^*(k,l)}{\sqrt{\sum_{k=1}^{M}\sum_{l=1}^{M} w(k,l)\cdot w(k,l)}\sqrt{\sum_{k=1}^{M}\sum_{l=1}^{M} w^*(k,l)\cdot w^*(k,l)}} \quad (15)$$

where *k* and *l* are the indices of the binary watermark image. The correlation between *W* and W^* is very high if $NC\,(W, W^*)$ is close to 1. On the other hand, the correlation between *W* and W^* is very low if $NC\,(W, W^*)$ is close to zero.

The bit error rate (BER) is used to measure the robustness of a watermarking scheme and is computed as

$$BER(W,W^*) = \frac{\sum_{k=1}^{M}\sum_{l=1}^{M} w(k,l)\oplus w^*(k,l)}{M \times M}. \quad (16)$$

where $\oplus$ is the exclusive or (XOR) operation.

The following signal processing attacks were performed to assess the robustness of the proposed scheme.

1) Additive White Gaussian noise (AWGN): AWGN is added to the watermarked audio signal.

2) Cropping: Segments of 1000 samples (10 × 100) are removed from the watermarked audio signal at ten different positions and subsequently replaced by segments of the watermarked audio signal attacked additive white Gaussian noise.

3) Re-sampling: The watermarked audio signals are down-sampled back to 22.05 kHz and then up-sampled back to 44.1 kHz.

4) Re-quantization: Each sample of watermarked audio signals is re-quantized from 16 bits to 8 bits.

5) Low-pass filtering: Low-pass filter with 10 kHz cut-off frequency is applied to the watermarked audio signals.

6) MP3 compression: MPEG-1 layer 3 compression was applied. The watermarked audio signal is compressed at a bit rate of 32 kbps and then decompressed back to the wave format.

It is very important to select the significant sub band from the DCT coefficients for embedding watermark. The sub band with highest power is the significant sub band for embedding watermark. In this study, 256 DCT coefficients are divided into 16 sub bands (*m* = 16) with 16 DCT coefficients (*r* = 16) in each sub band. The 1st 16 DCT coefficients belong to the 1st sub band, the next 16 DCT coefficients belong to the 2nd sub band, and the next 16 DCT coefficients belong to the 3rd sub band and so on. We observed that most of the cases, the sub band with the highest power was found in the 1st band. This is because the DCT has energy compaction characteristics. For this reason, most of the energies were concentrated in low frequency region. Figure 5 shows the robustness results (NC values) of the proposed method against different attacks for the audio signal "Speech" where the highest power, the 2nd highest power, the 3rd highest power, and the 4th highest power represent embedding watermark bits into the DCT sub band with the highest power, the 2nd highest power, the 3rd highest power, the 4th highest power, respectively of each frame. We observed that embedding watermark bits into DCT sub band with the highest power provides better robustness than embedding watermark bits into DCT sub band with the 2nd highest power, the 3rd highest power, and the 4th highest power. In this situation, DCT sub band with the highest power can work better than the other DCT sub bands with lower power. This is because DCT sub band with the highest power indicates the optimal DCT sub band for embedding watermark.
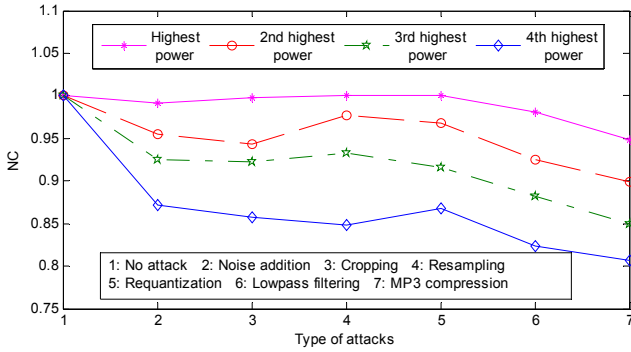
Fig. 5. Robustness result of the proposed scheme where watermark is embedded in different DCT sub bands for the audio signal "Speech".

Type of attacks	NC	BER (%)	Extracted watermark
No attack	1	0	
Noise addition	0.9919	0.9766	
Cropping	0.9976	0.2386	
Re-sampling	1	0	
Re-quantization	1	0	
Low-pass filtering	0.9810	2.3438	
MP3 compression	0.9481	6.4213	

Tab. 4. Extracted watermark image with NC and BER for the audio signal 'Speech'.

Audio Signal	Types of attacks	NC	BER (%)
Pop	No attack	1	0
	Noise addition	0.9951	0.5865
	Cropping	0.9980	0.1962
	Re-sampling	1	0
	Re-quantization	1	0
	Low pass filtering	0.9870	1.5625
	MP3 compression	0.9593	5.9256
Jazz	No attack	1	0
	Noise addition	0.9919	0.9766
	Cropping	0.9968	0.2841
	Re-sampling	1	0
	Re-quantization	1	0
	Low pass filtering	0.9822	2.1484
	MP3 compression	0.9497	6.2158
Classical	No attack	1	0
	Noise addition	0.9943	0.6836
	Cropping	0.9978	0.2145
	Re-sampling	1	0
	Re-quantization	1	0
	Low pass filtering	0.9894	1.2695
	MP3 Compression	0.9543	5.9832

Tab. 5. NC and BER of the extracted watermark for different audio signals.

Table 4 shows the robustness results of the proposed scheme in terms of NC and BER against several attacks for the audio signal 'Speech'. The minimum NC value and the maximum BER value are 0.9481 and 6.4231, respectively. The extracted watermark images are visually similar to the original watermark. This clearly shows a good performance of the proposed scheme against different kinds of attacks. Table 5 shows similar results for the audio signal 'Pop', 'Jazz', and 'Classical', respectively. The NC values are all above 0.94 and the BER values are all below 7 %, demonstrating the high robustness of our proposed scheme against different attacks.

4.3 Security

For a secured watermarking scheme robustness against attack is important. To enhance the security, the proposed method utilizes chaotic encryption. Since the proposed watermark embedding and detection processes depend on the secret key K, it is impossible to maliciously detect the watermark without these keys.

4.4 Data Payload

The data payload is defined as the number of bits that can be embedded into the original audio signal within a unit of time and is measured by bits per second (bps). The data payload P can be represented as follows:

$$P = \frac{B}{T}(\text{bps}) \tag{17}$$

where T is the duration of the original audio signal in seconds and B is the number of watermark bits to be embedded into the original audio signal. Usually, the data payload for any watermarking method should be more than 20 bps [15]. The data payload of the proposed scheme is 172.3906 bps.

Due to the diversity of the watermarking methods, a general comparison between the proposed scheme and the several recent methods sorted by data payload is given in Tab. 6, which is based on the reported result in [7], [8], [11], [12], [15–18], [21]. In addition, re-sampling, re-quantization, and MP3 compression are compared in Tab. 6. The maximum BER for each attack is mentioned in Tab. 6. For example, in our proposed scheme, the maximum BER of attacked watermarked sound for resampling (22.05 kHz), requantization (8 bits/sample) and MP3 compression (32 kbps) attack is 0, 0, and 6.42, respectively. According to the reported result in Wavelet-based entropy [8], the maximum BER of resampling (22.05 kHz) and MP3 compression (128 kbps) attack is 9.1 and 6.7, respectively. According to the reported result in SVD-DCT [16], the maximum BER of resampling (22.05 kHz), requantization (8 bits/sample) and MP3 compression (32 kbps) attack is 0, 0, and 3, respectively. We observed that our proposed method shows better robustness than some recent methods such as DCT-SVD-

Reference	Algorithm	Payload (bps)	Re-sampling BER (%)	Re-quantization BER (%)	MP3 compression BER (%)
Proposed	**DCT-SVD-EO-LO**	**172.39**	**0 (22.05 kHz)**	**0 (8 bits/sample)**	**6.42 (32 kbps)**
[21]	DCT-SVD-LPT	172.39	1.56 (22.05 kHz)	0 (8 bits/sample)	8.67 (32 kbps)
[8]	Wavelet-based entropy	172.28~86.14	9.1 (22.05 kHz)	--	6.7 (128 kbps)
[7]	DWT-based energy proportion	114.82	6.92 (22.05 kHz)	--	5.71 (80 kbps)
[12]	EMD	46.9~50.3	3 (22.05 kHz)	0 (8 bits/sample)	1 (32 kbps)
[15]	DWT-SVD	45.90	2 (22.05 kHz)	0 (8 bits/sample)	1 (32 kbps)
[16]	SVD- DCT	43	0 (22.05 kHz)	0 (8 bits/sample)	3 (32 kbps)
[17]	Histogram	3	0 (--)	0 (8 bits/sample)	15 (128 kbps)
[18]	DWT-based Histogram	2	0 (16 kHz)	0 (8 bits/sample)	17.50 (64 kbps)
[11]	LWT	--	16.50 (36.750 kHz)	22.09 (8 bits/sample)	51.73 (128 kbps)

Tab. 6. A general comparison of audio watermarking algorithms with proposed scheme sorted by data payload.

LPT [21], Wavelet-based entropy [8], Histogram [17], DWT-based Histogram [18], LWT [11] and shows comparable robustness with DWT-based energy proportion [7] against MP3 compression. On the other hand, the proposed method shows less robustness than some other methods such as EMD [12], DWT-SVD [15], and DCT-SVD [16] for MP3 compression. Besides, the proposed method has better robustness than the state-of-the-art methods [7], [8], [11], [12], [15–18] in terms of re-sampling and re-quantization attacks. Also, it has higher data payload than the state-of-the-art methods [7], [8], [11], [12], [15–18].

4.5 Error Analysis

Two types of error may occur while searching the watermark sequence: a false positive error (FPE), or a false negative error (FNE). A false positive error occurs when an unwatermarked audio signal is declared as watermarked audio signal by the detector, on the other hand, a false negative error occurs when a watermarked audio signal is declared as unwatermarked audio signal by the detector. It is difficult to give an exact probabilistic model for false positive and false negative errors. Here, a simplified model based on binomial probability distribution is used in our proposed method to calculate the FPE probability and FNE probability which is same as the model utilized by the method reported in [15].

Let s be the total number of watermark bits and t be the total number of matching bits. The FPE probability P_{f_p} and FNE probability P_{f_n} can be calculated as:

$$P_{f_p} = 2^{-s} \sum_{t=\lceil 0.8s \rceil}^{s} \binom{s}{t}, \qquad (18)$$

$$P_{f_n} = \sum_{t=0}^{\lceil 0.8s \rceil - 1} \left[\binom{s}{t} (P)^t (1-P)^{s-t} \right] \qquad (19)$$

where $\binom{s}{t}$ is the binomial coefficient and P is the bit error rate (BER) probability of the extracted watermark. We observed that (18) is independent of BER. This indicates that FPE is independent of attacks. As we select $s = 1024$, by substituting the value of s (18) gives $P_{f_p} = 2.6209 \times 10^{-88}$. Figure 6 shows the FPE probability for $s \in (0,100)$. It is noted that P_{f_p} approaches 0 when s is larger than 30.

The approximate value of P can be obtained from BER under different attacks. From Tab. 5 and 6, we observed that all BER values are less than 0.07. Thus P is taken as 0.93. By substituting the value of s and P, (19) gives $P_{f_n} = 5.4708 \times 10^{-42}$. Figure 7 shows the FNE probability for $s \in (0,100)$. It is noted that P_{f_n} approaches 0 when s is larger than 30.
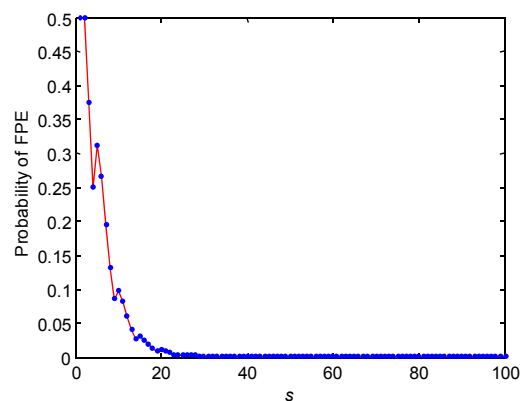

Fig. 6. Probability of FPE for various values of s.
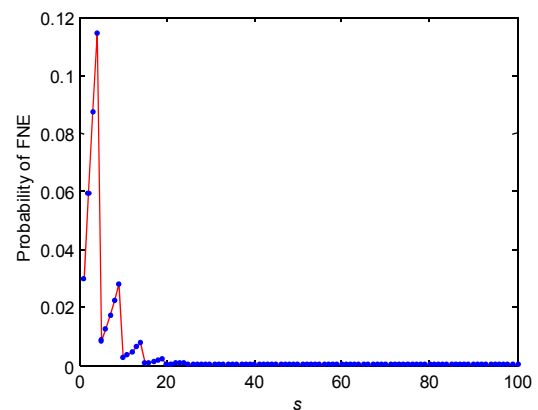

Fig. 7. Probability of FNE for various values of s.

The main difference between the proposed method and the method reported in [15] is that they provide different result for PFE and FNE probability because the value of s, t, and P are different for each method. For example, the P_{f_n} of the proposed method is 5.4708×10^{-42}, whereas the P_{f_n} of the method reported in [15] is 6.47×10^{-34}.

5. Conclusion

In this paper, we have introduced an audio watermarking scheme in DCT domain based on SVD, EO and LO. Simulation results demonstrate that the proposed scheme provides high robustness against different attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression. This is because watermark is embedded into the largest singular value of the exponential coefficients obtained from the DCT sub-band with highest power of each audio frame and slight variations of the largest singular values cannot affect the quality of the sound and also these values may change very little against different attacks. In addition, the proposed scheme achieves very low false positive and false negative error probability rates. Moreover, it provides good performance in terms of imperceptibility, robustness, and data payload compared with some recent state-of-the-art watermarking methods. These results verify the effectiveness of our proposed watermarking scheme for audio copyright protection.

There are several directions for future work on the proposed scheme introduced in this paper. In the future work, we will include synchronization code [5] and error correcting codes [22] to improve the robustness of the proposed scheme. Moreover, psychoacoustic model can be adopted to improve the imperceptibility of the proposed scheme. Furthermore, a comparison between the proposed scheme and several the state-of-the-art methods will be carried out in terms of imperceptibility and robustness with same audio signal and same parameter. In addition, computational complexity of the proposed scheme will be calculated.

References

[1] CVEJIC, N., SEPPANEN, T. (Eds.) *Digital Audio Watermarking Techniques and Technologies. Applications and Benchmarks.* IGI Global, 2007. DOI: 10.4018/978-1-59904-513-9

[2] COX, I., MILLER, M., BLOOM J., FRIDRICH, J., KALKAR, T. Digital Watermarking and Steganography. 2nd ed. *The Morgan Kaufmann Series in Multimedia Information Systems*, Elsevier, 2007. ISBN-13: 978-0123725851

[3] BASSIA, P., PITAS, I., NIKOLAIDIS, N. Robust audio watermarking in the time domain. *IEEE Transaction on Multimedia*, 2001, vol. 3, no. 2, p. 232–241. DOI: 10.1109/6046.923822

[4] LIE, W.-N., CHANG, L.-C. Robust high quality time domain audio watermarking based on low frequency amplitude modification. *IEEE Transaction on Multimedia*, 2006, vol. 8, no. 1, p. 46–59. DOI: 10.1109/TMM.2005.861292

[5] MEGÍAS, D., SERRA-RUIZ, J, FALLAHPOUR, M. Efficient self–synchronized blind audio watermarking system based on time domain and FFT amplitude modification. *Signal Processing*, 2010, vol. 90, no. 12, p. 3078–3092. DOI: 10.1016/j.sigpro.2010.05.012

[6] ZENG, Z., QIU, Z. Audio watermarking in DCT: embedding strategy and algorithm. In *Proceedings of the 11th IEEE International Conference on Signal Processing (ICSP'08)*. 2008, vol. 1, p. 2193–2196. DOI: 10.1109/ICOSP.2008.4697583

[7] CHEN, S. T., HUANG, H. N., CHEN, C. J., WU, G. D. Energy-proportion based scheme for audio watermarking. *IET Signal Processing*, 2010, vol. 4, no. 5, p. 576–587. DOI: 10.1049/iet-spr.2009.0184

[8] CHEN, S. T., HUANG, H. N., CHEN, C. J., TSENG, K. K., TU, S. Y. Adaptive audio watermarking via the optimization point of view on the wavelet-based entropy. *Digital Signal Processing*, 2013, vol. 23, no. 3, p. 971–980. DOI: 10.1016/j.dsp.2012.12.013

[9] CHEN, S. T., WU, G. D., HUANG, H. N. Wavelet-domain audio watermarking scheme using optimisation-based quantization. *IET Signal Processing*, 2010, vol. 4, no. 6, p. 720–727. DOI: 10.1049/iet-spr.2009.0187

[10] WANG, R., XU, D., CHEN, J., DU, C. Digital audio watermarking algorithm based on linear predictive coding in wavelet domain. In *Proceedings of the 7th IEEE International Conference on Signal Processing (ICSP'04)*. 2004, vol. 1, p. 2393–2396. DOI: 10.1109/ICOSP.2004.1442262

[11] ERÇELEBI, E., BATAKÇI, L. Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. *Digital Signal Processing*, 2009, vol. 19, no. 2, p. 265–277. DOI: 10.1016/j.dsp.2008.11.007

[12] KHALDI, K., BOUDRAA, A. O. Audio watermarking via EMD. *IEEE Transaction on Audio, Speech and Language Processing*, 2013, vol. 21, no. 3, p. 675–680. DOI: 10.1109/TASL.2012.2227733

[13] EL-SAMIE, F. E. A. An efficient singular value decomposition algorithm for digital audio watermarking. *International Journal of Speech Technology*, 2009, vol. 12, no. 1, p. 27–45. DOI: 10.1186/s13636-014-0037-2

[14] NUAIMY, W. A., EL-BENDARY, M. A. M., SHAFIK, A., et al. An SVD audio watermarking approach using chaotic encrypted images. *Digital Signal Processing*, 2011, vol. 21, no. 6, p. 764–779. DOI: 10.1016/j.dsp.2011.01.013

[15] BHAT, V. K., SENGUPTA, I., DAS, A. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. *Digital Signal Processing*, 2010, vol. 20, no. 6, p. 1547–1558. DOI: 10.1016/j.dsp.2010.02.006

[16] LEI, B. Y., SOON, I. Y., LI, Z. Blind and robust audio watermarking scheme based on SVD-DCT. *Signal Processing*, 2011, vol. 91, p. 1973–1984. DOI: 10.1016/j.sigpro.2011.03.001

[17] XIANG, S., HUANG, J. Histogram based audio watermarking against time scale modification and cropping attacks. *IEEE Transaction on Multimedia*, 2007, vol. 9, no. 7, p. 1357–1372. DOI: 10.1109/TMM.2007.906580

[18] XIANG, S., KIM, H. J., HUANG, J. Audio watermarking robust against time scale modification and MP3 compression. *Signal Processing*, 2008, vol. 88, no. 10, p. 2372–2387. DOI: 10.1016/j.sigpro.2008.03.019

[19] ERFANI, Y., SIAHPOUSH, S. Robust audio watermarking using improved TS echo hiding. *Digital Signal Processing*, 2009, vol. 19, no. 5, p. 809–814. DOI: 10.1016/j.dsp.2009.04.003

[20] Y. XIANG, D. PENG, I. NATGUNANATHAN, W., ZHOU. Effective pseudonoise sequence and decoding function for

imperceptibility and robustness enhancement in time-spread echo-based audio watermarking. *IEEE Transaction on Multimedia*, 2011, vol. 13, no. 1, p. 2–13. DOI: 10.1109/TMM.2010.2080668

[21] DHAR, P. K., SHIMAMURA, T. Blind SVD-based audio watermarking using entropy and log-polar transformation. *Journal of Information Security and Applications*, 2015, vol. 20, p. 74–83. DOI: 10.1016/j.jisa.2014.10.007

[22] THIEDE, T., TREURNIET, W. C., BITTO, R., et al. PEAQ - The ITU standard for objective measurement of perceived audio quality. *Journal of Audio Engineering Society*, 2000, vol. 48, no. 1/2, p. 3–29.

[23] BAUDRY, S., DELAIGLE, J.-F., SANKUR, MACQ, B. B., MAÎTRE, H. Analyses of error correction strategies for typical communication channels in watermarking. *Signal Processing*, 2001, vol. 81, no. 6, p. 1239–1250. DOI: 10.1016/S0165-1684(01)00041-X

About the Authors...

Pranab Kumar DHAR (corresponding author) received the Bachelor of Science (B.Sc.) degree in Computer Science and Engineering from Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh in 2004. He received Master of Science (M.Sc.) degree in Computer Engineering from University of Ulsan, South Korea in 2010. He received his Ph.D. degree from the Graduate School of Science and Engineering, Saitama University, Japan in 2014. In 2005, he joined the Dept. of Computer Science and Engineering, CUET, Chittagong, Bangladesh as a Lecturer, where he is currently serving as an Associate Professor. Dr. Pranab is an author of one book. His research interest includes multimedia security, digital watermarking, multimedia data compression, sound synthesis, digital signal processing, and digital image processing.

Tetsuya SHIMAMURA received the B.E, M.E., and Ph.D. degrees in Electrical Engineering from Keio University, Yokohama, Japan, in 1986, 1988, and 1991, respectively. In 1991, he joined Saitama University, Japan, where he is currently a Professor. He was a visiting researcher at Longhborough University, U.K. in 1995 and at Queen's University of Belfast, U.K. in 1996, respectively. Prof. Dr. Shimamura is an author and coauthor of 6 books. He serves as an editorial member of several international journals and is a member of the organizing and program committees of various international conferences. His research interests are in digital signal processing and its application to speech, image, and communication systems.