CrossMark

A Blind Audio Watermarking Scheme Employing DCT–HT–SD Technique

Aniruddha Kanhe[1] · Aghila Gnanasekaran[2]

Abstract

In this paper, a blind audio watermarking technique using discrete cosine transform (DCT), Hadamard transform (HT) and Schur decomposition (SD) is presented. The embedding of watermark is carried out by modifying the DCT coefficients through HT and SD. Watermark is embedded in the selected high-energy low-frequency frames of audio signal. The watermark bits are adaptively embedded in the unitary matrix obtained from SD. To make the proposed watermarking technique fully blind, embedding of watermark is performed by effectively utilizing the ratio between the first column elements of unitary matrix. Experimental results confirm that the proposed technique is robust to various signal processing attacks, such as re-sampling, re-quantization, noise addition, filtering, scaling and compression. The comparison of payload and robustness reveals that the proposed watermarking technique outperforms the existing techniques.

Keywords Blind audio watermarking · Discrete cosine transform · Hadamard transform · Schur decomposition · Signal processing attacks

1 Introduction

The increase in usage of internet technology has made the sharing, illegal copying, modifying and tampering of digital audio very easy. Variety of solutions are proposed for the protection of digital audio from such illegal practices. Digital audio

---

✉ Aniruddha Kanhe
aniruddhakanhe@nitpy.ac.in

Aghila Gnanasekaran
aghila@nitpy.ac.in

1 Department of Electronics and Communication Engineering, National Institute of Technology Puducherry, Karaikal, Puducherry, India

2 Department of Computer Science and Engineering, National Institute of Technology Puducherry, Karaikal, Puducherry, India

Birkhäuser

watermarking gives the promising solution for such challenges [26,38,39]. Diverse audio watermarking techniques are proposed in recent years for different applications. These techniques are mainly classified as blind, semi-blind and non-blind audio watermarking techniques. In blind watermarking techniques, the watermark is extracted without the original source. In semi-blind watermarking, the information other than the original source is required to extract the watermark. Whereas, in non-blind watermarking original source is required for the extraction of watermark. Additionally, audio watermarking is also used for forensic and security applications [23,32], where the high-quality audio signals are recorded using optical devices [13,14] to ensure minimal background noise. In recent years, blind audio watermarking is gaining more attention despite of the difficulty in realizing such techniques [16,18].

The watermark is embedded in cover audio by modifying the coefficients obtained after applying different transforms such as discrete wavelet transform (DWT) [8,10,16,18,26], fast Fourier transform (FFT) [11], Haar wavelet transform, discrete cosine transform (DCT) [24,37], singular value decomposition (SVD) [22] and integer wavelet transform [9,21]. In [8], a DWT-based audio data hiding technique is presented where the copyright information is encrypted using advance encryption standards (AES) [12] before embedding. In [6], an SVD-based audio watermarking is proposed, robust to signal processing attacks. The original audio is segmented into non-overlapping frames and transformed to matrix. The SVD operation is performed on this matrix where the watermark data are suitably embedded using quantization index modulation (QIM). The watermark can be extracted without the original audio, thereby maintaining the false negative error probability to zero. In [7], a bandwidth extension-based speech watermarking technique is proposed. Various time and frequency domain parameters are extracted from the high-frequency component of speech and embedded in the narrow band speech bit stream by modifying the least significant bit to embed the watermark. A semi-blind DWT–SVD-based watermarking technique is proposed in [3]. The audio signal is decomposed into four-level DWT followed by SVD decomposition. The watermark bits are embedded in the singular matrix obtained after SVD decomposition. The watermarking technique is robust against various signal processing attacks but cannot resist synchronization attack such as random cropping, filtering, pitch shifting and amplitude scaling.

In [11], Fibonacci number-based audio watermarking scheme is proposed and the robustness of the scheme is tested against various signal processing attacks such as filtering, compression and noise addition. The FFT spectrum of the short frames of audio signal is adaptively modified based on Fibonacci numbers to embed the watermark. In [16], DWT-based blind audio watermarking technique is proposed using rational dither modulation (RDM). The audio signal is decomposed in the 5th-level approximation subband using DWT. The watermark bits are embedded by modifying the approximation coefficients using quantization steps derived from previous watermark vectors. The robustness against synchronization attack is achieved using periodic property of RDM.

In [35], an empirical mode decomposition-based audio watermarking scheme is proposed. The watermark bits are embedded by changing the final residual to maintain the imperceptibility. In [22], SVD-based audio watermarking technique is proposed for stereo audio signal. The watermark is embedded in the singular matrix by modifying

the ratio of singular values using QIM. The SVD–QIM-based audio watermarking technique proposed in [22] is robust against re-sampling, re-quantization and other signal processing attacks.

In [15], multilevel DWT–DCT-based blind audio watermarking technique robust to random cropping and time-scale modification attack is proposed. The frame synchronization is achieved by the repetitive pattern of 11th approximation subband and the watermark bits are embedded in first to ninth detail subbands. In [10], synchronization attack resilient DWT–SVD–QIM-based audio watermarking technique is proposed. The audio signal is decomposed into short frames and the watermark bits are embedded in the singular matrix obtained after applying SVD over the DWT coefficients of short audio frames. The robustness against synchronization attack is achieved by embedding synchronization codes along with the watermark bits in the singular matrix using QIM. In [37], the multi-resolution characteristics of DWT and energy compaction capability of DCT are explored to embed the watermark bits. The DCT of fourth-level DWT coefficients are obtained to embed the binary bit stream.

In [18], a blind audio watermarking technique is proposed using DWT and adaptive mean modulation (AMM) technique. The watermark bits and synchronization codes are embedded in second-level approximation and detail subbands, respectively. The BCH codes are used as synchronization code to achieve the robustness against synchronization attack. In [26], DWT–LU (lower upper)-based audio watermarking technique is proposed for high payload applications. To achieve the high imperceptibility, genetic algorithm (GA) is used to identify suitable audio frames for embedding the watermark and synchronization bits. The authors of this paper proposed a DCT–SVD-based audio data hiding technique in voiced frames [24]. The voiced frames are identified by short-time energy and zero crossing rate. The DCT of selected voiced frames are computed followed by SVD to embed the watermark bits adaptively in singular matrix.

In this paper, a novel blind audio watermarking technique is presented employing DCT, Hadamard transform and Schur decomposition resilient to various signal processing and cropping attacks. The watermark is embedded in DCT coefficients of high-energy low-frequency frames (HELF) of cover audio signal using Schur decomposition. To increase the imperceptibility of watermarked audio, the watermark bits are embedded after applying Hadamard transform on DCT coefficients through Schur decomposition. The novelty of this paper comes from the utilization of HELF frames of cover audio signal for embedding the watermark bits using DCT approach along with HT and SD. To the best of our knowledge, there no earlier reports on DCT–HT–SD-based data hiding technique for audio watermarking. Experimental results of the proposed technique shows substantial improvement in the performance evaluation parameters, imperceptibility and robustness. The robustness of proposed technique is tested against noise addition, filtering, re-sampling, re-quantization, MP3 compression, amplitude scaling and cropping attack. The proposed watermarking technique achieves a payload capacity of 500 bps with negligible bit error rate in the signal processing attack environment.

The rest of the paper is organized as follows. Motivation of the proposed technique is presented in Sect. 2. The preliminary technical background is presented in Sect. 3. The

proposed embedding and extraction technique is presented in Sect. 4. Experimental results followed by conclusion are presented in Sects. 5 and 6, respectively.

2 Motivation

Recently we have proposed an audio data hiding scheme based on DCT–SVD technique in [24]. The SVD-based technique generally provides high payload capacity. The important properties due to which SVD is popularly used in audio data hiding are: (i) the slight modification in singular value matrix does not affect the overall quality of signal. (ii) Singular value matrix is invariant to signal processing attacks. (iii) SVD follows the linear algebraic properties [3,5]. In this paper, Schur decomposition is used instead of SVD since the Schur decomposition is an intermediate step of SVD and has lower computational complexity.

In [28], a watermarking technique based on Hadamard transform and Schur decomposition is proposed. Their scheme is proposed for color image watermarking application. In this paper, the proposed audio watermarking scheme utilizes the DCT coefficients of HELF frame for embedding the watermark bits, to increase the robustness and imperceptibility of watermarked audio signals. The DCT is employed in the proposed technique due to its energy compaction characteristics [36]. Further, the HELF frames are chosen for embedding because the poor representation of DCT coefficients of low-energy frames will degrade the perceptual quality of audio signal [30].

3 Preliminary

3.1 Discrete Cosine Transform

DCT decomposes a finite sequence into sum of cosine functions of different frequencies [1]. The 1-D DCT of a signal $x(n)$ having length N is defined as:

$$X(k) = w(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad k = 0, 1, \ldots, N-1. \qquad (1)$$

where

$$w(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{if } k = 0 \\ \sqrt{\frac{2}{N}} & \text{Otherwise} \end{cases}$$

In the proposed audio watermarking technique, the data are embedded in the DCT coefficients of low-frequency components of the cover audio. The inherent property of DCT to maintain the shape of low-frequency components even after re-sampling is exploited in the proposed technique to resist against re-sampling attack. The DCT

sequence of re-sampled signal is proportional to original DCT sequence, where the proportionality constant is the square root of re-sampling factor [15].

3.2 Hadamard Transform

The Hadamard transform decomposes a signal into set of orthogonal, rectangular waveforms called Walsh function [2]. Hadamard transform of a sequence $\{x(k)\} = \{x(0), x(1), \ldots, x(n-1)\}$ is defined as

$$\{X(g)\} = \frac{1}{n}[H(g)]\{x(g)\} \tag{2}$$

where $\{x(g)\}^T = \{x(0), x(1), \ldots, x(n-1)\}$, $g = \log_2 n$ is the representation of signal $\{x(k)\}$, $H(g)$ represents the Hadamard matrix of dimension $n \times n$ also $[H(g)] = [H(g)]^* = [H(g)]^T = [H(g)]^{-1}$ and satisfies the orthogonality, i.e., $[H(g)]^T[H(g)] = I(g)(I = Identity\ matrix)$ and $\{X(g)\}^T = \{X(0)X(1), \ldots, X(n-1)\}$ represents the HT coefficients. The signal can be recovered from the inverse HT:

$$\{x(g)\} = [H(g)]\{X(g)\} \tag{3}$$

The HT is used in the proposed audio watermarking technique due to its equivalency with multidimensional discrete Fourier transform (DFT) and also it is invariant to any cyclic shift [27]. In addition, the specific characteristics of Hadamard transform such as low computational complexity, suitability for data hiding and ease of implementation [2,28] have made the proposed watermarking technique unique and advantageous. In the proposed watermarking technique, 2D-HT is used for hiding the watermark data. The 2D-HT of $[x]_{n \times n}$ is obtained by Eq. 4:

$$[X]_{n \times n} = \frac{1}{n}[H][x][H] \tag{4}$$

The 2D inverse HT is given as [33]

$$[x]_{n \times n} = \frac{1}{n}[H][X][H] \tag{5}$$

3.3 Schur Decomposition

Schur decomposition is an important tool which can decompose a given $n \times n$ matrix $A \in \mathbb{R}$, into set of orthonormal basis $B \in \mathbb{R}$ and upper-triangular matrix S with entries in $\mathbb{R}$, where $\mathbb{R}$ denotes the real vector space. The Schur decomposition equally holds true for complex vector space $\mathbb{C}$ also [4].
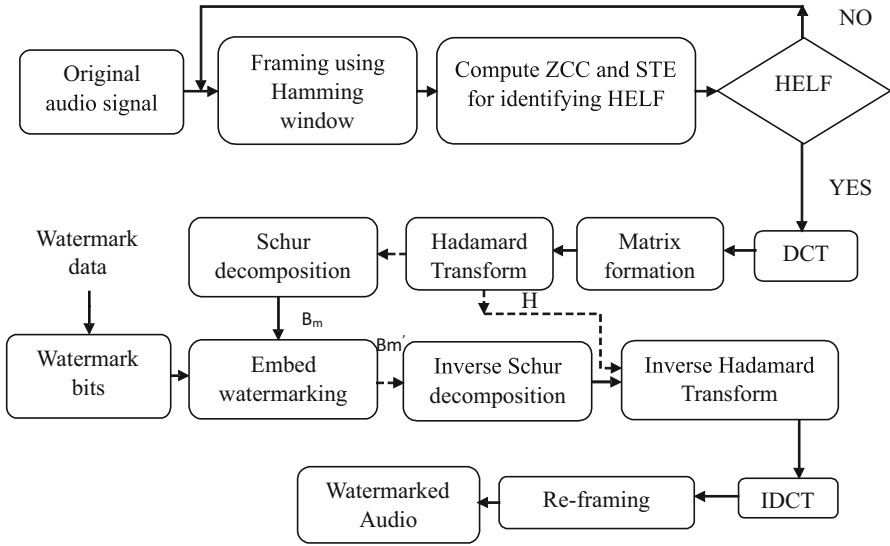
Fig. 1 Proposed DCT–HT–SD-based audio watermark embedding process

$$A = \begin{bmatrix} \vdots & & \vdots \\ \mathbf{b_1} & \cdots & \mathbf{b_n} \\ \vdots & & \vdots \end{bmatrix} \times S \times \begin{bmatrix} \vdots & & \vdots \\ \mathbf{b_1} & \cdots & \mathbf{b_n} \\ \vdots & & \vdots \end{bmatrix}^{-1} = B \times S \times B^{-1}$$

The strong correlation between the elements of the column vector $\mathbf{b_1}$ is exploited in the proposed audio watermarking technique for hiding the watermark data [34].

4 Proposed Embedding Technique

The proposed embedding process is illustrated in Fig. 1. The various steps involved in the proposed watermark embedding process is explained as follows.

Step 1 The original audio signal is separated into non-overlapping frames using Hamming window function. The zero crossing rate and short-time energy are computed using Eqs. 6 and 7 [31,40]. If $s(n)$ represents a signal with L ms duration and N numbers of samples per frame, then the short-time energy STE is given by Eq. 6. In the proposed watermarking technique, the value of L is considered as 10 ms because the audio signals are reasonably considered as stationary and periodic [31]. The value of N is computed as $N = f_s \times L$, where f_s is sampling frequency.

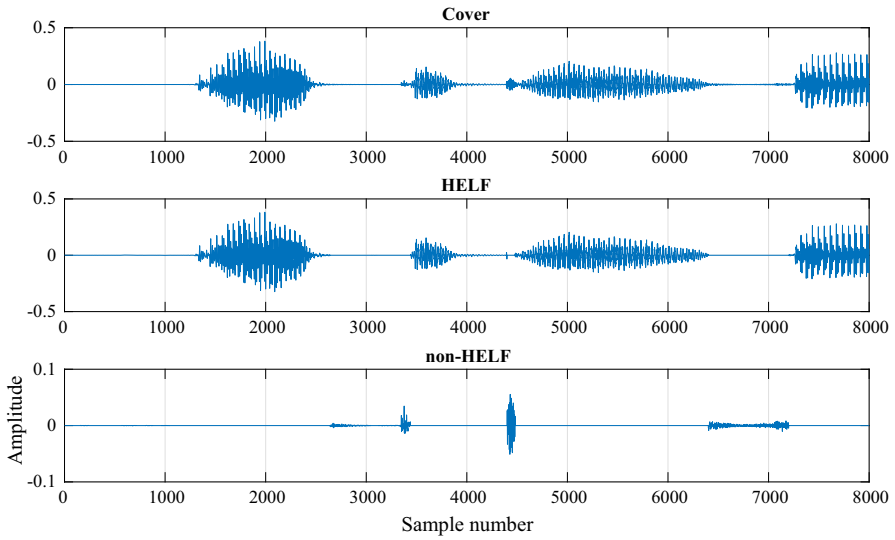$$\text{STE} = \sum_{m=n-(N-1)}^{n} [s(m)w(m-n)]^2 \qquad (6)$$

Fig. 2 Separation of HELFs from the original audio signal. The audio signal is sampled at 8 kHz

where

$$w(n) = \begin{cases} 0.54 - 0.46\cos\frac{2\pi n}{N-1} & \text{if } 0 \le n \le N - 1 \\ 0 & \text{otherwise} \end{cases}$$

The ZCR of speech signal computed by Eq. 7:

$$ZCR = \sum_{n=0}^{N-1} 0.5|\text{sign}(s[n]) - \text{sign}(s[n-1])| \tag{7}$$

where

$$\text{sign}(s[n]) = \begin{cases} +1 & \text{if } s[n] \ge 0 \\ -1 & \text{otherwise} \end{cases}$$

The frame having high STE and low ZCR is identified as a HELF and embedding is done on such selected frames only. The separation of HELF frames from the audio taken from the NOIZEUS database [19,20,29] is shown in Fig. 2.

Step 2 The HELF frame is designated as $x(mn)$ and it is further divided into m sub-frames, each having n samples, $x(mn) = \{x_1(n), x_2(n), \ldots, x_m(n)\}$. The 1D-DCT operation is performed on the HELF sub-frame $x_m(n)$ using Eq. 1 to obtain $X_m(k)$ where, $k = 0, 1, 2, \ldots, 15$. The DCT of a HELF frame is shown in Fig. 3. Obtained DCT coefficients of each sub-frame are arranged in a matrix form of dimension 4×4 to apply 2D HT and SD, and subsequently to embed the watermark bits.
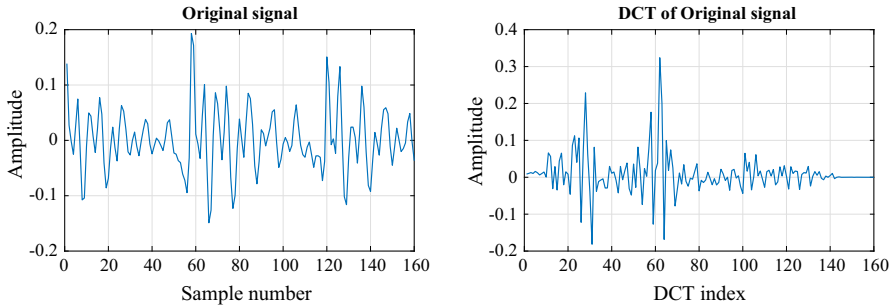
Fig. 3 Two HELFs from the original audio signal and its DCT. The audio signal is sampled at 8 kHz

Step 3 Let $[X_m(k)]$ denote the matrix having DCT coefficients of mth sub-frame.

$$[X_m(k)] = \begin{bmatrix} X_m(0) & X_m(1) & X_m(2) & X_m(3) \\ X_m(4) & X_m(5) & X_m(6) & X_m(7) \\ X_m(8) & X_m(9) & X_m(10) & X_m(11) \\ X_m(12) & X_m(13) & X_m(14) & X_m(15) \end{bmatrix}$$

After generating the matrix of DCT coefficients as shown above, the HT operation is performed using Eq. 4:

$$\left[X_m^H(k)\right] = \frac{1}{k}[H][X_m(k)][H] \tag{8}$$

where, $[X_m^H(k)]$ denotes the HT of mth sub-frame DCT matrix and $[H]$ is the Hadamard matrix.

Step 4 The SD operation is performed on the matrix $[X_m^H(k)]$ to obtain $[B_m]$ and $[S_m]$.

$$\left[X_m^H(k)\right] = [B_m][S_m][B_m]^{\mathrm{T}} \tag{9}$$

Step 5 Each watermark bit W is embedded in one sub-frame by modifying the elements of column $\mathbf{b_1} = \{b_1(1), b_1(2), b_1(3), b_1(4)\}$ of $[B_m]$. The elements of $\mathbf{b_2}$, $\mathbf{b_3}$ and $\mathbf{b_4}$ are uncorrelated with each other hence these vectors are not considered for embedding the watermark bits [34]. The watermark bits are inserted using below mentioned scheme [34]:

$$b_1'(2) = \begin{cases} \mathrm{sign}(b_1(2)) \times (1+\delta)\mathrm{avg} & \text{if } W = 1 \\ \mathrm{sign}(b_1(2)) \times (1-\delta)\mathrm{avg} & \text{if } W = 0 \end{cases} \tag{10}$$

$$b_1'(3) = \begin{cases} \mathrm{sign}(b_1(3)) \times [\mathrm{abs}(b_1(3)) - [\delta \times \mathrm{avg} \times 0.5]] & \text{if } W = 1 \\ \mathrm{sign}(b_1(3)) \times [\mathrm{abs}(b_1(3)) + [\delta \times \mathrm{avg} \times 0.5]] & \text{if } W = 0 \end{cases} \tag{11}$$

Fig. 4 Effect of change in δ on the watermarked audio for $\delta = 0.03, 0.09$ and 0.3

$$b_1'(4) = \begin{cases} \text{sign}(b_1(4)) \times [\text{abs}(b_1(4)) - [\delta \times \text{avg} \times 0.5]] & \text{if } W = 1 \\ \text{sign}(b_1(4)) \times [\text{abs}(b_1(4)) + [\delta \times \text{avg} \times 0.5]] & \text{if } W = 0 \end{cases} \quad (12)$$
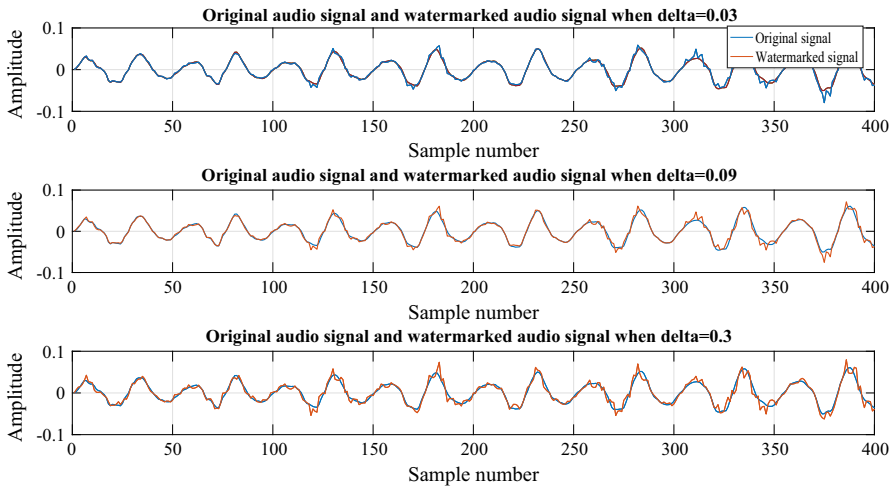
where $\text{sign}(y)$ denotes the sign of y, $\text{abs}(y)$ denotes the absolute value of y, $\text{avg} = (|b_1(3)| + |b_1(4)|)/2$ and δ is the control factor. The reason for not considering $(b_1(1))$ for embedding the watermark is that the correlation between $(b_1(1))$ and any other element of $\mathbf{b_1}$ is least [34]. The effect of variation in δ on the proposed method for $\delta = 0.03, 0.09$ and 0.3 is shown in Fig. 4. It is clear from Fig. 4 that for $\delta = 0.03$ the overlapping of watermarked audio signal on original audio signal is maximum when compared with the cases for $\delta = 0.09$ and $\delta = 0.3$. The lower value of δ reduces the distortion but increases the bit error rate during the extraction process.

After embedding the watermark bits, the modified $[B_m]$ matrix is obtained. The inverse SD is performed on modified $[B_m]$ using $[S_m]$ followed by inverse HT using $[H]$ to obtain the modified DCT coefficients. The IDCT operation is performed on these modified coefficients. The HELF frames are recreated and arranged to obtain the watermarked audio.

5 Proposed Extraction Technique

The proposed extraction process is shown in Fig. 5. The steps involved in proposed watermark extraction process is explained as follows. The initial steps involved in extraction of watermark bits are same as explained in embedding process.

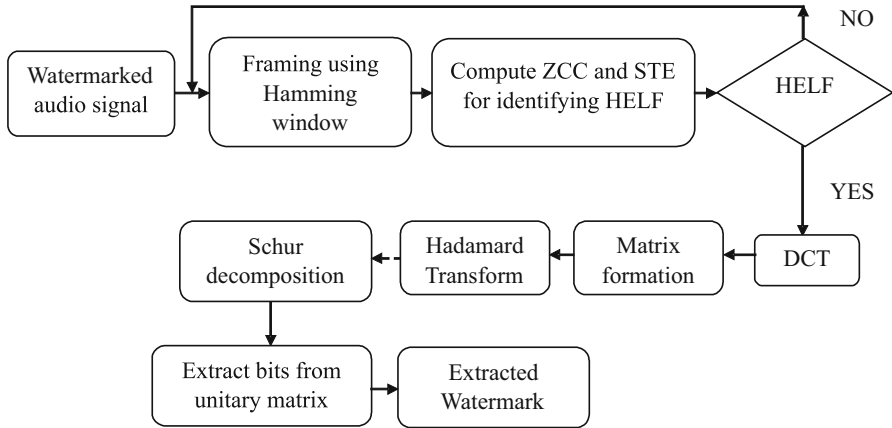*Step 1* The HELFs are identified from the watermarked audio signal.

Fig. 5 Proposed DCT–HT–SD-based audio watermark extraction process

Step 2 The DCT operation is performed on each HELF frame followed by HT.

Step 3 After performing HT and SD operation on the watermarked mth sub-frame i.e., $[\hat{X}_m^H(k)]$, the $[\hat{B}_m]$ and $[\hat{S}_m]$ are obtained respectively.

$$\left[\hat{X}_m^H(k)\right] = [\hat{B}_m][\hat{S}_m][\hat{B}_m]^{\mathrm{T}} \tag{13}$$

The watermark bit $\hat{W}$ is extracted from each sub-frame by using Eqs. 14 and 15.

$$\hat{W} = 0; \quad if \begin{cases} \mathrm{abs}(\hat{b}_1(3)) < \widehat{\mathrm{avg}} \ \&\& \ \mathrm{abs}(\hat{b}_1(4)) > \widehat{\mathrm{avg}} \\ \qquad\qquad or \\ \mathrm{abs}(\hat{b}_1(3)) > \widehat{\mathrm{avg}} \ \&\& \ \mathrm{abs}(\hat{b}_1(4)) < \widehat{\mathrm{avg}} \end{cases} \tag{14}$$

$$\hat{W} = 1; \quad if \begin{cases} \mathrm{abs}(\hat{b}_1(3)) < \widehat{\mathrm{avg}} \ \&\& \ \mathrm{abs}(\hat{b}_1(4)) < \widehat{\mathrm{avg}} \\ \qquad\qquad or \\ \mathrm{abs}(\hat{b}_1(3)) > \widehat{\mathrm{avg}} \ \&\& \ \mathrm{abs}(\hat{b}_1(4)) > \widehat{\mathrm{avg}} \end{cases} \tag{15}$$

where, $\hat{b}_1(3)$ and $\hat{b}_1(4)$ are the elements of column $\hat{\mathbf{b}}_1 = \{\hat{b}_1(1), \hat{b}_1(2), \hat{b}_1(3), \hat{b}_1(4)\}$ of $[\hat{B}_m]$ and $\widehat{\mathrm{avg}} = (|\hat{b}_1(3)|+|\hat{b}_1(4)|)/2$.

6 Experimental Results

The performance of proposed audio watermarking technique including imperceptibility and robustness is verified using NOIZEUS audio database [19,20,29]. All the audio signals are sampled at 8 kHz. For the fair comparison with other watermarking methods, the watermark bits-stream of 1s and 0s are placed alternately to cover entire signal [16]. A plot depicting the cover audio signal and watermarked audio signal is shown in Fig. 6 where the embedding portions are highlighted with box. The watermark bits are
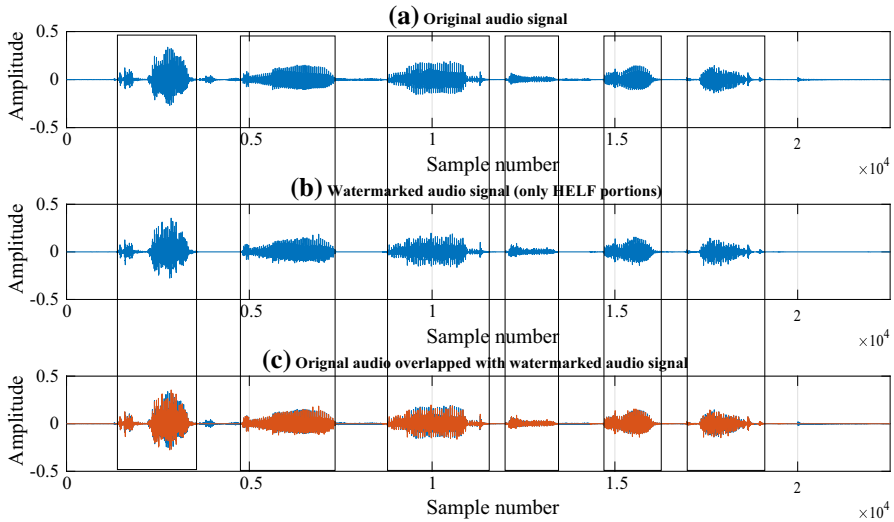
Fig. 6 **a** Plot depicting original audio signal, **b** watermarked audio signal with only HELF frames where watermark is embedded, **c** overlapping of original audio with watermarked audio signal. The HELF frames are highlighted using a rectangle to show that non-HELF frames are not used of embedding

embedded only in HELF frames hence in Fig. 6b non-HELF portion for watermarked audio is not shown. The plot in Fig. 6a, b is overlapped in Fig. 6c to show the similarity between the original audio and watermarked audio. The parameters involved in the DCT–HT–SD experiment are: $L = 10$ ms, $\delta = 0.031$ and each sub-frame consists of 16 samples. The proposed audio watermarking technique is compared with six other schemes, which are abbreviated as DWT–LU [26], DWT–RDM [16], MDWT–VWM [15], DWT–AMM [18], DWT–SVD [10], DWT–Fibonacci [11] and DWT–SVD [3].

The imperceptibility and robustness of proposed watermarking technique is evaluated using the signal-to-noise ratio (SNR), and normalized correlation coefficient (NCC) and bit error rate, respectively. The SNR, NCC and BER are evaluated using Eqs. 16, 17 and 18, respectively [26].

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_n \hat{s}^2(n)}{\sum_n \left(\hat{s}(n) - s(n) \right)^2} \right) \qquad (16)$$

where $\hat{s}(n)$ and $s(n)$ denote the watermarked audio and original audio signals, respectively.

$$\text{NCC} = \left\{ \frac{\sum_{i=1}^{M} X(i) \hat{X}(i)}{\sqrt{\sum_{i=1}^{M} X^2(i)} \times \sqrt{\sum_{i=1}^{M} \hat{X}^2(i)}} \right\} \qquad (17)$$

where $X(i)$ denotes original audio signal, $\hat{X}(i)$ denotes watermarked audio signal and M denotes the total number of samples in the audio signal.
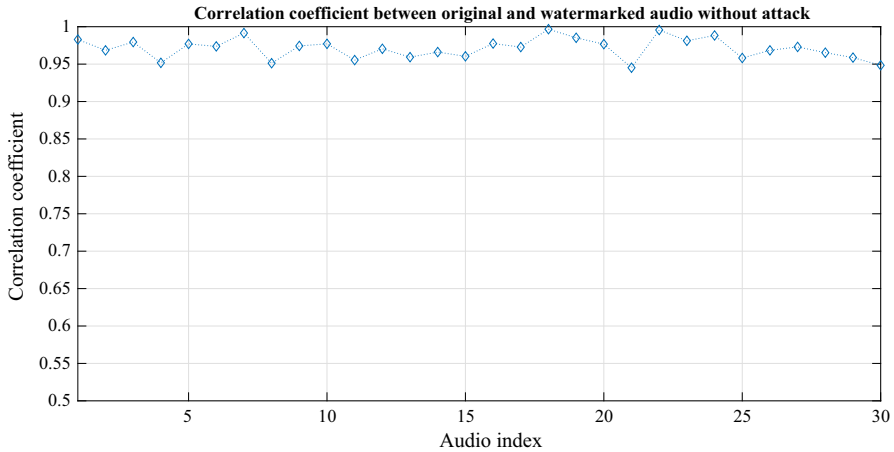
Fig. 7 Correlation coefficient between original and watermarked audio signal

$$\mathrm{BER} = \frac{\sum_{i=1}^{N} W(i) \oplus \hat{W}(i)}{N} \tag{18}$$

where $W(i)$, $\hat{W}(i)$ denote the embedded and extracted watermark bits, N denotes the number watermark bits and $\oplus$ denotes exclusive OR operator, respectively. To test the similarity between original audio signal and watermarked audio signal, the NCC is computed for all the audio as shown in Fig. 7. The higher NCC achieved confirms that the changes introduced because of embedding watermark are small and at acceptable limits.

The imperceptibility and payload results of proposed watermarking techniques without any attack are listed in Table 1.

It can be seen from Table 1 that SNR is much higher than the minimum requirement set by the International Federation of Phonographic Industry (IFPI) [25]. The average SNR obtained is 41.82 dB which is twice than the requirement, i.e., 20 dB. The NCC and BER values confirm that the higher accuracy is achieved from the proposed watermarking technique. The payload in each test is also mentioned in Table 1 to clearly indicate the performance of the proposed watermarking technique.

Comparison of SNR and payload of proposed technique with other recent techniques is shown in Table 2. The SNR achieved by proposed technique is the highest among all other techniques because of selecting the HELF frames for embedding. Although the payload of the proposed watermarking technique is third highest among all the techniques under comparison, the SNR is highest instead. The proposed DCT–HT–SD-based audio watermarking technique gives payload of 500.85 bps because the cover audio is sampled at 8 kHz and divided into 10 ms frames and each frame is further divided in 5 sub-frames having 16 samples per sub-frame. In each sub-frame, one bit of watermark data is embedded and the payload (bps) is calculated using Eq. 19

$$Payload = \frac{N_{\mathrm{wb}}}{L_{\mathrm{o}}} \tag{19}$$

Table 1 Imperceptibility and payload results without attack

Audio	SNR (dB)	NCC	BER (%)	Payload (bps)
sp01	46.45	0.9429	0.0	500.69
sp02	40.57	0.9284	0.0	500.02
sp03	43.51	0.9395	0.0	500.92
sp04	35.65	0.9117	0.0	500.95
sp05	44.14	0.937	0.0	501.50
sp06	42.41	0.9337	0.0	500.89
sp07	49.51	0.9515	0.0	501.12
sp08	33.12	0.9112	0.0	500.79
sp09	42.38	0.9343	0.0	500.00
sp10	43.63	0.9371	0.0	501.61
sp11	35.63	0.9154	0.0	500.00
sp12	40.87	0.9305	0.0	500.00
sp13	37.45	0.9193	0.0	500.00
sp14	40.96	0.9261	0.0	501.15
sp15	38.39	0.9204	0.0	500.79
sp16	42.93	0.9375	0.0	500.51
sp17	41.8	0.9328	0.0	501.27
sp18	52.69	0.9567	0.0	501.40
sp19	47.37	0.9452	0.0	500.52
sp20	42.89	0.9367	0.0	501.16
sp21	35.17	0.9052	0.0	500.29
sp22	52.18	0.9556	0.0	500.42
sp23	46.19	0.9412	0.0	501.67
sp24	48.82	0.9483	0.0	501.30
sp25	37.86	0.9182	0.0	501.64
sp26	40.36	0.9284	0.0	501.66
sp27	41.74	0.933	0.0	500.00
sp28	37.63	0.9253	0.0	500.85
sp29	37.65	0.9187	0.0	501.74
sp30	34.62	0.9084	0.0	500.81

where N_{wb} denotes the number of watermark bits embedded, and L_o is the length of the original audio signal in seconds [10].

The robustness of the proposed watermarking technique against various attacks is assessed by calculating BER between original watermark W and extracted watermark $\widehat{W}$. The different types of attacks considered in this study are mentioned in Table 3. The type of the attacks and their corresponding abbreviations along with the description are mentioned in Table 3.

The average BERs of other recent audio watermarking techniques are compared with the proposed technique in Table 4.

	Watermarking techniques	SNR (dB)	Payload (bps)
Table 2 Comparison of SNR and payload of proposed DCT–HT–SD watermarking technique with different watermarking techniques	DWT–LU [26]	30.94	1280
	DWT–AMM [18]	30.61	100
	DWT–RDM [16]	20.23	344.53
	DWT–SVD [3]	38.17	516
	DWT–Fibonacci [11]	35	683
	EMD [35]	39.46	–
	SVD–QIM [22]	–	187.5
	MDWT–VWM [15]	15.95	86.13
	DWT–DCT–AMM [37]	23.49	172.27
	Proposed DCT–HT–SD	41.82	500.85

Table 3 Attack types and description

Attack types	Description
Re-sampling attack (RSA)	Watermarked audio is sampled with different sampling frequency and re-sampled using original sampling frequency
Re-quantization attack (RQA)	Watermarked audio is quantized with different level and re-quantized using original quantization level
MP3 compression attack (MCA)	Watermarked audio is compressed with different rate to corrupt the watermark bits
AWGN (AA)	Additive white Gaussian noise is added to the watermarked audio different SNR to corrupt the watermark data bits
Low pass filtering attack (LPFA)	A low pass filter of cutoff frequency 1 kHz is applied to the watermarked signal
High pass filtering attack (HPFA)	A high pass filter of cutoff frequency 50 Hz is applied to the watermarked signal
Random cropping attack-1 (RCA-1)	Replacing the first 200 samples of watermarked audio by 200 samples of watermarked audio processed by low pass filter [6]
Random cropping attack-2 (RCA-2)	Eliminating the first 200 samples of watermarked audio signal [15]
Amplitude scaling attack (ASA)	Scaling the amplitude of watermarked signal by 0.7

All the watermarking techniques compared in Table 4 can restore the watermark without any error when attacks are absent. The SVD–QIM [22] is more susceptible to re-sampling attack compared to the proposed DCT–HT–SD technique. The reason for achieving 0 BER against re-sampling attack in the proposed watermarking technique is that the watermark is embedded in low-frequency DCT coefficients of audio samples and DCT possess a property to retain the shape of low-frequency components even after re-sampling [15]. The various methods compared in Table 4 show less BER for RQA and MCA because all the methods embed the watermark bits in low-frequency components. The watermark is embedded artfully in low-frequency components, or in the approximation coefficients of lowest subband after computing DWT. Since,

Table 4 Average BER results of the proposed DCT–HT–SD watermarking technique compared with other watermarking techniques

Attacks	DWT–SVD [3]	SVD–QIM [6]	DWT–RDM [16]	SVD–QIM [22]	DWT–VM [17]	DWT–LU [26]	MDWT–VWM [15]	DWT–AMM [18]	DCT–SVD [24]	DWT–DCT–AMM [37]	Proposed method
None	0	0	0	0.10	0	0	0	0	0	0	0
RSA	0	0	0	4.88	0	0	0.90	0	0.003	0.01	0
RQA	0	0	0	–	0	–	0	0.03	0	0.14	0
MCA 64 kbps	0.0820	0.561	0	24.56	0.05	0.02	0.01	–	0.0013	0.08	0
MCA 32 kbps	0.2901	2.204	1.05	–	2.75	0.02	0.01	–	0.0014	0.01	0.01
AA 20dB	0	0	0	–	0	–	0.03	9.32	0	2.27	0
AA 30dB	0	0	0	10.25	0.04	0	0.04	1.38	0	0.22	0
LPFA (4 kHz)	0.1893	0.512	0.08	0.31	0.98	0.03	0.04	0.03	–	0.01	0
HPF (50 Hz)	0.358	–	–	–	–	–	–	8.93	–	–	18.62
RCA-1	–	0	–	–	–	0.1126	–	–	–	–	1.02
RCA-2	–	–	–	–	–	–	0.89, 46.14	–	–	–	8.10
ASA	–	–	0	0.38	0	–	0	0	–	0.01	0.001

the low-frequency components undergo less perceptible distortion compared to the high-frequency components.

The BER in case of RQA, MCA and AA are lowest or second lowest because the intensity of noise added due to the attack is low compared to watermark noises. The proposed technique achieves 0 BER for LPFA because only the low-frequency components are used for embedding the watermark. The BER for HPFA is high because the attack filters out the low-frequency components where watermark bits are embedded. In case of ASA, the BER of proposed technique is lowest because the watermarking is performed using orthonormal basis of Schur decomposition, and the orthonormal basis are independent of amplitude scaling [4]. The slight increase in the BER in case of RCA-1 is primarily due to the frame misalignment and it can be reduced further by incorporating synchronization codes. The BER for RCA-1 attack is higher than the BER of DWT–LU [26], because in the proposed techniques the cropping percentage is 2.5% whereas, in DWT–LU [26] the cropping percentage is 1%. The BER values in case of RCA-2 attack for MDWT–VWM [15] techniques with and without using synchronization codes are 0.89 and 46.14, respectively. Whereas, the BER in case of RCA-2 attack for the proposed technique is limited to lower value without using synchronization codes is due to the fact that the initial part of speech contains high-frequency low-energy frames (silence or unvoiced sound) [31], and in the proposed watermarking technique these frames are not used for embedding the watermark data.

7 Conclusion

In this paper, a novel blind audio watermarking technique using DCT–HT–SD is presented. The distinctive features of DCT, HT and SD are strategically explored to achieve a robust and imperceptible watermarking technique. The imperceptibility is achieved by selecting the HELF frames of watermarked audio. The embedding of watermark is performed using the average of the elements of B matrix obtained after SD to resist various signal processing attacks.

The proposed method achieves a payload of 500.85 bps. Experimental results validate that the proposed technique maintains the perceptual quality of watermark audio. The comparison of robustness results with other watermarking technique indicates that the proposed technique is robust to re-sampling, re-quantization, AWGN, MP3 compression, LPFA and amplitude scaling attacks. The proposed technique gives negligible BER in signal processing attack environment. The future research will focus on enhancing the proposed audio watermarking technique to resist time-scale modification, jittering and synchronization attacks. The proposed watermarking technique can be made robust to such attacks by incorporating the synchronization codes while embedding watermark bits.

References

1. N. Ahmed, T. Natarajan, K.R. Rao, Discrete cosine transform. IEEE Trans. Comput. **C–23(1)**, 90–93 (1974). https://doi.org/10.1109/T-C.1974.223784

2. N. Ahmed, K. Rao, A. Abdussattar, BIFORE or Hadamard transform. IEEE Trans. Audio Electroacoust. **19**(3), 225–234 (1971). https://doi.org/10.1109/TAU.1971.1162193
3. A. Al-Haj, An imperceptible and robust audio watermarking algorithm. EURASIP J. Audio Speech Music Process. **2014**(1), 37 (2014). https://doi.org/10.1186/s13636-014-0037-2
4. S. Axler, *Linear Algebra Done Right* (Springer, Berlin, 1997)
5. W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding. IBM Syst. J. **35**(3.4), 313–336 (1996). https://doi.org/10.1147/sj.353.0313
6. K.V. Bhat, I. Sengupta, A. Das, A new audio watermarking scheme based on singular value decomposition and quantization. Circuits Syst. Signal Process. **30**(5), 915–927 (2011). https://doi.org/10.1007/s00034-010-9255-8
7. Z. Chen, C. Zhao, G. Geng, F. Yin, An audio watermark-based speech bandwidth extension method. EURASIP J. Audio Speech Music Process. **2013**(1), 10 (2013). https://doi.org/10.1186/1687-4722-2013-10
8. K. Datta, I. Sengupta, A redundant audio watermarking technique using discrete wavelet transformation, in *2010 Second International Conference on Communication Software and Networks* (2010), pp. 27–31. https://doi.org/10.1109/ICCSN.2010.57
9. F. Djebbar, B. Ayad, K.A. Meraim, H. Hamam, Comparative study of digital audio steganography techniques. EURASIP J. Audio Speech Music Process. **2012**(1), 25 (2012). https://doi.org/10.1186/1687-4722-2012-25
10. A.R. Elshazly, M.E. Nasr, M.M. Fouad, F.S. Abdel-Samie, High payload multi-channel dual audio watermarking algorithm based on discrete wavelet transform and singular value decomposition. Int. J. Speech Technol. **20**(4), 951–958 (2017)
11. M. Fallahpour, D. Megas, Audio watermarking based on fibonacci numbers. IEEE/ACM Trans. Audio Speech Lang. Process. **23**(8), 1273–1282 (2015). https://doi.org/10.1109/TASLP.2015.2430818
12. A. Farmani, H.B. Bahar, Hardware implementation of 128-bit AES image encryption with low power techniques on FPGA to VHDL. Majlesi J. Electr. Eng. **6**(4), 13–22 (2012)
13. A. Farmani, M. Farhang, M.H. Sheikhi, High performance polarization-independent quantum dot semiconductor optical amplifier with 22 db fiber to fiber gain using mode propagation tuning without additional polarization controller. Opt. Laser Technol. **93**, 127–132 (2017)
14. A. Farmani, M.H. Sheikhi, Quantum-dot semiconductor optical amplifier: performance and application for optical logic gates. Majlesi J. Telecommun. Dev. **6**(3), 85–89 (2017)
15. H.T. Hu, J.R. Chang, Efficient and robust frame-synchronized blind audio watermarking by featuring multilevel DWT and DCT. Cluster Comput. **20**(1), 805–816 (2017). https://doi.org/10.1007/s10586-017-0770-2
16. H.T. Hu, L.Y. Hsu, A DWT-based rational dither modulation scheme for effective blind audio watermarking. Circuits Syst. Signal Process. **35**(2), 553–572 (2016). https://doi.org/10.1007/s00034-015-0074-9
17. H.T. Hu, L.Y. Hsu, Incorporating spectral shaping filtering into DWT-based vector modulation to improve blind audio watermarking. Wirel. Pers. Commun. **94**(2), 221–240 (2017). https://doi.org/10.1007/s11277-016-3178-z
18. H.T. Hu, S.J. Lin, L.Y. Hsu, Effective blind speech watermarking via adaptive mean modulation and package synchronization in DWT domain. EURASIP J. Audio Speech Music Process. **2017**(1), 10 (2017)
19. Y. Hu, P.C. Loizou, Subjective comparison and evaluation of speech enhancement algorithms. Speech Commun. **49**(7), 588–601 (2007)
20. Y. Hu, P.C. Loizou, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Speech Audio Process. **16**(1), 229–238 (2008). https://doi.org/10.1109/TASL.2007.911054
21. G. Hua, J. Huang, Y.Q. Shi, J. Goh, V.L. Thing, Twenty years of digital audio watermarking—a comprehensive review. Signal Process. **128**, 222–242 (2016). https://doi.org/10.1016/j.sigpro.2016.04.005
22. M.J. Hwang, J. Lee, M. Lee, H.G. Kang, SVD-based adaptive QIM watermarking on stereo audio signals. IEEE Trans. Multimed. **20**(1), 45–54 (2018)
23. M. Imran, Z. Ali, S.T. Bakhsh, S. Akram, Blind detection of copy-move forgery in digital audio forensics. IEEE Access **5**, 12843–12855 (2017). https://doi.org/10.1109/ACCESS.2017.2717842
24. A. Kanhe, G. Aghila, A DCT-SVD-based speech steganography in voiced frames. Circuits Syst. Signal Process. **37**(11), 5049–5068 (2018). https://doi.org/10.1007/s00034-018-0805-9

25. S. Katzenbeisser, F.A. Petitcolas (eds.), *Information Hiding Techniques for Steganography and Digital Watermarking*, 1st edn. (Artech House, Norwood, 2000)
26. A. Kaur, M.K. Dutta, An optimized high payload audio watermarking algorithm based on LU-factorization. Multimed. Syst. (2017). https://doi.org/10.1007/s00530-017-0545-x
27. H. Kunz, On the equivalence between one-dimensional discrete Walsh–Hadamard and multidimensional discrete Fourier transforms. IEEE Trans. Comput. **28**, 267–268 (1979). https://doi.org/10.1109/TC.1979.1675334
28. J. Li, C. Yu, B.B. Gupta, X. Ren, Color image watermarking scheme based on quaternion Hadamard transform and Schur decomposition. Multimed. Tools Appl. **77**(4), 4545–4561 (2018). https://doi.org/10.1007/s11042-017-4452-0
29. J. Ma, Y. Hu, P.C. Loizou, Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J. Acoust. Soc. Am. **125**(5), 3387–3405 (2009)
30. W.K. McDowell, W.B. Mikhael, A.P. Berg, Efficiency of the KLT on voiced & unvoiced speech as a function of segment size, in *2012 Proceedings of IEEE Southeastcon* (2012), pp. 1–5. https://doi.org/10.1109/SECon.2012.6197063
31. L.R. Rabiner, R.W. Schafer, *Digital Processing of Speech Signals*, vol. 100 (Prentice-Hall, Englewood Cliffs, 1978)
32. D. Renza, D.M. Ballesteros Larrota, C. Lemus, Authenticity verification of audio signals based on fragile watermarking for audio forensics. Expert Syst. Appl. **91**, 211–222 (2018). https://doi.org/10.1016/j.eswa.2017.09.003
33. V. Santhi, P. Arulmozhivarman, Hadamard transform based adaptive visible/invisible watermarking scheme for digital images. J. Inf. Secur. Appl. **18**(4), 167–179 (2013). https://doi.org/10.1016/j.istr.2013.01.001
34. Q. Su, Y. Niu, X. Liu, Y. Zhu, Embedding color watermarks in color images based on schur decomposition. Opt. Commun. **285**(7), 1792–1802 (2012). https://doi.org/10.1016/j.optcom.2011.12.065
35. X. Tang, Z. Ma, X. Niu, Y. Yang, Robust audio watermarking algorithm based on empirical mode decomposition. Chin. J. Electron. **25**(6), 1005–1010 (2016). https://doi.org/10.1049/cje.2016.06.007
36. X.Y. Wang, H. Zhao, A novel synchronization invariant audio watermarking scheme based on DWT and DCT. IEEE Trans. Signal Process. **54**(12), 4835–4840 (2006). https://doi.org/10.1109/TSP.2006.881258
37. Q. Wu, M. Wu, A novel robust audio watermarking algorithm by modifying the average amplitude in transform domain. Appl. Sci. **8**(5), 2076–3417 (2018)
38. S. Xiang, H.J. Kim, J. Huang, Audio watermarking robust against time-scale modification and MP3 compression. Signal Process. **88**(10), 2372–2387 (2008). https://doi.org/10.1016/j.sigpro.2008.03.019
39. S. Xiang, L. Yang, Y. Wang, Robust and reversible audio watermarking by modifying statistical features in time domain. Adv. Multimed. (2017). https://doi.org/10.1155/2017/8492672
40. P. Yao, K. Zhou, Application of short time energy analysis in monitoring the stability of arc sound signal. Measurement **105**, 98–105 (2017). https://doi.org/10.1016/j.measurement.2017.04.015