Kanhe and Gnanasekaran *EURASIP Journal on Audio, Speech, and Music Processing* (2018) 2018:16
https://doi.org/10.1186/s13636-018-0139-3

EURASIP Journal on Audio, Speech, and Music Processing

RESEARCH Open Access

Robust image-in-audio watermarking technique based on DCT-SVD transform

Aniruddha Kanhe[1*] and Aghila Gnanasekaran[2]

Abstract

In this paper, a robust and highly imperceptible audio watermarking technique is presented based on discrete cosine transform (DCT) and singular value decomposition (SVD). The low-frequency components of the audio signal have been selectively embedded with watermark image data making the watermarked audio highly imperceptible and robust. The imperceptibility of proposed methods is evaluated by computing signal-to-noise ratio and by conducting subjective listening tests. The robustness of proposed technique is evaluated by computing bit error rate and average information loss in retrieved watermark image subjected to MP3 compression, AWGN, re-sampling, re-quantization, amplitude scaling, low-pass filtering, and high-pass filtering attacks with high data payload of 6 kbps. The information-theoretic approach is used to model the proposed watermarking technique as discrete memoryless channel. The Shannon's entropy concept is used to highlight the robustness of proposed technique by computing the information loss in retrieved watermarked image.

Keywords: Audio watermarking, DCT, SVD, Voiced frames, Unvoiced frames

1 Introduction

The rapid developments in the field of digital audio technology have increased the ease to store, distribute and reproduce the audio files. This leads to an inherent security risk of illegal data usage and copyright violation. The digital audio watermarking technique provides a promising solution to protect such copyright violation [1]. The digital watermark is a check to illegal copying of data and identifies the copyright infringement [2].

Digital audio watermarking is the process of embedding owner's signature or copyright information in audio signal (cover media). The watermark data can be text or image (logo) and can be utilized for protection of copyright, authentication, and deterrent illegal copying of audio files [3]. The performance of audio watermarking techniques is evaluated on three major categories: imperceptibility, robustness, and payload as shown in Fig. 1.

Imperceptibility The quality of the audio signal to be restored after adding watermark. The imperceptibility is quantified by signal-to-noise ratio (SNR) and by conducting subjective listening test.

Robustness It reflects the ability of correctly retrieving the watermark bits, with and without attack. The robustness is evaluated by computing the bit error rate and average information loss (AIL) considering various signal processing attacks.

Payload The embedding capacity of watermarking algorithm defines the payload. It represents the number of bits embedded per second in original audio signal.

The audio watermarking techniques have been classified into time domain and frequency domain. The time-domain techniques mainly utilize least significant bit (LSB) substitution and echo-hiding techniques. In LSB technique, the audio signal is sampled at 8 or 16 kHz and divided into frames, and the LSB of each frame is replaced by the watermark bit [4]. To increase the robustness and imperceptibility, various modifications in LSB technique have been proposed by changing the embedding positions. The time-domain watermarking techniques are found in [5–11]. Generally, time-domain watermarking techniques are simple and less complex but suffers with low robustness [12, 13].

*Correspondence: aniruddhakanhe@nitpy.ac.in
[1]Department of Electronics and Communication Engineering, National Institute of Technology Puducherry, Karaikal, India
Full list of author information is available at the end of the article

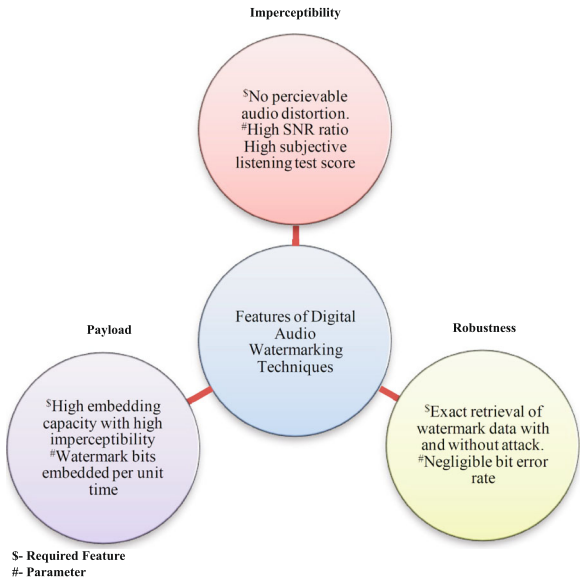

Fig. 1 Performance evaluation factors for audio watermarking techniques

In frequency domain audio watermarking techniques, various transforms such as discrete wavelet transform (DWT), fast Fourier transform (FFT), modified discrete cosine transform (MDCT), and cepstral coefficient transforms are used, and the watermark bits are embedded in the transform coefficients [3, 14–17]. Different transforms are cascaded to increase the robustness and imperceptibility of the audio watermarking techniques.

In [18], a DWT-DCT based audio watermarking scheme is proposed where the watermark bits are embedded in the low-frequency components by adaptive quantization technique. The multiresolution characteristics of DWT and energy compaction characteristics of DCT are explored for increasing the robustness of the watermarking scheme. In [19], lifting wavelet transform (LWT) and QR decomposition-based audio watermarking scheme is presented. The watermark is embedded using quantization of transform coefficients to increase the robustness of the watermarking scheme. A DCT-based data-hiding technique is presented in [20], where the DCT coefficients are modified using a scaling factor, to embed the data bits. In [21], speech bandwidth extension-based audio watermarking method is presented where time-domain and frequency-domain parameters of high-frequency speech signal are embedded in the narrow-band speech. In [22], watermark bits are embedded adaptively by performing SVD transform on short-time Fourier transform (STFT) coefficients of original audio. In [23], to increase the payload and robustness of the audio watermarking technique, the watermark bits are embedded in the off-diagonal elements of singular matrix obtained by decomposing the DWT coefficients using SVD.

In [24], SVD-based blind audio watermarking technique is proposed. The audio signal is divided into non-overlapping frames followed by the SVD operation. The binary watermark image is inserted in the singular matrix. The watermarking scheme is tested against various signal processing attacks. In [3], the audio signal is divided into short frames after computing the FFT. The watermark bits are embedded in the audio signal by modifying the FFT samples using Fibonacci numbers. The watermarking technique provides the robustness against common signal processing attacks with high payload capacity. A DWT and rational dither modulation-based audio watermarking technique is presented in [25]. The watermark bits are embedded in the 5th-level approximation subband to increase the robustness. The scheme provides robustness against various signal processing attacks with lower payload capacity.

In [26], SVD- and QIM-based adaptive watermarking scheme is presented for stereo audio signals. The audio signal is transformed into frequency domain, and multichannel SVD operation is performed to obtain the singular values. The watermark is embedded in the singular values using QIM scheme. In [27], energy-balanced vector modulation scheme is proposed for embedding the watermark bits in the DWT coefficients. The spectral shaping filters are incorporated to reduce the error spectrum. In [28], a DWT and lower upper (LU) factorization-based audio watermarking technique is presented with high payload capacity. The audio signal is divided into small samples, and the genetic algorithm is used to find the sample for hiding. LU decomposition is used to hide the watermark bits in the 5th-level DWT low-frequency components. In [29], DWT-DCT-based audio watermarking technique is presented. The audio signal is decomposed using multilevel DWT where, 1st–9th detail subbands are used for embedding the watermark and 11th approximation subband is used for inserting synchronization data. The watermark is embedded in DCT coefficients of detailed subbands using rational dither modulation scheme. In [30], adaptive mean modulation (AMM)-based speech watermarking technique in DWT domain is proposed. The watermark bits and synchronization codes are embedded in 2nd-level approximation and detail subbands, respectively. QIM is used for embedding the watermark bits in DWT coefficients of voiced frames, by adaptively changing the quantization steps. In [31], a DWT-SVD-QIM-based audio watermarking technique is presented for stereo audio signals. The 2nd-level approximation DWT coefficients of original audio signals are decomposed by SVD transform, and watermark bits are embedded in singular matrix using QIM. The watermark image is encrypted using Arnold chaotic map algorithm before embedding in the original audio signal. In [32], a blind audio watermarking

algorithm is presented using DWT-DCT transform. The fourth-level detail coefficients of original audio signal are decomposed using DCT. The watermark is embedded by modifying the average amplitude of DCT coefficients. It has been found in literature that, frequency domain watermarking techniques can achieve high robustness and imperceptibility [12].

In this paper, we propose a robust audio watermarking technique using DCT and SVD decomposition. In the proposed technique, the audio signal is sampled into short frames and the frames are identified as voiced and unvoiced frames by computing short-time energy (STE) and zero-crossing count (ZCC). The frames having high STE and low ZCC are marked as voiced frames, and the DCT coefficients are obtained for such frames. These coefficients are arranged in matrix form, and SVD operation is performed on these matrices. The watermark bits are embedded into the non-diagonal elements of singular matrix obtained by SVD operation to achieve high robustness and payload.

To the best of our knowledge, this is the first image-in-audio watermarking technique based on DCT-SVD transform in low-frequency audio frames. The novelty of this paper comes from modifying and testing of an image watermarking-based DCT-SVD [33] approach for audio watermarking and providing a statistical frame work to quantify the loss of entropy from watermark image under various signal processing attacks. Embedding watermark bits in low-frequency voiced frames increases imperceptibility and robustness against signal processing attacks, compared to the fragile approach of embedding in all frames of the original audio signal. The experimental results show that the proposed method provides high embedding capacity of 6 kbps, and robustness against common signal processing attacks by limiting the BER to $< 0.3\%$ even for strong perturbations. We propose a new metric called average information loss (AIL) from watermark image due to signal processing attacks, based on statistical measures. The complete watermarking technique is modeled mathematically as a discrete memoryless channel (DMC), and Shannon's average information is computed to check the loss of information.

The rest of the paper is organized as follows: Section 2 gives introduction about extraction of STE and ZCC, to mark voiced and unvoiced frames followed by proposed embedding and extraction procedure. Experimental results for robustness, imperceptibility and payload are presented in Section 3. Section 4 includes the conclusions of the proposed work.

2 Proposed technique

In this work, the audio signal used for experimentation is speech signal. Generally, the speech signal has been classified into voiced and unvoiced parts. The voiced part of speech consists of high-energy and low-frequency component whereas the unvoiced part of speech contains low-energy and high-frequency component [34]. The voiced frames are selected for embedding watermark bits because the distortion created in high-energy frames are less audible compared to low-energy frames. Further, the modification in DCT coefficients of high-energy frames introduces minimal distortion compared to low-energy frames and hence provides better scope of embedding the watermark.

The embedding and extraction of proposed watermarking technique is presented in this section. The proposed audio watermarking technique consists of three main procedural blocks: frames marking/separation block, embedding block, and extraction block.

2.1 Frames marking/separation

In the proposed method the audio frame is marked as voiced frame when the STE is high and ZCC is low. In contrast, when the STE is low and ZCC is high, the frame is marked as unvoiced frame [35].

The flow chart to separate voiced and unvoiced frames using STE and ZCC is shown in Fig. 2. The speech signal is sampled at 8 kHz and divided into non-overlapping frames having L samples per frame. The STE and ZCC are computed using Eqs. 2 and 3, respectively.

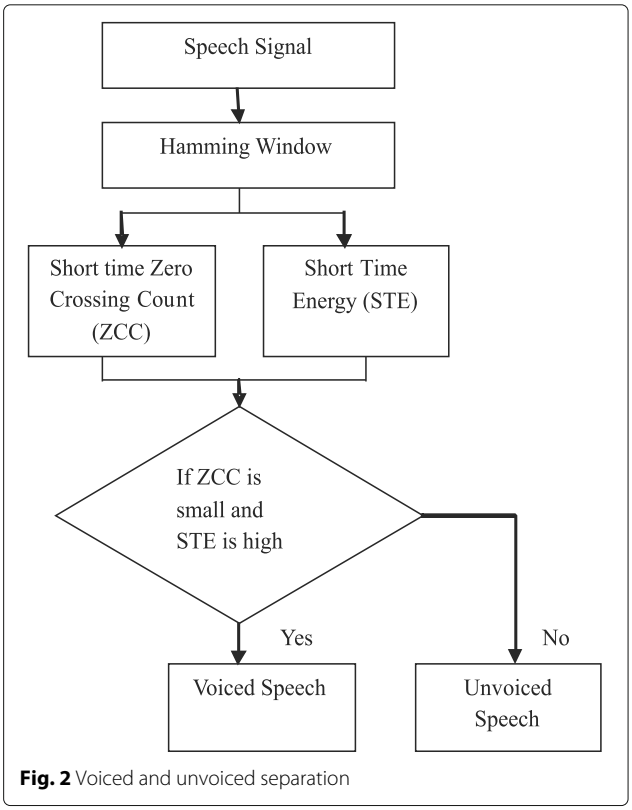

Fig. 2 Voiced and unvoiced separation

2.1.1 STE

The short-time energy of speech signal reflects the amplitude variation in it. The sampled speech signal is divided into number of frames by multiplying with Hamming window function. Individually, in each frame, the square of every sample is added together to get STE. The Hamming window function $w(n)$ used in the proposed technique for dividing the speech audio signal into frames is given in Eq. 1:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\frac{2\pi n}{L-1} & \text{for } 0 \le n \le L-1 \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

and if $s(n)$ represents a signal, then the short-time energy E_m is given by Eq. 2

$$E_m = \sum \{s(n)w(m-n)\}^2 \quad (2)$$

2.1.2 ZCC

The ZCC counts the number of times the signal crosses the zero of the time axis in each frame, which basically reflects frequency. As voiced speech contains low-frequency components, the ZCC for voiced signal will be considerably lower than its unvoiced counterpart. The DC offset is removed before calculating the ZCC. Consider a frame of speech signal $s[n]$ containing L samples, then the ZCC is given by Eq. 3:

$$\text{ZCC} = \sum_{n=0}^{L-1} 0.5|\text{sign}(s[n]) - \text{sign}(s[n-1])| \quad (3)$$

where

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

The threshold values of STE and ZCC for marking voiced and unvoiced frames are made available at receiving end to correctly decode the watermark image. After separating the voiced and unvoiced frames from the original audio signal, DCT and SVD of voiced frames are computed for embedding the watermark bits as explained in the next subsection.

2.2 Embedding procedure

The watermark bits are embedded in voiced parts of original audio signal by computing DCT followed by SVD as depicted in Fig. 3.

2.2.1 DCT

The energy compaction characteristics of DCT makes it suitable for proposed audio watermarking algorithm [18]. Consider $x(n)$ is the input voiced frame, then the 1-D DCT of length N can be given by:

$$X(k) = w(k)\sum_{n=0}^{N-1} x(n)\cos\frac{(2n+1)k\pi}{2N}, k = 0, 1, \ldots, N-1. \quad (4)$$

where

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

The DCT operation is performed over each voiced frame. The original audio signal is sampled at 8 kHz and frame size is taken as 10 ms. Each frame is further divided into five subframes having 16 samples in each subframe. The DCT operation is performed on subframe to generate 16 DCT coefficients. The DCT coefficients of each subframe are arranged in a 4×4 matrix designated as $[A]$. The next step is to perform SVD operation on $[A]$.

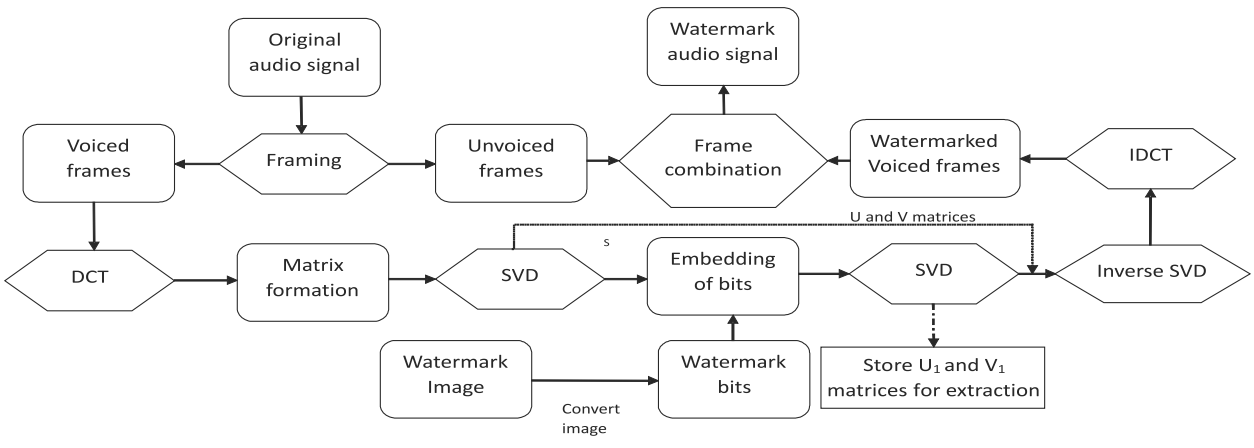

Fig. 3 Embedding procedure: The block diagram show the major steps involved in embedding watermark image in original audio signal using proposed technique

$$[A] = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 & a_7 \\ a_8 & a_9 & a_{10} & a_{11} \\ a_{12} & a_{13} & a_{14} & a_{15} \end{bmatrix}$$

2.2.2 SVD

SVD is a powerful mathematical tool that decomposes a given matrix [A] into combination of three matrices Eq. 5:

$$[A] = [U][S][V]^T \tag{5}$$

where $[U]$ and $[V]^T$ are orthogonal matrices and $[S]$ is a singular value matrix. The S matrix of SVD decomposition is invariant to common signal processing operation. This property of SVD decomposition makes it more suitable for the proposed audio watermarking algorithm. The watermark bits are embedded in non-diagonal elements of $[S]$ matrix.

$$[S] = \begin{bmatrix} s_0 & 0 & 0 & 0 \\ 0 & s_5 & 0 & 0 \\ 0 & 0 & s_{10} & 0 \\ 0 & 0 & 0 & s_{15} \end{bmatrix}.$$

2.2.3 Watermark embedding

In the proposed watermarking algorithm, the binary images are used as watermark as shown in Fig. 4.

The watermark image of size $m \times n$ is converted in binary sequence of $K(= m \times n)$ bits as shown below

$$B = b_1 b_2 b_3 b_4 \ldots b_K$$

The non-diagonal elements of $[S]$ matrix are replaced by the binary watermark bits using a scaling factor α as mentioned in Eq. 6.

$$[S_n] = [S] + \alpha \times [W] \tag{6}$$

where

$$[W] = \begin{bmatrix} 0 & b_1 & b_2 & b_3 \\ b_4 & 0 & b_5 & b_6 \\ b_7 & b_8 & 0 & b_9 \\ b_{10} & b_{11} & b_{12} & 0 \end{bmatrix}.$$

Fig. 4 Watermark images used for embedding

$[W]$ is the watermark bit matrix with diagonal elements as zero and $[S_n]$ is the modified singular matrix.

$$[S_n] = \begin{bmatrix} s_0 & b_1' & b_2' & b_3' \\ b_4' & s_5 & b_5' & b_6' \\ b_7' & b_8' & s_{10} & b_9' \\ b_{10}' & b_{11}' & b_{12}' & s_{15} \end{bmatrix}.$$

The b_i' in the $[S_n]$ matrix denotes the scaled watermark bits. The SVD operation is performed on $[S_n]$ to get the orthogonal matrices $[U_1]$ and $[V_1]$ which will be utilized for the extraction of watermark. The SVD operation is performed on $[S_n]$ followed by inverse DCT transform using Eq. 7 to generate watermarked voiced frames.

$$x(n) = \sum_{k=0}^{N-1} w(k)X(k)\cos\left[\frac{(2n+1)k\pi}{2N}\right], n = 0, 1, 2, \ldots, N-1. \tag{7}$$

These steps are repeated for all the voiced frames of original signal as per Algorithm 1.

2.3 Extraction procedure

To extract the watermark image, the voiced and unvoiced frames are separated from watermarked audio followed by the DCT and SVD operation as shown in Fig. 5. The

Algorithm 1 Watermark embedding

Divide the audio signal into frames of 10ms duration compute ZCC and STE
↓ represents low value and ↑ represents high value
if ZCC is ↓ and STE is ↑ **then**
  frame = voiced frame
  divide into sub-frames $S_v(n)$
  else
  frame = unvoiced frame $S_{uv}(n)$
end if
for each voiced frame **do**
  $S_v(k) = \mathbf{DCT}[S_v(n)]$
  $[A] \Leftarrow [S_v(k)]$ ▷ Transform the coefficient into 4×4 matrix
  Perform **SVD** operation
  $\mathbf{SVD}[A] = [U][S][V]^T$
  Generate $[W]$ with diagonal element zero
  $[S_n] = [S] + \alpha \times [W]$
  $[U_1][S_1][V_1]^T \Leftarrow \mathbf{SVD}[S_n]$
  Store the values of $[U_1]$ and $[V_1]^T$ for extraction
  Perform Inverse **SVD** operation
  $[A']_{4\times4} \Rightarrow [U][S_n][V]^T$
  Perform Inverse **DCT** operation
  $S_{vn}'(n) \Leftarrow \mathbf{IDCT}[S_v'(k) \leftarrow [A']]$
end for
  combine $S_{vn}'(n)$ **and** $S_{uv}(n)$

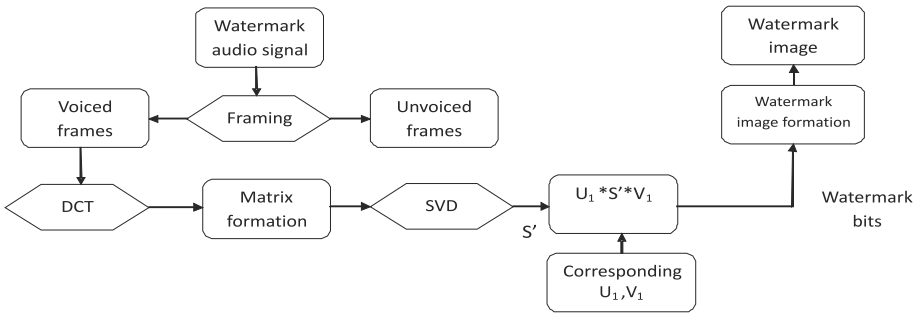

Fig. 5 Extraction procedure: The block diagram shows the major steps involved in extraction of watermark image from watermarked audio signal using proposed DCT-SVD technique

watermarked audio signal is divided into non-overlapping frames of L samples per frames and marked as voiced and unvoiced frames, as mentioned in Subsection 2.1. Each voiced frame is divided into subframes with 16 samples in each subframe. It is to be noted that the length of frame L and the threshold for marking voiced and unvoiced is made available to the receiver end as key. The DCT operation is performed on each watermarked voiced subframe $S_v^s(n)$ to obtain $S_v^s(k)$. The obtained DCT coefficients are arranged in 4×4 matrix designated as $[B]$.

Using the pre-stored matrices U_1 and V_1, SVD operation is performed on $[B]$ to obtain $[S_{vn}^s]$ as mentioned in Algorithm 2 to get $[D_w]$. The watermark bits from $[D_w]$ are extracted by examining the non-diagonal elements using a decision-making scheme as shown below:

$$b_i = \begin{cases} 1 & \text{for} \quad D_{w(ij)} \geq \epsilon \\ 0 & \text{for} \quad D_{w(ij)} < \epsilon \end{cases} \tag{8}$$

where

$$\epsilon = \text{avg}[D_{w(ij)}] \quad \forall \quad i \neq j$$

These steps are repeated for all the voiced frames of watermarked audio signal to extract watermark bits.

3 Results

The proposed audio watermarking technique is tested on NOI ZEUS speech database [36–38] and MIR-1K music database (https://sites.google.com/site/unvoicedsoundseparation/mir-1k). The NOIZEUS database contains 30 sentences from IEEE sentence database, recorded in a sound proof booth using Tucker Davis Technologies recording system. The database contains 15 male and 15 female speakers and include all phonemes in the American English language. MIR-1K database contains 1000 song audio from 110 karaoke pop songs performed by both male and female amateurs. The singing voice from the music signal has been is separated to utilize specific voicing characteristics of speech. The separation of singing voice before embedding the watermark is performed using principal component analysis [39].

Algorithm 2 Watermark extraction

1: Divide the watermark audio signal into frames of $10ms$ duration compute ZCC and STE
2: $\downarrow$ represents low value and $\uparrow$ represents high value
3: **if** ZCC is $\downarrow$ and STE is $\uparrow$ **then**
4: frame = voiced frame
5: divide into sub-frames $S_v^s(n)$
6: **else**
7: frame = unvoiced frame $S_{uv}^s(n)$
8: **end if**
9: **for** each voiced frame **do**
10: $S_v^s(k) = \textbf{DCT}[S_v^s(n)]$
11: $[B] \Leftarrow [S_v^s(k)]$ ▷ Transform the coefficient into 4×4 matrix
12: Perform **SVD** operation
13: $[U][S_{vn}^s][V]^T \Leftarrow \textbf{SVD}[B]$
14: use the $[U_1]$ and $[V_1]$ to get $[D_w]$
15: $[D_w] \Leftarrow [U_1] \times [S_{vn}^s] \times [V_1]^T$
16: compute the average of non-diagonal elements of $[D_w]$
17: $\epsilon = avg[D_{w(ij)}] \quad \forall \quad i \neq j$
18: Compare the non-diagonal elements
19: **if** non diagonal element $< \epsilon$ **then**
20: watermark bit = 0
21: **else**
22: watermark bit = 1
23: **end if**
24: **end for**
25: **combine all the extracted bits to get the watermark image**

The imperceptibility and robustness of the proposed audio watermarking technique is evaluated using SNR, subjective listening test, and BER. SNR of the proposed work is listed in Table 1 and compared with the DWT [3], DWT-SVD [23], and DWT-FFT [40] techniques. It is evident from Table 1 that SNR of the proposed technique is higher than the SNR obtained by [3, 23, 40]. The

Table 1 Comparison SNR values with other techniques

Methodology	Proposed method	Proposed method	DWT [3]	DWT-SVD [23]	DWT-FFT [40]
SNR (dB)	87.41	85.32	61	37.50	34.45
Database	Speech	Music	Speech	Speech	Speech

reason of achieving the significant improvement in SNR values is because of embedding the watermark data in DCT coefficients of voiced frames only.

Blind subjective listening test is performed on the watermarked signal to estimate the audio quality. The test is performed with five individuals of age group 17–21 years in a closed room with good quality earphones. Each individual is provided with randomly selected ten original and watermarked audio signals and were asked to grade the quality on a scale of five. The grade starts with 1 for perceptible distortion and goes up to 5 for high imperceptibility. The average of grades provided by the listeners with maximum payload are presented in Table 2. The comparison with another DCT-based technique [20] indicates that the proposed technique maintains the imperceptibility.

The values of SNR and subjective listening score indicates that the proposed audio watermarking technique is highly imperceptible. The spectrogram of original audio signal and watermarked audio signal are shown in Fig. 6 to support the results of high imperceptibility of proposed audio watermarking algorithm.

The robustness of the proposed audio watermarking technique is verified by the computing BER. The watermarked audio is processed through re-sampling, requantization, AWGN, MP3 compression, amplitude scaling, low-pass filtering, and high-pass filtering operations to corrupt the watermark image. In re-sampling attack, the watermarked audio signal is sampled with a frequency different from the original sampling frequency and resampled back to the original frequency. Similarly, in requantization attack, the watermarked audio is quantized to different level to destroy the watermark. In AWGN attack, white Gaussian noise is added to the watermarked audio signal and the error between retrieved watermark and original watermark image is calculated. Similarly, in MP3 compression attack, the watermarked audio is compressed by MP3 standard and de-compressed to destroy the watermark embedded in the audio. In low-pass filtering (LPF) attack, the watermarked signal is passed through a filter of cutoff frequency of 4 kHz. In high-pass filtering (HPF) attack, the watermarked signal is passed through a filter of cutoff frequency of 50 Hz. In amplitude scaling attack (ASA), the amplitude of watermarked signal is scaled by 0.7. The BER values of the proposed watermarking technique obtained in various attack cases are listed in Fig. 7 and Table 3, with maximum payload. The BER values confirm that the proposed method is robust against the common signal processing attacks.

Table 2 Subjective listening test results

Audio type	DCT based [20]	Proposed technique
	Grades	Grades ("mean [± standard deviation"])
Speech	4.87	4.61 [± 0.21]
Music	–	4.57 [± 0.16]

The BER comparison between proposed audio watermarking technique with other frequency domain watermarking techniques for re-sampling attack, requantization attack, AWGN, and MP3 compression is shown in Table 3.

Compared to the proposed DCT-SVD method, SVD-QIM [26] shows $<$ 100% accuracy in recovering the watermark bits in the absence of any attacks. Such a drawback is common in short-length watermark and correlation-based detection scheme [25]. In our implementation, the watermark is recovered using pre-stored orthogonal matrices. The proposed DCT-SVD technique shows the second lowest BER for re-sampling attack because embedding is done in the low-frequency components and DCT possesses a property to retain the shape of low-frequency components [29]. The BER in case of AWGN attack is 0 because the extraction of watermark bits mainly depends on the change in DCT coefficients, and since the change in DCT is comparatively low, the watermark bits can be estimated accurately [41]. In the two cases of MP3 attacks, the BER is significantly low. The reason for such a low BER is that the intensity of noises added due to attack is considerably low compared to watermark noise. The proposed DCT-SVD technique shows robustness against LPF attack. The reason for such an observation can be attributed to the fact that the embedding is done in low-frequency frames only. For the attack of HPF with a cutoff frequency 50 Hz, the BER is highest because the watermark is primarily embedded in low-frequency spectrum. The HPF neglects the low-frequency components and corrupts the watermark. The robustness against AS attack is achieved because the extraction of watermark is dependent on orthonormal matrices, and the orthonormal matrices are invariant to amplitude scaling attack.

The performance of proposed technique is also evaluated by computing the average information loss during the watermarking. In this paper, the average information is computed by modeling the overall system as a discrete memoryless channel whose input is watermarked audio

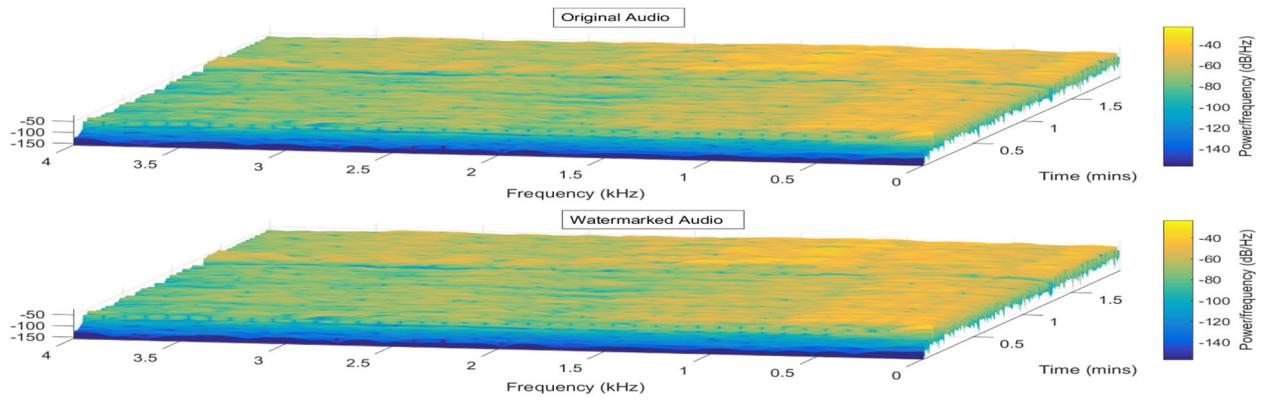

Fig. 6 Spectrogram plot of original and watermarked audio

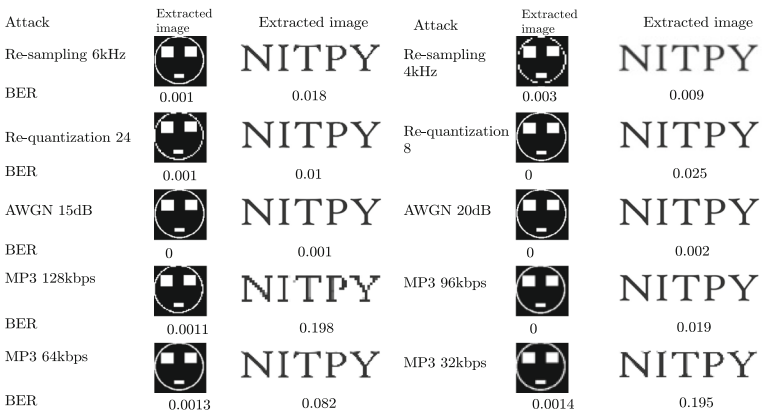

Fig. 7 Extracted watermark images with BER for speech audio signal for re-sampling, re-quantization, AWGN, and MP3 compression attacks

Table 3 Average BER results of the proposed DCT-SVD-based watermarking technique compared with other watermarking techniques

Attacks	Proposed method	DWT-SVD [23]	SVD-QIM [24]	DWT-RDM [25]	SVD-QIM [26]	DWT-VM [27]	DWT-LU [28]	DWT-DCT [29]	DWT-AMM [30]
None	0	0	0	0	0.10	0	0	0	0
Re-sampling	0.03	0	0	0	4.88	0	0	0.90	0
Re-quantization	0	0	0	0	–	0	–	0	0.03
AWGB 20 dB	0	0	0	0	–	0	–	0.03	9.32
AWGN 30 dB	0	0	0	0	10.25	0.04	0	0.04	1.38
MP3 Compression 64 kbps	0.0013	0.0820	0.561	0	24.56	0.05	0.02	0.01	–
MP3 Compression 32 kbps	0.0014	0.2901	2.204	1.05	–	2.75	0.02	0.01	–
LPF (4 kHz)	0	0.1893	0.512	0.08	0.31	0.98	0.03	0.04	0.03
HPF (50 Hz)	40.62	0.358	–	–	–	–	–	–	8.93
AS	0.31	–	–	0	0.38	0	–	0	0

and output is the retrieved watermark image as depicted in Fig. 8. The AIL metric introduced in this paper can be further used for the empirical computation of lower and upper bounds of robustness-related entropy based on the theoretical model proposed in [29, 42].

We consider the formulation of proposed watermark technique as a generic model of communication problem [42]. M denotes the watermark image embedded in audio data O_a^N transmitted to decoder through channel. $A(o_w|n_w)$ is the channel statistical characteristics subjected to various signal processing attacks provided with an input data O^N. K^N is the common side information shared by both encoder and decoder, and $\hat{M}$ is the retrieved watermark image. Referring to Fig. 8, suppose the watermark data is associated with a random variable M, which takes the symbol from a finite source alphabet.

$$\Psi = \{m_1, m_2, \ldots m_i\}$$

with probabilities

$$Q(M = m_i) = q_i \qquad i = 1, 2, \ldots L$$

And the original watermarked audio source O_a is a random variable which takes the symbol from a finite source alphabet

$$\Omega = \{o_{a1}, o_{a2}, \ldots, o_{aN}\}$$

with probabilities

$$P(O_a = o_{aj}) = p_j \qquad j = 1, 2, \ldots N$$

The retrieved watermark at the decoder is associated with a random variable $\hat{M}$, which takes the symbol from another alphabet

$$\hat{\Psi} = \{\hat{m_1}, \hat{m_2}, \ldots \hat{m_k}\}$$

with probabilities

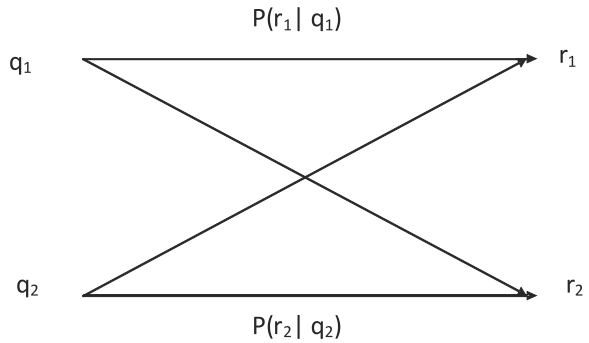

Fig. 9 Approximation of proposed watermarking technique as a binary channel. q_1, q_2 and r_1, r_2 are the input and output probabilities of binary watermark image, respectively

$$Q(M = \hat{m_k}) = r_k \qquad k = 1, 2, \ldots L$$

$$\sum_{i=1}^{L} q_i = 1; \qquad \sum_{j=1}^{N} p_j = 1; \qquad \sum_{k=1}^{L} r_k = 1;$$

The above communication model can be simplified to binary channel as shown in Fig. 9. The watermark image considered is binary image; therefore, the source alphabet of original watermark image and retrieved watermark image contains only two symbols $\{0, 1\}$ [43].

Then, the average information associated with random variable M and $\hat{M}$ can be expressed as:

$$H(M) = \sum_{i=1}^{i=L} q_i \log \frac{1}{q_i} \text{ binits/symbol} \tag{9}$$

$$H(\hat{M}) = \sum_{k=1}^{k=L} r_k \log \frac{1}{r_k} \text{ binits/symbol} \tag{10}$$

where q_i and r_k denote the probability of occurrence of m_i and $\hat{m}_i$, respectively. In the proposed technique, binary watermark image is used; hence, for the Eqs. 9 and 10, the value of $L = 2$. Then, the average information loss can be expressed as

$$AIL = |H(M) - H(\hat{M})|$$

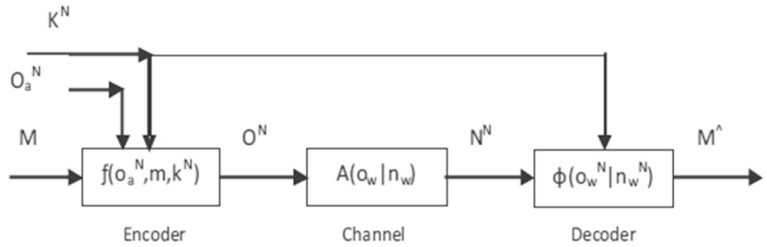

Fig. 8 Modeling of proposed watermarking technique as discrete memoryless channel

Table 4 BER and AIL results

	Re-sampling		Re-quantization		AWGN		MP3 compression (kbps)			
	6 kHz	4 kHz	24	8	15 dB	20 dB	128	96	64	32
BER	0.01	0.03	0	0	0	0	0	0.001	0.0013	0.0014
AIL	0.12	0.13	0	0	0	0	0	0.01	0.01	0.01

The AIL values for the proposed technique subjected to re-sampling, re-quantization, and AWGN attack are given in Table 4. It is evident form the AIL results that the proposed technique is robust to various signal processing attacks since the average information loss is negligible.

The payload represents the number of bits embedded within 1 s of the original audio. In the proposed technique, 12 bits of secret data were embedded in every 16 samples, where the sampling rate is 8 kHz, and frame size is fixed to 10 ms and each frame have 80 samples. In each voiced frame, the number of watermark bits embedded were 60.

Hence, the payload is 60 bits per 10 ms per voiced frames. The payload of proposed method and its comparison with different wavelet domain-based audio watermarking methods are given in Table 5.

It is evident that the proposed technique provides the high embedding capacity up to 6 kbps. The watermark bits are embedded only in the low-frequency high-energy voiced frames since the unvoiced frames are low-energy frames, and the poor representation of the of DCT coefficients of unvoiced frames will degrade the SNR and subjective listening quality of watermarked audio [44].

4 Conclusions

In this paper, we proposed a novel audio watermarking technique based on DCT and SVD transform. The proposed technique embeds the watermark bits adaptively in selected frames having low frequency and high energy. The watermark bits are embedded in DCT coefficients of selected frames by performing SVD operation. The watermark bits are embedded in non-diagonal elements of SVD matrix. Experiments are conducted to evaluate the performance of the proposed audio watermarking technique and compared with recent frequency-domain audio watermarking techniques.

The high-SNR values confirm that the proposed technique is highly imperceptible. The robustness of proposed audio watermarking technique is evaluated by computing BER and AIL for re-sampling, re-quantization, AWGN, and MP3 compression attacks with high data payload. The proposed watermarking scheme achieves comparable, if not better, results compared with other recently developed techniques for various attacks considered in this work.

Future research work may include the enhancement of proposed technique to withstand with random cropping attack, pitch shifting attack, and time-scale modification attack. The proposed technique can be made robust against these attacks by embedding synchronization codes with watermark bits.

Table 5 Payload results

Techniques	Payload (*bps*)	Database
SVD-QIM [24]	196	Music
DWT-SVD [23]	1.03k	Speech and music
Fibonacci-FFT [44]	3k	Speech and music
DWT-RDM [25]	344.53	Music
SVD-QIM [26]	187.5	Speech
DWT-VM [27]	818.26	Music
DWT-LU [28]	1.28k	Speech and music
DWT-DCT [29]	86.13	Music
DWT-AMM [30]	200	Music
DWT-SVD-QIM [31]	1.6k	Music
Proposed technique	6k	Speech and music

Abbreviations

AIL: Average information loss; AWGN: Additive white Gaussian noise; DCT: Discrete cosine transform; DMC: Discrete memoryless channel; DWT: Discrete wavelet transform; FFT: Fast Fourier transform; LSB: Least significant bit; LWT: Lifting wavelet transform; MDCT: Modified discrete cosine transform; SNR: Signal-to-noise ratio; STE: Short-time energy; STFT: Short-time Fourier transform; SVD: Singular value decomposition; ZCC: Zero-crossing count

Availability of data and materials

The data supporting the conclusions of this article are included within the article.

Authors' contributions

AK has conducted the research, analyzed the data, and authored the paper. GA has provided the guidance for the research and has revised the paper. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Electronics and Communication Engineering, National Institute of Technology Puducherry, Karaikal, India. [2]Department of Computer Science and Engineering, National Institute of Technology Puducherry, Karaikal, India.

Received: 16 February 2018 Accepted: 3 September 2018
Published online: 01 October 2018

References

1. M. Arnold, M. Schmucker, S. D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*. (Artech House, Inc., Norwood, 2003)
2. H. J. Kim, Y. H. Choi, J. Seok, J. Hong, Audio watermarking techniques. Intell. Watermarking Tech. **7**, 185 (2004)
3. M. Fallahpour, D. Megias, 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
4. W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding. IBM Syst. J. **35**(3.4), 313–336 (1996)
5. N. Cvejic, T. Seppanen, in *2002 IEEE Workshop on Multimedia Signal Processing*. Increasing the capacity of LSB-based audio steganography, (St. Thomas, 2002), pp. 336–338. https://doi.org/10.1109/MMSP.2002.1203314, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1203314&isnumber=27103
6. P. Bassia, I. Pitas, N. V, Robust audio watermarking in the time domain. IEEE Trans. Multimed. **3**(2), 232–241 (2001). https://doi.org/10.1109/6046.923822
7. W.-N. Lie, L.-C. Chang, Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. IEEE Trans. Multimed. **8**(1), 46–59 (2006)
8. D. Cai, K. Gopalan, in *IEEE International Conference on Electro/Information Technology*. Audio watermarking using bit modification of voiced or unvoiced segments, (Milwaukee, 2014), pp. 491–494. https://doi.org/10.1109/EIT.2014.6871813, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6871813&isnumber=6871745
9. Y. Erfani, S. Siahpoush, Robust audio watermarking using improved TS echo hiding. Digit. Signal Process. **19**(5), 809–814 (2009)
10. G. Hua, J. Goh, V. L. L. Thing, Cepstral analysis for the application of echo-based audio watermark detection. IEEE Trans. Inf. Forensics Secur. **10**(9), 1850–1861 (2015). https://doi.org/10.1109/TIFS.2015.2431997
11. A. Kanhe, G. Aghila, C. Y. S. Kiran, C. H. Ramesh, G. Jadav, M. G. Raj, in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Robust Audio steganography based on Advanced Encryption standards in temporal domain, (Kochi, 2015), pp. 1449–1453. https://doi.org/10.1109/ICACCI.2015.7275816, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7275816&isnumber=7275573
12. G. Hua, J. Huang, Y. Q. Shi, J. Goh, V. L. 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
13. 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
14. M. Fallahpour, D. Megias, High capacity audio watermarking using FFT amplitude interpolation. IEICE Electron. Express. **6**(14), 1057–1063 (2009)
15. R. K. Jha, B. Soni, K. Aizawa, Logo extraction from audio signals by utilization of internal noise. IETE J. Res. **59**(3), 270–279 (2013). https://doi.org/10.4103/03772063.2013.10876505
16. M. Fallahpour, D. Megias, Robust high-capacity audio watermarking based on FFT amplitude modification. IEICE Trans. Inf. Syst. **93**(1), 87–93 (2010)
17. S. V. Dhavale, R. S. Deodhar, D. Pradhan, L. M. Patnaik, State transition based embedding in cepstrum domain for audio copyright protection. IETE J. Res. **61**(1), 41–55 (2015). https://doi.org/10.1080/03772063.2014.987704
18. 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
19. J. Li, T. Wu, in *2015 International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*. Robust audio watermarking scheme via QIM of correlation coefficients using LWT and QR decomposition, (Chengdu, 2015), pp. 1–6. https://doi.org/10.1109/ICCSS.2015.7281138, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7281138&isnumber=7281133
20. A. Kanhe, G. Aghila, in *Proceedings of the International Conference on Informatics and Analytics, ICIA-16*. DCT Based Audio Steganography in Voiced and Un-voiced Frames (ACM, New York, 2016), pp. 47:1–47:4. https://doi.org/10.1145/2980258.2980360
21. 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
22. H. Ozer, Sankur, N. Memon, in *Proceedings of the 7th Workshop on Multimedia and Security. MM&Sec '05*. An SVD-based audio watermarking technique (ACM, New York, 2005), pp. 51–56. https://doi.org/10.1145/1073170.1073180
23. A.-H. Ali, 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
24. V. Bhat, I. Sengupta, A. Das, A new audio watermarking scheme based on singular value decomposition and quantization. Circ. Syst. Signal Process. **30**(5), 915–927 (2011)
25. H.-T. Hu, L.-Y. Hsu, A DWT-based rational dither modulation scheme for effective blind audio watermarking. Circ. Syst. Signal Process. **35**(2), 553–572 (2016). https://doi.org/10.1007/s00034-015-0074-9
26. 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). https://doi.org/10.1109/TMM.2017.2721642
27. 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
28. A. Kaur, M. K. Dutta, An optimized high payload audio watermarking algorithm based on LU-factorization. Multimedia Systems. **24**(3), 341–353 (2018). https://doi.org/10.1007/s00530-017-0545-x
29. H.-T. Hu, J.-R. Chang, Efficient and robust frame-synchronized blind audio watermarking by featuring multilevel DWT and DCT. Clust Comput. **20**(1), 805–816 (2017). https://doi.org/10.1007/s10586-017-0770-2
30. 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). https://doi.org/10.1186/s13636-017-0106-4
31. 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). https://doi.org/10.1007/s10772-017-9462-9
32. Q. Wu, M. Wu, A novel robust audio watermarking algorithm by modifying the average amplitude in transform domain. Appl. Sci. (2076-3417). **8**(5) (2018)
33. A. Sverdlov, S. Dexter, A. M. Eskicioglu, in *Signal Processing Conference, 2005 13th European*. Robust DCT-SVD domain image watermarking for copyright protection: embedding data in all frequencies (IEEE, 2005), pp. 1–4
34. L. Rabiner, B.-H. Juang, *Fundamentals of Speech Recognition*. (Prentice-Hall, Inc., Upper Saddle River NJ USA, 1993)
35. R. Bachu, S. Kopparthi, B. Adapa, B. D. Barkana, in *Electrical Engineering Department School of Engineering*. Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal (University of Bridgeport, 2005), pp. 1–7
36. Y. Hu, P. C. Loizou, Subjective comparison and evaluation of speech enhancement algorithms. Speech Comm. **49**(7), 588–601 (2007)
37. Y. Hu, P. C. Loizou, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Speech Audio Process. **16**(1), 229–238 (2008)
38. 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)
39. Huang P.S., S. D. Chen, P. Smaragdis, M. Hasegawa-Johnson, in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singing-voice separation from monaural recordings using robust principal component analysis, (2012), pp. 57–60
40. S. Rekik, D. Guerchi, S.-A. Selouani, H. Hamam, Speech steganography using wavelet and fourier transforms. EURASIP J. Audio, Speech, Music Process. **2012**, 20 (2012). https://doi.org/10.1186/1687-4722-2012-20
41. Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech, Signal Process. **33**(2), 443–445 (1985). https://doi.org/10.1109/TASSP.1985.1164550

42. P. Moulin, J. A. O'Sullivan, Information-theoretic analysis of information hiding. IEEE Trans. Inf. Theory. **49**(3), 563–593 (2003). https://doi.org/10.1109/TIT.2002.808134
43. D.-Y. Tsai, Y. Lee, E. Matsuyama, Information entropy measure for evaluation of image quality. J. Digit. Imaging. **21**(3), 338–347 (2008)
44. W. K. McDowell, W. B. Mikhael, A. P. Berg, in *2012 Proceedings of IEEE Southeastcon*. Efficiency of the KLT on voiced amp: unvoiced speech as a function of segment size, (2012), pp. 1–5. https://doi.org/10.1109/SECon.2012.6197063