Arabian Journal for Science and Engineering (2022) 47:10003–10024
https://doi.org/10.1007/s13369-021-06431-8

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Secured Speech Watermarking with DCT Compression and Chaotic Embedding Using DWT and SVD

Kasetty Praveen Kumar[1] · **Aniruddha Kanhe**[1]

Received: 31 March 2021 / Accepted: 22 November 2021 / Published online: 27 January 2022

Abstract
In this paper, a secured watermarking algorithm based on chaotic embedding of speech signal in discrete wavelet transform (DWT) domain of cover audio is proposed. The speech signal to be embedded is compressed using discrete cosine transform (DCT) by finding the suitable number of DCT coefficients such that the perceptual quality of decompressed signal is preserved. The chaotic map is used to select the cover audio frames randomly instead of performing sequential embedding. The cover audio is decomposed using DWT followed by singular value decomposition (SVD), and the DCT coefficients of the speech signal are embedded in the singular matrix of the cover audio. The proposed watermarking algorithm achieves good imperceptibility with an average SNR and ODG of 46 dB and −1.07, respectively. The proposed algorithm can resist to various signal processing attacks such as noise addition, low-pass filtering, requantization, resampling, amplitude scaling, and MP3 compression. Experimental results show that the secret speech is reconstructed with an average perceptual evaluation of speech quality (PESQ) score of 4.26 under no attack condition, and above 3.0 under various signal processing attacks. Further, the correlation between original and reconstructed secret speech signal is close to unity. In addition, the loss in the generality of the information of the reconstructed speech signal is tested and is found minimum even the watermarked audio is subjected to various signal processing attacks. The proposed algorithm is also tested for false positive test to ensure the security of watermarking algorithm.

Keywords Chaotic embedding · DCT · DWT · SVD

1 Introduction

The security of digital multimedia data is one of the major challenge in communicating sensitive information in various fields such as medical diagnosis and military. Various techniques have been proposed by the researchers based on cryptography, secret sharing, and information hiding to achieve the secured communication. In cryptography, a secret key is used to encode the plaintext into a meaningless and unreadable format known as ciphertext. The decoding process of the ciphertext is done with a valid key that is shared with authorized persons only. The techniques proposed in [1,2] uses the mechanism of cryptography for secured transmission of audio.

In secret sharing technique, the secret data to be communicated are divided into multiple parts known as shares. In a (k, n) secret sharing scheme, the secret message is divided into 'n' shares, and at least 'k' shares are required to regenerate the secret message [3,4]. Bharthi et al. [5] proposed a verifiable (n, n) secret audio sharing scheme where the audio is divided into stream of amplitudes and signs. These streams are further divided into shares and a key is embedded into them to avoid the reconstruction of original audio by unauthorized users.

There exists certain limitations with cryptography and secret sharing-based techniques. (1) Certainty of the existence of secret data cannot be avoided, (2) Retrieval of secret data is difficult when an intentional signal processing attack, such as noise addition, compression, and cropping, is performed.

The process of concealment of secret data in a cover medium (either audio, image or text) is known as informa-

✉ Kasetty Praveen Kumar
praveenk783@gmail.com

Aniruddha Kanhe
aniruddhakanhe@nitpy.ac.in

[1] Department of Electronics and Communication Engineering, National Institute of Technology Puducherry, Karaikal 609609, India

tion/data hiding which is commonly known as watermarking or steganogarphy. These data hiding techniques conceal the secret data, and exhibit good robustness to the signal processing attacks. So that, the secret message can be recovered with minimum error. In addition, it ensures that the perceptual distortion introduced due to embedding is minimum.

In this paper, a novel audio watermarking algorithm is proposed where both the cover medium and secret data are audio signals. The audio watermarking techniques can be classified into time domain and transform domain [6–8]. In time domain, the secret data can be embedded in various ways such as direct modification of amplitudes, substitution of least significant bits (LSB), and insertion of echo signals. The time-domain approaches are easier and faster for implementation, but these are not robust to the signal processing attacks [7].

In transform domain, the cover audio is transformed using different frequency transformation techniques such as fast Fourier transform (FFT), discrete cosine transform (DCT), discrete wavelet transform (DWT), and lifting wavelet transform (LWT). Then, the secret data are embedded in this transformed coefficients to achieve good imperceptibility and robustness to the attacks.

To the best of authors' knowledge, a few works have been reported in [9–17] for hiding secret speech in audio. Xu et al. [9] proposed a secure speech communication scheme in which the secret speech is compressed using a compressive sensing method and converted it into a binary stream before embedding. The cover audio is transformed using DCT followed by LWT, and binary bits are embedded in LWT coefficients using scalar costa scheme. At extraction side, it uses a pre-trained K-SVD-based dictionary to decompress the extracted speech signal. The experimental results show that the watermarked audio achieves segmental signal-to-noise ratio (SNR) of 32.34 dB with mean opinion score (MOS) of 3.616. Similarly, the reconstructed speech signal achieves segmental SNR of 13.06 dB and is enhanced to 14.50 dB after performing wavelet de-noising. In addition, the results demonstrated that the scheme is robust to additive white Gaussian noise (AWGN) of 20 dB and low-pass filtering (LPF) with normalized correlation coefficient (NCC) of 0.91 and 0.99, respectively.

Shahidi et al. [10] proposed an audio steganography scheme using integer LWT. The secret speech samples are converted into binary stream and then embedded in LSB positions of second level LWT coefficients of cover audio by a dynamic stego key. The proposed scheme achieves hiding capacity of 25% of the cover audio size and SNR of watermarked audio is about 45 dB. Moreover, it shows robustness against adaptive noise up to 60 dB AWGN with NCC equal to 0.96.

Ballesteros et al. [11] proposed a steganography model in which the secret speech and the cover speech are decomposed using DWT and then the wavelet coefficients of speech signal are sorted as per the order of cover audio's wavelet coefficients. The embedding of these sorted coefficients of speech signal is performed by modifying the first five LSB positions of wavelet coefficients of cover audio. The order in which the coefficients are sorted is used as a key at extraction side to recover the secret speech. Due to its large key size, the watermarking system achieves higher security.

Ali et al. [12] compressed the secret audio using fractal coding technique in which the best match of secret audio with the cover audio is computed and obtained the corresponding fractal parameters. These fractal parameters are then converted into binary and chaotically embedded in LSB positions of the cover audio. This method achieved good imperceptibility of extracted secret speech with an average NCC equal to 0.99. The same authors proposed a modified version of the above technique in [16], wherein the fractal parameters are embedded in LWT coefficients of cover audio using uniform coefficient modulation scheme. The proposed scheme extracts the secret speech with an average correlation of 0.99 under no attack condition. In addition, the robustness test results showed that this scheme can resist to AWGN attack of 30 dB, echo addition, and cropping attack. The computational time of these fractal coding-based techniques is higher due to its asymmetric property. The encoding process is time-consuming during the range-domain matching process, while the decoding is simple and faster.

Ballesteros et al. [13] proposed a technique, where the secret speech samples are scrambled such that it imitates a super-Gaussian noise signal with similar statistics to that of secret speech. The scrambled samples of the secret speech are embedded into LSB positions of cover audio based on an adaptive parameter. At extraction side, the secret speech is extracted by using a key that contains the original positions of scrambled speech signal. The proposed scheme achieves 99.7% transparency of watermarked audio with an average SNR of 23 dB, and the secret speech is extracted with a correlation of 0.974. The authors' have reported that the robustness evaluation of the scheme is considered as a future work.

Bharthi et al. [14] proposed an audio stenography technique in which the amplitudes and signs of the secret audio signal samples are separated and embedded in LSB positions of cover audio samples in a non-deterministic manner using a key. The experimental results showed that the secret speech is extracted with an average correlation of 0.97 under no attack condition.

Alsabhany et al. [15] proposed an adaptive multi-level phase coding audio steganography technique based on FFT and LSB. This method achieved perceptual transparency of 35 dB SNR of watermarked audio, and robustness to AWGN attack with an average bit error rate of 0.3.

The limitations of the techniques discussed above are as follows: The technique proposed in [9] requires a pre-trained

dictionary to decompress the extracted secret speech signal. Fractal encoding in [12,16] and speech scrambling in [13] are the iterative procedures that consume more time to process the secret speech signal before embedding. In addition, the watermarking algorithms proposed in [9,10,14,16] do not provide high robustness against the signal processing attacks.

The authors' of this paper proposed a steganography technique in [17] where the secret speech is divided into non-overlapping frames and SVD is applied on each frame. The singular values of secret speech frame are embedded in singular value matrix of DWT coefficients of cover audio. This scheme achieves good imperceptibility and robustness to the signal processing attacks, but it is a non-blind technique which requires partial information of secret speech at extraction stage.

To overcome the above said limitations, a watermarking algorithm for embedding the secret speech in the cover audio has been proposed that provides good imperceptibility and robustness. The major contributions of this paper are:

- DCT-based compression is used for compressing the speech signal, and the suitable number of DCT coefficients are obtained for embedding.
- Random numbers are generated for chaotic embedding of the secret data in cover audio to increase the security of watermarking.
- DWT–SVD-based watermarking is proposed in which the secret data are embedded in the singular value matrix of cover audio.
- Robustness of the proposed method is tested by computing NCC, perceptual evaluation of speech quality (PESQ) score, and the loss in generality of the information of reconstructed speech signal.

The organization of this paper is as follows: Preliminaries are discussed in Sect. 2, proposed method of embedding and extraction are explained in Sect. 3. Experimental results are presented in Sect. 4, followed by conclusion in Sect. 5.

2 Preliminaries

2.1 Compression of Speech Signal Using DCT

In this paper, DCT-based compression technique is used due to its energy compaction property. The information present in the higher-order DCT coefficients is negligible which can be eliminated without huge effect on speech signal [18]. Consider a sequence $x(n)$ of length M; the 1-D DCT of the sequence is computed as:

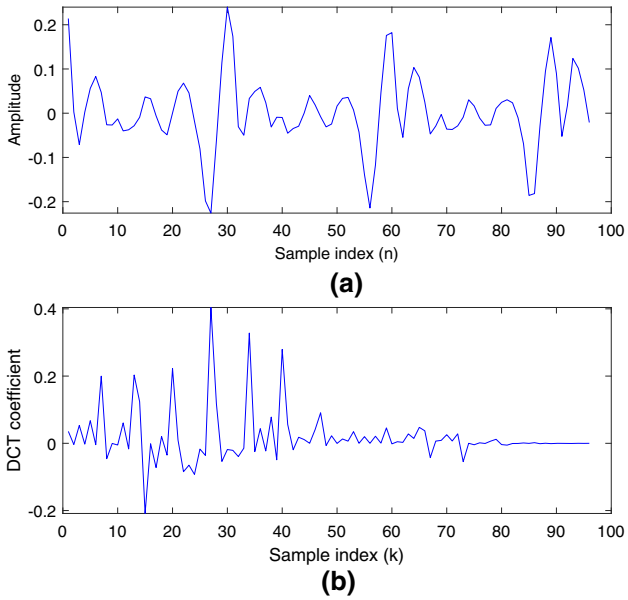

Fig. 1 **a** Original speech frame, **b** DCT coefficients of speech frame

$$X(k) = w(k) \sum_{n=0}^{M-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2M}\right), \tag{1}$$

where $k = 0, 1, 2, \ldots, M-1$ and

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

Similarly, the inverse DCT is computed as:

$$x(n) = \sum_{k=0}^{M-1} w(k) X(k) \cos\left(\frac{(2n+1)k\pi}{2M}\right), \tag{2}$$

where $n = 0, 1, 2, \ldots, M-1$. Figure 1 shows a frame of speech signal and it's DCT coefficients of length 96 samples. It can be observed that the energy of the signal is concentrated in the lower-order coefficients, whereas higher-order DCT coefficients are negligible.

Therefore, by neglecting the higher-order DCT coefficients, the compressed version of signal can be obtained by considering the suitable number of DCT coefficients such that the correlation coefficient (CC) between original and decompressed signal is closer to unity and a minimum PESQ score of 4.0. Algorithm 1 shows the procedure for finding the suitable number of DCT coefficients of secret speech for embedding in the cover audio.

Algorithm 1 Compression Algorithm

Input: Speech signal
Output: Compression factor
1: Divide the speech signal x into non-overlapping frames of length M samples.
2: Initialize compression factor, $CF = 0$
3: **do**
4: $CF = CF + 1/l$
5: **for** i=1 to N_s **do** ▷ N_s is no.of frames
6: Apply 1D-DCT on speech frame, $X_d = \text{DCT}(x(i))$ ▷ X_d is DCT of i^{th} frame of 'x'
7: Compute the number of DCT coefficients required, $count = CF * M$
8: Apply M point inverse DCT, $xcomp = \text{IDCT}(X_d(1 : count), M)$
9: **end for**
10: Merge all the frames to get decompressed speech
11: Calculate PESQ score of decompressed speech with reference to original speech signal
12: **while** PESQ< 4.0
13: return CF

In this algorithm, the compression factor (CF) is initialized to zero, and in each iteration, the CF is incremented by a step size of '$1/l$' ($l \leq M$) for finding the suitable number of DCT coefficients for compression. For each value of CF, PESQ score for decompressed speech signal is measured. If the PESQ $\geq$ 4.0, then the algorithm will return the corresponding CF value. As per the International Telecommunication Union—Telecommunication Standardization Sector (ITU-T) P.862 standard [19], PESQ score lies between -0.5 and 4.5, where, 1 indicates poor quality and 4.5 indicates excellent.

The small value of 'l' terminates the algorithm quickly, whereas large value of 'l' provides better resolution. So, a proper value of 'l' must be chosen based on the frame size M.

2.2 Chaotic Map for Random Embedding

Chaotic systems are type of dynamic systems whose random states and irregularities depends on its governing equations (difference equation which is termed as chaotic map) that are highly sensitive to the initial conditions. Due to the various properties exhibited by these systems, they can be used as a basis to generate random numbers in an efficient manner [1,20]. As the chaotic system is highly sensitive to initial conditions of chaotic map, a very small change in the initial condition will diverge the output. There are various 1-D chaotic maps existing in the literature such as logistic map, tent map, Bernoulli shift, and Chen map to generate the random numbers [1].

In this paper, a series of random numbers are generated using a logistic chaotic map to choose the cover audio frames for embedding. The usage of chaotic map makes data embedding in cover audio frames randomly instead of embedding

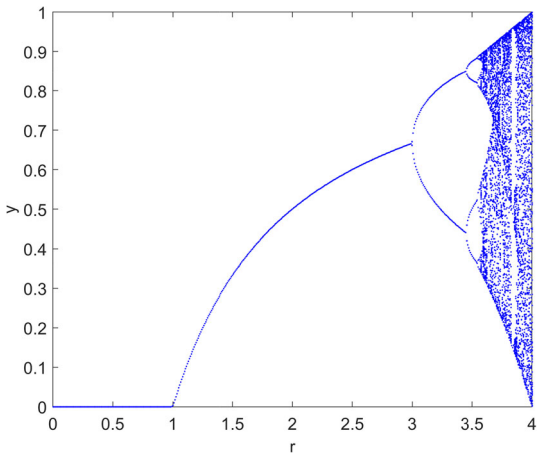

Fig. 2 Bifurcation diagram of logistic map

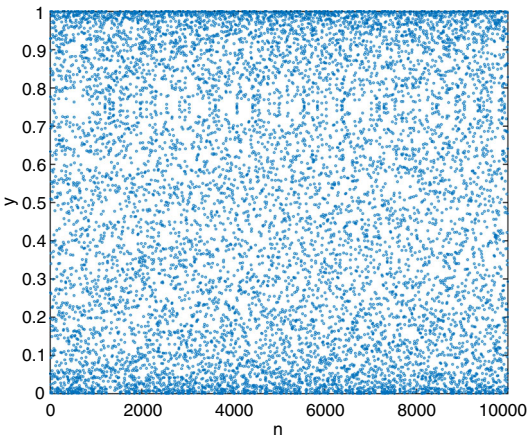

Fig. 3 Distribution of y for $y_0 = 0.052$ and $r = 4$

in sequential manner, which in turn increases the security of watermarking.

Logistic map is one of the most popular models for discrete nonlinear dynamic systems which is analogous to the logistic equation first proposed by Pierre Francois Verhulst [21]. A logistic map is given by,

$$y_{n+1} = r y_n (1 - y_n), \tag{3}$$

where $y_n \in (0, 1)$ is the ratio of the existing population to the maximum possible population after n years (i.e., y_n is the random value after n number of iterations), $y_0 \in (0, 1)$ is the initial population, and $r \in (0, 4]$ is the parameter that controls the rate of population.

The bifurcation diagram of logistic map for $y_0 = 0.052$ is shown in Fig. 2. It can be observed that the randomness in the value of y depends on the control parameter r. If $r \in (0, 1)$, then the value of y is close to 0 and is independent of initial population. If $r \in [1, 3)$, then the value of y approaches to $r/1-r$ and it is also independent of initial population, and for

$r \in (3.5, 4]$, the logistic map shows chaotic characteristics. Figure 3 shows the randomness of y distributed between zero and one for 10,000 iterations with initial value $y_0 = 0.052$ and $r = 4$. It is observed that y value always ranges between 0 and 1, but it is required to generate the random integers to choose the cover audio frames for embedding. It can be achieved by modifying the result of logistic map as follows: [22],

$$y_i' = y_i * 1000 \pmod{p} + 1 \quad i = 0, 1, 2 \ldots, n \tag{4}$$

where y_i' is the random integer, y_i is the random number generated from Eq. (3) with y_0 and r values, and p is the largest interval upto which the random integer is to be generated.

Finally, the set of y_i' are sorted in ascending order, and the repetition of integers is removed by using MATLAB functions $sort()$ and $unique()$, respectively, as shown in Eq. (5).

$$\text{loc} = unique(sort(y_i')), \tag{5}$$

where loc contains the random integers. These are used as frame indices of the cover audio in which the secret data is to be embedded.

2.3 Discrete Wavelet Transform

The DWT represents the signal's characteristics in both time and frequency domain using a scaling function $\phi(t)$ and a wavelet function $\psi(t)$ which are defined as follows [23]:

$$\phi(t) = \sqrt{2} \sum_n h_0(n) \phi(2t - n), \tag{6}$$

$$\psi(t) = \sqrt{2} \sum_n h_1(n) \phi(2t - n), \tag{7}$$

where $h_0(n)$ and $h_1(n)$ are the coefficients of low-pass and high-pass analysis filter, respectively.

If the signal $f(t) \in L^2(\mathbb{R})$, then it can be expressed as a series expansion of scaling and wavelet functions given by,

$$f(t) = \sum_{k=-\infty}^{\infty} c(k)\phi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j,k)\psi_{j,k}(t), \tag{8}$$

where $c(k)$ is the low-pass coefficients, and $d(j,k)$ is jth-level high-pass coefficients of the wavelet transform. The first term in the above expansion gives lower resolution or approximation of signal $f(t)$, and second term gives higher resolution of signal for each value of j.

Figure 4 shows the multi-level DWT decomposition of signal $f(t)$ employing the filter bank $h_0(n)$ and $h_1(n)$, where the output of low-pass filter is called as approximate coefficients and output of high-pass filter as detailed coefficients.

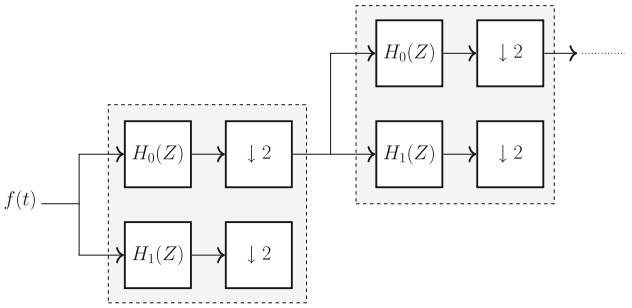

Fig. 4 Multi-level decomposition of DWT

These are obtained by calculating inner-product of $f(t)$ and $\phi_k(t)$, $f(t)$ and $\psi_{j,k}(t)$, respectively, as follows:

$$c(k) = \langle f(t), \phi_k(t) \rangle = \int f(t)\phi(t-k)\mathrm{d}t, \tag{9}$$

$$d(j,k) = \langle f(t), \psi_{j,k}(t) \rangle = \int f(t) 2^{j/2} \psi(2^j t - k)\mathrm{d}t, \tag{10}$$

where $\langle \cdot \rangle$ indicates inner product, $\phi_k(t) = \phi(t-k)$, and $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$.

Similarly, the inverse DWT is computed as follows:

$$c(j+1,k) = \sum_{n'=-\infty}^{\infty} c(j,n')\tilde{h}_0(k-2n') + \sum_{n'=-\infty}^{\infty} d(j,n')\tilde{h}_1(k-2n'), \tag{11}$$

where $\tilde{h}_0(n)$ and $\tilde{h}_1(n)$ are the coefficients of low-pass and high-pass synthesis filter, respectively. The corresponding filter coefficients are obtained as: $\tilde{h}_i(n) = h_i(L-1-n)$, where $i = 0, 1$ and 'L' is length of the filter.

In this paper, DWT is chosen for watermarking due its following advantages: (1) As L_2 norm is preserved in DWT domain, the amount of distortion introduced in the high-energy approximate coefficients due to embedding is minimum. Hence, good imperceptibility of watermarked audio can be achieved [24]. (2) Watermarking in the approximate coefficients of a signal are found to be robust against signal processing attacks [25].

2.4 Singular Value Decomposition

SVD decomposes a matrix $[A]$ of size $m \times m$ into combination of three matrices as shown below:

$$A = [U][S][V]^T = \sum_{i=1}^{m} \sigma_i u_i v_i^T, \tag{12}$$

where u_i and v_i are the column components of U and V matrices, respectively. These u_i and v_i components are obtained by

computing the eigenvectors of the matrices AA^T and A^TA, respectively. S is known as singular matrix that contains singular values ($\sigma_i > 0$) of A arranged diagonally such that $\sigma_1 > \sigma_2 > \cdots > \sigma_m$ and zeros along non-diagonal positions.

In this paper, SVD is chosen for watermarking the cover audio signal due to its unique properties. The embedding of data in singular matrix provides good imperceptibility and robustness to the watermarked signal because the singular matrix [S] represents the energy of a signal [26]. If any data are embedded into [S] matrix, then the variation in its singular values is minimum [27].

3 Proposed Method of Watermarking and Extraction

3.1 Embedding Algorithm

The process of embedding the secret speech signal in cover audio is shown in Fig. 5, and the steps are explained as follows:

Step 1: Pre-processing

The cover audio signal is divided into N_c number of non-overlapping frames with the frame size of N samples. The secret speech signal is divided into N_s number of non-overlapping frames with the frame size of M samples and then apply 1-D DCT on each frame.

The DCT coefficients of each secret audio frame is further divided into 'l' number of segments to obtain the DCT coefficients for embedding in cover audio frame as discussed in the Sect. 2.1. To select the cover audio frames chaotically for embedding the secret speech, random integers are generated using Eqs. (3), (4), and (5) with initial conditions of y_0 and r. These values serve as a key at the extraction side for logistic chaotic map.

Step 2: DWT

Initially, the cover audio frames are selected chaotically from the set of random integers generated in step 1, and then second-level DWT is performed on each cover audio frame. The approximate coefficients of the corresponding cover audio frame are further divided into 'l' segments.

Consider that c_k represents the second-level approximate coefficients of the kth cover audio frame of size 'N' samples.

$$c_k = c_k(0), c_k(1), \ldots, c_k(N/4). \tag{13}$$

Then, c_k^i represents the ith segment of the kth frame as follows:

$$c_k^i = c_k^i(0), c_k^i, \ldots, c_k^i(N/4l), \tag{14}$$

where $i = 1, 2, \ldots, l$. These $N/4l$ coefficients are arranged in a $m \times m$ matrix to perform SVD operation on it.

Step 3: SVD

SVD operation is performed on the coefficient matrix $[c_k^i]_{m\times m}$ and decompose it into $[U_k^i]$, $[S_k^i]$, and $[V_k^i]$ matrices to embed the data in the singular matrix.

$$[c_k^i]_{m\times m} = [U_k^i][S_k^i][V_k^i]^T. \tag{15}$$

Step 4: Embedding into cover audio frame

The DCT coefficients of ith segment of kth secret speech frame are arranged in a matrix [W] of size $m \times m$ such that the zeros are arranged diagonally, while the DCT coefficients in non-diagonal positions are as follows:

$$[W] = \begin{bmatrix} 0 & X_1 & X_2 & X_3 \\ X_4 & 0 & X_5 & X_6 \\ X_7 & X_8 & 0 & X_9 \\ X_{10} & X_{11} & X_{12} & 0 \end{bmatrix}_{4\times 4}, \tag{16}$$

where $X_1, X_2, \ldots, X_{12}$ are the DCT coefficients of the secret speech signal.

The watermark is embedded into the singular matrix $[S_k^i]$ such that $[\hat{S}_k^i] = [S_k^i] + \alpha[W]$, where α is a scaling parameter that ranges in between zero and one. The small value of 'α' leads to better imperceptibility, whereas large value leads to better robustness. Hence, it must be chosen to optimize the trade-off between imperceptibility and robustness. In this paper, 'α' is chosen empirically and is set to 0.08.

The SVD operation is again performed on $[S_k^i]$ matrix to obtain $[\hat{U}_k^i]$ and $[\hat{V}_k^i]$ matrices which are used during extraction phase.

Step 5: Generation of watermarked audio frame

Inverse SVD operation is performed on $[\hat{S}_k^i]$ by multiplying with corresponding $[U_k^i]$ and $[V_k^i]^T$ matrices to obtain the modified approximate coefficients $[\hat{c}_k^i]$ of the cover audio frame. These modified coefficients are arranged into a vector form, and then second-level inverse DWT is performed to get the watermarked audio frame.

This embedding process is repeated for all the chaotically selected frames and then merged to get the watermarked audio.

3.2 Extraction Algorithm

In extraction phase, the initial conditions of logistic map is given as a key for finding the embedding locations in the watermarked audio to extract the secret speech. The process of extracting the secret speech signal is shown in Fig. 6. The watermarked audio is divided into non-overlapping frames and are selected chaotically using the random integers generated from logistic map. The 2nd-level DWT is applied on

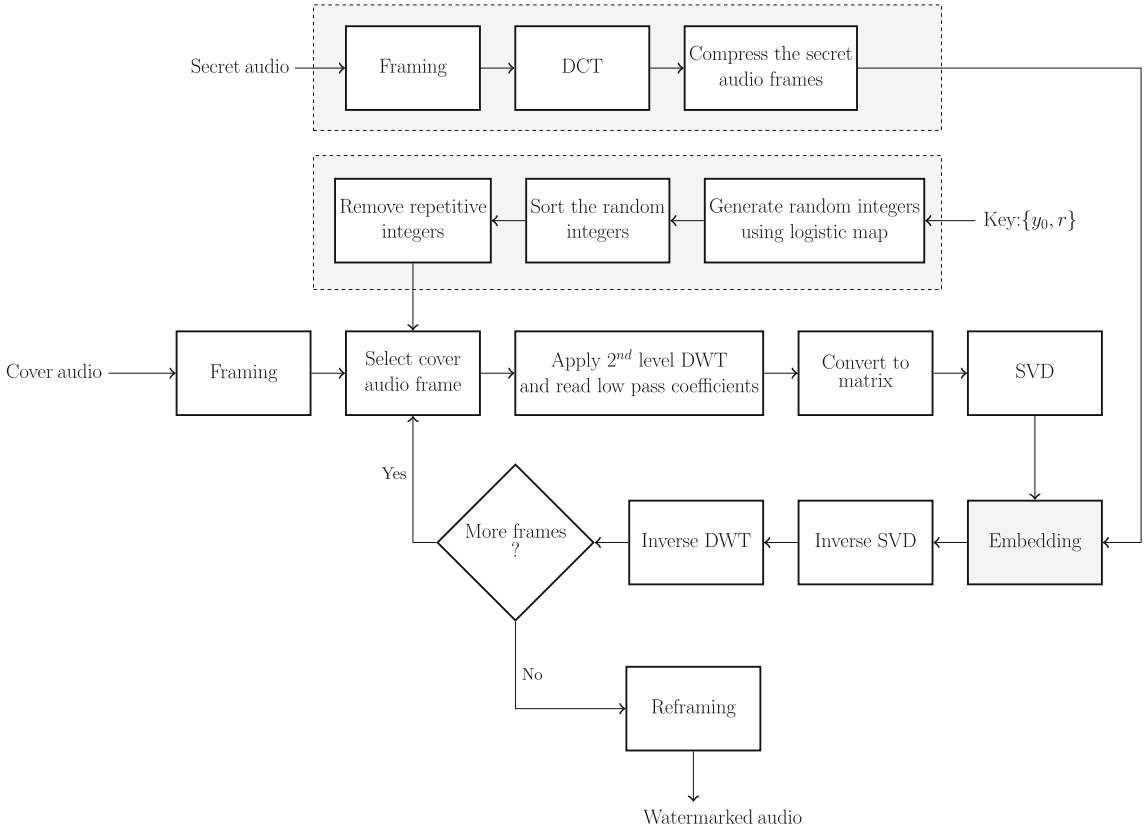

Fig. 5 Embedding algorithm

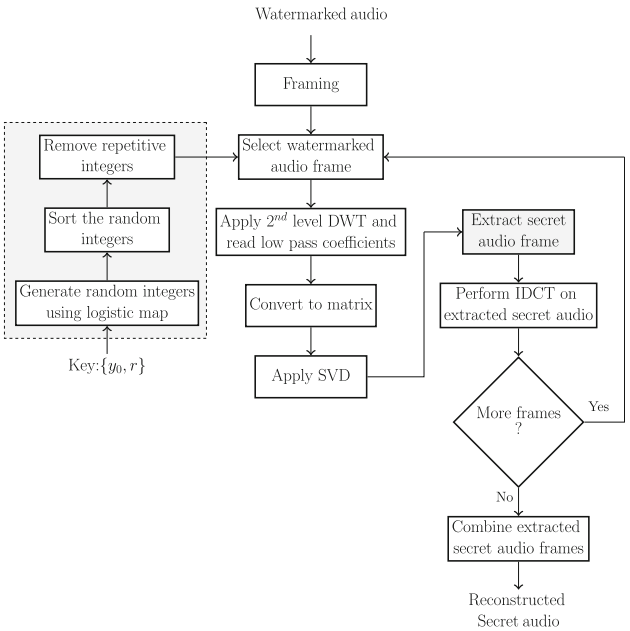

Fig. 6 Extraction algorithm

these selected frames, and the approximate coefficients are arranged in a matrix $[\dot{c}_k^i]$.

The SVD operation is performed on $[\dot{c}_k^i]$ and decompose into three matrices, namely $[\dot{U}_k^i]$, $[\dot{S}_k^i]$, and $[\dot{V}_k^i]$. The secret data are extracted by performing the inverse SVD on $[\dot{S}_k^i]$ using pre-stored matrices $[\hat{U}_k^i]$ and $[\hat{V}_k^i]$ which results a matrix $[D_k^i]$, as follows:

$$[D_k^i] = [\hat{U}_k^i][\dot{S}_k^i][\hat{V}_k^i]^T. \tag{17}$$

The watermark data are extracted from the matrix $[D_k^i]$ as follows:

$$D_w = \frac{D_k^i(p,q)}{\alpha} \quad \forall\, p \neq q, \tag{18}$$

where D_w is extracted DCT coefficients of secret speech. The above process for all l segments of the kth watermarked audio frame is repeated, and then M-point inverse DCT is performed using Eq. (2) to reconstruct the secret speech frame. Finally, all the extracted frames are merged to reconstruct the secret speech.

Table 1 Transparency of secret speech signal Sp13 for various CF values

CF	SNR_d (dB)	CC	PESQ	Error
1/8	0.7549	0.3994	1.23	6
2/8	3.2869	0.7286	2.28	6
3/8	7.7710	0.9126	2.93	4
4/8	17.1599	0.9903	3.51	4
5/8	19.7298	0.9947	3.96	0
6/8	25.7976	0.9987	4.29	0
7/8	46.2781	1.0000	4.48	0
8/8	310.4433	1.0000	4.50	0

4 Results and Discussions

The proposed watermarking algorithm is tested on unified speech and audio (USAC) database [28] which contains five music files sampled at 48 kHz with 16-bit quantization. The speech signal is chosen as secret audio from NOIZEUS database [29] that contains 30 speech (15 male and 15 female voices) signals sampled at 8 kHz with 16-bit quantization. In this section, compression test results on secret audio and performance of proposed audio watermarking technique are discussed.

4.1 Compression of Secret Audio

In this paper, the DCT-based compression on secret speech signal is performed by considering the frame size $M = 96$, and the compression factor CF is incremented by $1/8$ in each iteration. Therefore, 12 number of DCT coefficients are considered for compression, and the corresponding PESQ score is measured in each iteration.

In the proposed algorithm, the compressed secret speech is embedded in cover audio. Upon extraction, in addition to PESQ score, the decompressed speech signal was also tested to verify whether the reconstructed secret speech is compatible to existing speech to text conversion algorithms which may be employed in several applications such as speech recognition. This is performed by converting the decompressed speech signal into text by using the Microsoft Azure Speech application program interface (API) [30] built using Java script. The original text of the speech signals is taken from [29] and compared with the results of the speech-to-text API to count the number of erroneous words.

Table 1 shows the transparency test results of the secret speech signal 'Sp13' for various values of CF. It can be observed that for CF $= 6/8$, PESQ score is greater than 4.0 and the number of erroneous words are found to be zero. It is also observed that CC is closer to unity and signal-to-noise ratio (SNR) of decompressed speech signal is 25 dB which is an acceptable value as per International Federation of Phonographic Industry (IFPI) standard [31].

Table 2 ITU-R grade and ODG for audio quality evaluation

Grade	ODG	Quality of watermarked audio
5	0.0	Imperceptible
4	− 1.0	Perceptible, but not annoying
3	− 2.0	Slightly annoying
2	− 3.0	Annoying
1	− 4.0	Very annoying

In the similar manner, the compression is performed on all the speech signals of NOIZEUS database and it is found that compression factor, CF $= 6/8$, gives the optimized result. Therefore, 72 number of DCT coefficients from each speech frame are taken and then embedded into cover audio.

4.2 Performance of Proposed Audio Watermarking Algorithm

The performance of watermarking algorithm is evaluated by conducting 150 number of tests for both imperceptibility and robustness by embedding each secret speech into each of the cover audio. In the proposed watermarking algorithm, the frame size of the cover audio is chosen as 512 samples. The initial conditions for logistic map were chosen as $y_0 = 0.052$ and $r = 3.95$ to generate random numbers, and the random integers are generated using Eq. (4) with 'p' equals to the number of cover audio frames.

4.2.1 Imperceptibility

The imperceptibility of watermarked audio is quantified by measuring SNR as per Eq. (19),

$$\mathrm{SNR}_w\ (\mathrm{dB}) = 10 \log \left[\frac{\sum |s(n)|^2}{\sum |s(n) - s'(n)|^2} \right], \tag{19}$$

where SNR_w is SNR of the watermarked audio. $s(n)$, and $s'(n)$ are cover and watermarked audio signals, respectively. In addition to SNR, objective difference grade (ODG) score is measured by using perceptual evaluation of audio quality (PEAQ) measurement technique [32] specified by International Telecommunication Union—Radio—communication Sector (ITU-R) BS.1387 standard [33]. This ODG score quantifies the perceptual quality of watermarked audio as shown in Table 2.

The SNR and ODG values of the watermarked audio signals with speech signal 'Sp01' as secret audio is shown in Table 3. It is observed that the watermarked audio achieves an average SNR of 47 dB and an ODG score of −0.8. In similar manner, each speech signal from NOIZEUS database are

Table 3 SNR and ODG values with Sp01 as secret audio

Cover audio	SNR_w (dB)	ODG
Chorus	48.37	− 0.12
Classical	42.41	− 1.85
Jazz	47.89	− 0.78
Pop1	46.99	− 1.05
Pop2	53.22	− 0.50

embedded in all the five cover audio signals and measured the SNR_w and ODG score as mentioned. Table 4 shows the average SNR_w and ODG scores of all five watermarked audio signals corresponding to each secret speech signal. The average SNR of all 150 watermarked audios is 46 dB ±0.7242 and ODG score is -1.07 ± 0.0070. These results show that the distortion introduced due to embedding the secret speech in cover audio is negligible, and the perceptual quality of watermarked audio is preserved.

The comparison of imperceptibility results of watermarked audio is shown in Table 5. It is observed that the proposed algorithm achieves better SNR than the techniques present in [9,10,13]. The fractal encoding-based watermarking techniques in [12,16] achieve significantly better SNR than the proposed algorithm.

According to IFPI standard [31], the SNR of watermarked audio should be greater than 20 dB to achieve good imperceptibility. The proposed approach achieves an average SNR of 46 dB that meets the said criterion.

The imperceptibility of watermarked audio is also quantified by measuring the ODG score. From Table 2, it is observed that the ODG score lying in the range of −1 to 0 indicates that watermarked audio is imperceptible. From the imperceptibility test results mentioned in Table 4, it is found that the average ODG score of the watermarked audio is 1.07. From these experimental results, it is observed that the proposed approach maintains the minimum criteria required for achieving better imperceptibility of watermarked audio.

4.2.2 Robustness

Robustness test measures the ability of watermarked audio to reconstruct the secret data when it is subjected to various signal processing attacks. Various signal processing attacks that are considered in this paper are mentioned below:

1. Amplitude scaling (AS): The amplitude of watermarked audio is scaled by a factor of 0.75.
2. Additive white Gaussian noise (AWGN): A white Gaussian noise of SNR 20 dB is added to the watermarked audio.
3. Low-pass filtering (LPF): A second-order low-pass filter with cut-off frequency of 24 kHz is applied to the watermarked audio.
4. Requantization (RQ): The 16-bit watermarked audio is quantized to 8 bits/sample and back to 16 bits/sample.
5. MP3 compression: Watermarked audio signal is compressed to MPEG-I layer III (MP3) format at the rate of 128 kbps and again converted back to .wav format.
6. Resampling (RS): Watermarked audio is downsampled and then upsampled to original sampling frequency.

To evaluate the robustness of watermarking algorithm, NCC, SNR and PESQ score are measured in this paper. The purpose of NCC is to find the similarity between the original and reconstructed secret speech signals which is computed as per Eq. (20).

$$\mathrm{NCC} = \frac{\sum S_{\mathrm{o}}(i) S_{\mathrm{r}}(i)}{\sqrt{\sum S_{\mathrm{o}}^2(i)} \sqrt{\sum S_{\mathrm{r}}^2(i)}}, \tag{20}$$

where $S_{\mathrm{o}}(i)$ and $S_{\mathrm{r}}(i)$ represents original and reconstructed secret speech signals, respectively. Similarly, SNR and PESQ score are measured that gives the perceptual quality of reconstructed secret speech signal. In addition to these parameters, the number of erroneous words present in the text extracted from reconstructed secret speech is also determined.

The robustness test results of reconstructed speech signals for various attacks are given in Table 6, where SNR_{rs} is the SNR of the reconstructed secret speech, EW is the number of erroneous words present in the extracted text from secret speech. Considering the case for secret speech signal 'Sp01,' it is observed that the proposed algorithm is able to reconstruct the secret speech with high correlation of 0.98 under various attacks. Even though SNR_{rs} is lesser for some of the attacks, it is seen that there is no loss in the generality of the information. Figure 7 shows the PESQ scores of the reconstructed speech signal 'Sp01' from various watermarked audios against the signal processing attacks. It is found that the perceptual quality of speech signal lies in the acceptable range, and the average PESQ score of 3.81 is achieved.

In similar manner, a total of 150 number of robustness tests are conducted by reconstructing each speech signal from each cover audio under various signal processing attacks. Table 6 shows the robustness test results of all the thirty speech signals. It is observed that the number of erroneous words present in the extracted text from the reconstructed speech signal are zeros for most of the cases.

The robustness test results for LPF attack are shown in Figs. 8, 9, and 10 which show the PESQ score, SNR_{rs}, and NCC values of reconstructed speech signals, respectively. It is observed that the average PESQ score of the reconstructed speech is greater than 3.0, and the interquartile range (IQR)

Table 4 Imperceptibility test results (average of five watermarked audio signals)

Speech signal (.wav)	Sp01	Sp02	Sp03	Sp04	Sp05	Sp06	Sp07	Sp08	Sp09	Sp10	Sp11	Sp12	Sp13	Sp14	Sp15
SNR_w (dB)	47.78	46.12	47.54	46.51	46.78	45.39	47.06	45.95	45.50	45.69	44.64	45.73	45.32	44.89	45.29
ODG	− 0.86	− 1.16	− 0.97	− 1.09	− 1.00	− 1.08	− 1.01	− 1.14	− 1.03	− 1.17	− 1.18	− 1.15	− 1.12	− 1.09	− 1.24
Speech signal (.wav)	Sp16	Sp17	Sp18	Sp19	Sp20	Sp21	Sp22	Sp23	Sp24	Sp25	Sp26	Sp27	Sp28	Sp29	Sp30
SNR_w (dB)	46.61	46.95	47.49	45.06	45.46	45.01	47.35	46.69	45.73	45.84	46.50	46.86	46.30	45.95	46.03
ODG	− 0.96	− 1.01	− 1.04	− 1.07	− 1.00	− 1.13	− 1.10	− 0.99	− 1.17	− 1.01	− 1.17	− 0.98	− 1.11	− 1.08	− 1.07

Table 5 Comparison of imperceptibility results with relevant watermarking techniques

Method	Processing of secret speech	Embedding domain	SNR_w (dB)
[9]	Compressive sensing	DCT-LWT	32.34
[10]	Binary stream of samples	DWT	45.00
[12]	Fractal coding	Time domain	70.40
[13]	Scrambling the samples	Time domain	23.61
[16]	Fractal coding	LWT	50.40
Proposed	DCT-based compression	DWT–SVD	46.13

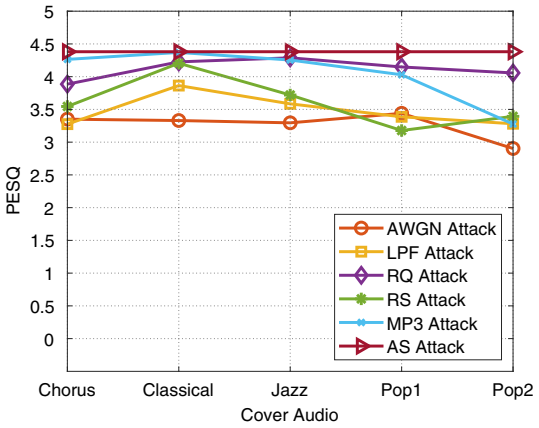

Fig. 7 PESQ scores of reconstructed speech (Sp01) against signal processing attacks

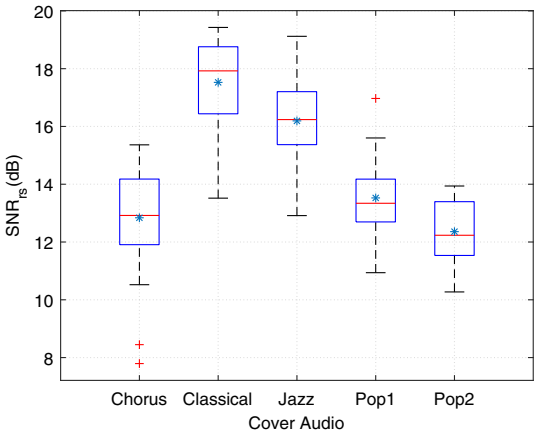

Fig. 9 SNR of reconstructed speech in terms of cover audio signal under LPF attack

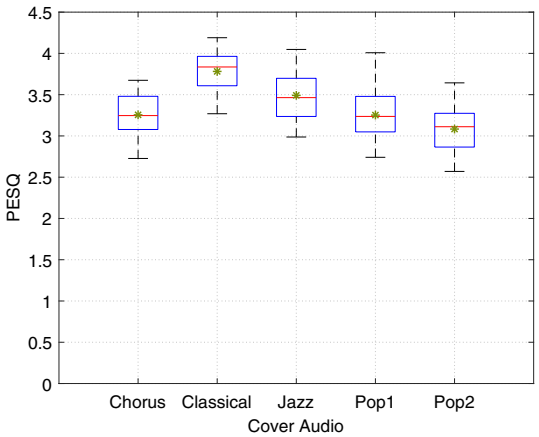

Fig. 8 PESQ scores of reconstructed speech in terms of cover audio signals under LPF attack

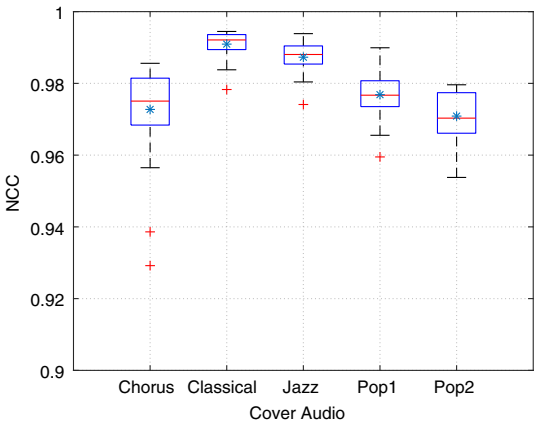

Fig. 10 Correlation between original and reconstructed speech in terms of cover audio signals under LPF attack

of PESQ scores is same for each box. Similarly, from Fig. 10, it is observed that the speech signal is reconstructed with the average correlation of 0.97 from each watermarked audio. These results show that the perceptual quality and correlation of the speech signal is high irrespective of the selection of cover audio signal which indicates that the proposed watermarking technique is able to resist the LPF attack. The reason is that the watermarking is performed in 2nd-level approximate coefficients of cover audio signal.

Table 7 shows the average values of SNR_{rs}, NCC, and PESQ of reconstructed speech signal for all 150 robustness tests on watermarked audio. It is observed that the proposed algorithm is able to extract the speech signal with high correlation close to unity and PESQ score of 3.78. This shows that the perceptual quality secret speech is retained under various signal processing attacks. Table 8 shows the comparison of SNR and correlation values between original and reconstructed secret speech with the relevant watermarking

Table 6 Robustness test results of the reconstructed secret speech signals

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp01																				
No attack	34.260	1.000	**4.381**	0	34.260	1.000	**4.381**	0	34.260	1.000	**4.381**	0	34.260	1.000	**4.381**	0	34.260	1.000	**4.381**	0
AWGN	9.277	0.953	3.351	0	12.840	0.979	3.330	0	10.737	0.964	3.295	0	12.661	0.976	**3.440**	0	8.597	0.948	2.903	0
LPF	12.298	0.972	3.275	0	19.036	0.994	**3.864**	0	17.568	0.991	3.585	0	13.013	0.975	3.387	0	13.613	0.979	3.281	0
RQ	19.101	0.994	3.886	0	23.845	0.998	4.226	0	27.014	0.999	**4.288**	0	22.308	0.997	4.148	0	22.507	0.997	4.057	0
RS	6.122	0.985	3.545	0	6.004	0.999	**4.205**	0	6.086	0.995	3.720	0	6.054	0.959	3.178	0	6.002	0.985	3.394	0
MP3	22.426	0.998	4.264	0	24.042	1.000	**4.370**	0	23.495	0.999	4.254	0	21.558	0.997	4.029	0	15.558	0.986	3.277	0
AS	12.017	1.000	**4.381**	0	12.017	1.000	**4.381**	0	12.017	1.000	**4.381**	0	12.017	1.000	**4.381**	0	12.017	1.000	**4.381**	0
For speech signal Sp02																				
No attack	19.736	0.995	**4.250**	0	19.736	0.995	**4.250**	0	19.736	0.995	**4.250**	0	19.736	0.995	**4.250**	0	19.736	0.995	**4.250**	0
AWGN	10.492	0.962	3.427	0	12.802	0.976	3.231	0	10.376	0.960	3.153	1	12.686	0.976	**3.596**	0	8.292	0.941	2.949	1
LPF	14.001	0.980	3.495	1	16.734	0.990	**3.964**	0	15.775	0.987	3.662	0	14.724	0.983	3.596	0	11.987	0.969	3.202	0
RQ	17.539	0.991	3.886	0	18.512	0.993	4.095	0	19.196	0.994	**4.212**	0	17.620	0.991	4.061	0	17.829	0.992	3.885	0
RS	5.921	0.987	3.747	0	5.902	0.993	**4.110**	0	5.889	0.989	3.753	0	5.920	0.970	3.478	0	5.956	0.979	3.408	0
MP3	18.350	0.994	4.187	0	18.618	0.994	**4.235**	0	18.319	0.994	4.193	0	18.253	0.994	4.153	0	16.156	0.988	3.515	0
AS	11.399	0.995	**4.250**	0	11.399	0.995	**4.250**	0	11.399	0.995	**4.250**	0	11.399	0.995	**4.250**	0	11.399	0.995	**4.250**	0
For speech signal Sp03																				
No attack	28.746	0.999	**4.289**	0	28.746	0.999	**4.289**	0	28.746	0.999	**4.289**	0	28.746	0.999	**4.289**	0	28.746	0.999	**4.289**	0
AWGN	11.197	0.965	**3.568**	0	13.981	0.982	3.429	0	10.601	0.961	3.395	0	12.909	0.978	3.605	2	7.426	0.935	3.060	0
LPF	15.364	0.986	3.660	0	18.866	0.994	**4.107**	0	17.861	0.992	3.761	0	14.294	0.982	3.625	0	13.701	0.978	3.549	0
RQ	23.839	0.998	4.110	0	23.436	0.998	4.209	0	26.067	0.999	**4.288**	0	22.834	0.997	4.144	0	18.399	0.993	3.882	0
RS	5.978	0.991	3.850	0	5.992	0.998	**4.174**	0	6.004	0.995	3.876	0	5.990	0.963	3.423	0	6.113	0.988	3.766	0
MP3	9.498	0.998	4.242	0	9.338	0.999	**4.278**	0	9.526	0.999	4.242	0	9.233	0.991	3.638	0	9.059	0.985	3.659	0
AS	11.955	0.999	**4.289**	0	11.955	0.999	**4.289**	0	11.955	0.999	**4.289**	0	11.955	0.999	**4.289**	0	11.955	0.999	4.289	0
For speech signal Sp04																				
No attack	22.584	0.997	**4.261**	0	22.584	0.997	**4.261**	0	22.584	0.997	**4.261**	0	22.584	0.997	**4.261**	0	22.584	0.997	**4.261**	0
AWGN	13.068	0.977	3.514	1	14.002	0.981	3.446	0	11.165	0.966	3.193	0	12.344	0.974	**3.563**	0	10.289	0.962	2.912	1
LPF	12.873	0.975	3.480	0	17.833	0.992	**3.983**	0	15.748	0.987	3.633	0	13.921	0.980	3.549	1	11.992	0.969	3.239	1
RQ	21.131	0.996	**4.209**	0	21.178	0.996	4.127	0	21.870	0.997	4.188	0	18.994	0.994	4.097	0	18.557	0.993	3.869	0
RS	5.986	0.985	3.729	0	5.975	0.996	**4.141**	0	6.017	0.991	3.866	0	6.123	0.971	3.380	0	6.072	0.977	3.519	0
MP3	20.221	0.996	4.228	0	20.616	0.997	**4.268**	0	18.496	0.994	4.201	0	19.609	0.996	4.176	0	15.459	0.986	3.447	0
AS	11.696	0.997	**4.261**	0	11.696	0.997	**4.261**	0	11.696	0.997	**4.261**	0	11.696	0.997	**4.261**	0	11.696	0.997	**4.261**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For Speech signal Sp05																				
No attack	24.943	0.998	**4.284**	0	24.943	0.998	**4.284**	0	24.943	0.998	**4.284**	0	24.943	0.998	**4.284**	0	24.943	0.998	**4.284**	0
AWGN	11.520	0.968	3.203	0	13.622	0.980	3.407	0	12.081	0.972	3.339	0	12.352	0.975	**3.525**	0	10.763	0.946	2.975	0
LPF	13.289	0.977	3.155	0	18.533	0.993	**3.883**	0	18.106	0.992	3.698	0	15.602	0.986	3.543	0	13.582	0.979	3.139	0
RQ	20.023	0.995	3.836	0	22.164	0.997	4.158	0	23.636	0.998	**4.276**	0	20.022	0.995	4.019	0	20.351	0.995	4.123	0
RS	6.043	0.987	3.421	0	5.988	0.997	**4.222**	0	5.991	0.995	4.002	0	6.004	0.970	3.389	0	6.000	0.985	3.265	0
MP3	21.677	0.998	4.226	0	21.667	0.998	**4.280**	0	21.671	0.998	4.276	0	21.293	0.998	4.161	0	15.962	0.988	3.337	0
AS	11.837	0.998	**4.283**	0	11.837	0.998	**4.283**	0	11.837	0.998	**4.283**	0	11.837	0.998	**4.283**	0	11.837	0.998	**4.283**	0
For Speech signal Sp06																				
No attack	21.119	0.996	**4.266**	0	21.119	0.996	**4.266**	0	21.119	0.996	**4.266**	0	21.119	0.996	**4.266**	0	21.119	0.996	**4.266**	0
AWGN	12.363	0.973	3.502	0	14.622	0.984	3.553	0	12.615	0.975	3.384	0	12.957	0.978	**3.566**	0	11.250	0.967	3.093	0
LPF	11.787	0.968	3.674	0	17.338	0.991	**4.048**	0	17.338	0.991	**4.048**	0	16.969	0.990	4.008	0	12.384	0.971	3.405	0
RQ	19.051	0.994	4.122	0	20.095	0.995	4.234	0	20.654	0.996	**4.248**	1	19.105	0.994	4.072	0	19.254	0.994	4.142	0
RS	5.927	0.984	3.928	0	5.936	0.995	**4.192**	0	5.935	0.994	4.121	0	6.044	0.950	3.376	0	5.951	0.985	3.673	0
MP3	19.021	0.995	4.251	0	19.560	0.996	4.260	0	19.465	0.996	**4.263**	0	18.287	0.994	4.136	0	17.740	0.993	3.684	0
AS	11.565	0.996	**4.266**	0	13.327	0.977	3.432	0	11.565	0.996	**4.266**	0	11.565	0.996	**4.266**	0	11.565	0.996	**4.266**	0
For speech signal Sp07																				
No attack	27.599	0.999	**4.344**	0	27.599	0.999	**4.344**	0	27.599	0.999	**4.344**	0	27.599	0.999	**4.344**	0	27.599	0.999	**4.344**	0
AWGN	12.132	0.973	3.503	0	12.896	0.977	3.259	0	12.024	0.972	3.245	0	12.547	0.976	**3.516**	0	7.354	0.933	3.019	0
LPF	12.933	0.976	3.229	0	18.311	0.993	**4.065**	0	17.090	0.990	3.874	0	14.062	0.980	3.502	0	13.611	0.979	3.390	0
RQ	23.217	0.998	3.918	0	22.112	0.997	4.240	0	25.191	0.998	**4.253**	0	21.449	0.996	4.064	0	19.848	0.995	4.140	0
RS	6.033	0.985	3.449	0	6.010	0.998	**4.239**	0	6.001	0.991	3.932	0	6.105	0.946	3.297	0	6.115	0.988	3.436	0
MP3	22.034	0.998	4.259	0	22.767	0.999	**4.332**	0	22.505	0.998	4.308	0	21.241	0.997	4.088	0	17.442	0.991	3.786	0
AS	11.929	0.999	**4.344**	0	11.928	0.999	4.343	0	11.930	0.999	4.343	0	11.929	0.999	**4.344**	0	11.929	0.999	4.343	0
For speech signal Sp08																				
No attack	21.725	0.997	**4.160**	0	21.725	0.997	**4.160**	0	21.725	0.997	**4.160**	0	21.725	0.997	**4.160**	0	21.725	0.997	**4.160**	0
AWGN	10.405	0.961	3.377	1	13.335	0.978	**3.390**	0	9.894	0.958	3.136	0	10.731	0.964	3.345	0	9.077	0.938	3.056	0
LPF	14.179	0.981	3.361	1	17.805	0.992	3.920	1	16.519	0.989	**3.808**	0	14.330	0.981	3.362	0	12.195	0.972	3.292	0
RQ	18.652	0.993	3.914	1	20.263	0.995	**4.107**	0	20.796	0.996	4.161	1	18.547	0.993	3.949	1	18.847	0.994	3.971	1
RS	6.000	0.989	3.632	0	5.966	0.995	**4.036**	0	5.941	0.990	3.827	0	5.983	0.974	3.342	0	6.110	0.977	3.485	0
MP3	19.721	0.996	4.110	0	19.882	0.996	**4.144**	0	19.813	0.996	4.116	0	19.386	0.995	4.046	0	17.441	0.991	4.026	0
AS	11.624	0.997	4.159	0	11.625	0.997	**4.160**	0	11.624	0.997	**4.160**	0	11.624	0.997	**4.160**	0	11.624	0.997	**4.160**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp09																				
No attack	17.649	0.991	**4.289**	0	17.649	0.991	**4.289**	0	17.649	0.991	**4.289**	0	17.649	0.991	**4.289**	0	17.649	0.991	**4.289**	0
AWGN	11.534	0.967	**3.275**	0	12.592	0.974	3.127	1	11.050	0.965	3.113	0	11.725	0.969	3.016	0	9.078	0.932	2.901	1
LPF	13.315	0.977	3.343	0	15.471	0.986	**3.800**	0	14.843	0.983	3.316	0	12.259	0.970	3.120	1	11.510	0.966	2.978	1
RQ	16.491	0.989	3.840	0	16.733	0.989	3.997	0	17.355	0.991	**4.223**	0	16.567	0.989	3.988	0	16.266	0.988	3.567	0
RS	5.767	0.984	3.652	0	5.789	0.990	**4.096**	0	5.811	0.988	3.609	0	5.811	0.958	3.141	1	5.969	0.974	2.990	0
MP3	16.764	0.991	4.224	0	16.766	0.991	4.228	0	16.758	0.991	**4.243**	0	16.326	0.989	4.012	0	14.643	0.983	3.306	0
AS	11.045	0.991	**4.289**	0	11.045	0.991	**4.289**	0	11.045	0.991	**4.289**	0	11.045	0.991	**4.289**	0	11.045	0.991	**4.289**	0
For speech signal Sp10																				
No attack	25.589	0.999	**4.279**	0	25.589	0.999	**4.279**	0	25.589	0.999	**4.279**	0	25.589	0.999	**4.279**	0	25.589	0.999	**4.279**	0
AWGN	12.970	0.977	**3.550**	0	13.848	0.981	3.264	0	11.631	0.970	3.263	0	12.444	0.975	3.448	0	7.167	0.932	2.681	0
LPF	14.506	0.983	3.536	0	18.679	0.993	**3.992**	0	17.314	0.991	3.625	0	13.580	0.978	3.292	0	13.396	0.977	3.303	0
RQ	22.218	0.997	4.172	0	22.610	0.997	4.188	0	23.674	0.998	**4.250**	0	20.540	0.996	4.013	0	20.012	0.995	4.100	0
RS	6.071	0.991	3.769	0	5.995	0.997	**4.170**	0	5.994	0.993	3.748	0	5.982	0.964	3.189	0	6.091	0.984	3.652	0
MP3	21.831	0.998	4.249	0	22.377	0.998	4.279	0	22.008	0.998	**4.284**	0	20.706	0.997	4.157	0	16.851	0.990	3.373	0
AS	11.865	0.999	**4.279**	0	11.865	0.999	**4.279**	0	11.865	0.999	**4.279**	0	11.865	0.999	**4.279**	0	11.865	0.999	**4.279**	0
For speech signal Sp11																				
No attack	24.999	0.998	**4.228**	0	24.999	0.998	**4.228**	0	24.999	0.998	**4.228**	0	24.999	0.998	**4.228**	0	24.999	0.998	**4.228**	0
AWGN	10.827	0.965	2.879	0	13.602	0.980	2.936	0	12.601	0.976	2.808	0	13.302	0.979	**2.989**	0	8.418	0.941	2.491	0
LPF	11.679	0.968	2.968	0	18.291	0.993	**3.669**	0	16.886	0.990	3.171	0	14.955	0.984	3.049	0	11.348	0.964	2.831	0
RQ	11.348	0.964	2.831	0	22.000	0.997	3.775	0	23.449	0.998	**4.061**	0	21.094	0.996	3.861	0	19.445	0.995	3.697	0
RS	6.009	0.981	3.192	0	6.000	0.997	**3.960**	0	5.987	0.994	3.429	0	5.973	0.975	2.903	0	6.033	0.973	2.904	0
MP3	21.282	0.997	3.965	0	21.787	0.998	**4.197**	0	21.649	0.998	3.931	0	20.924	0.997	3.758	0	16.209	0.988	3.148	0
AS	11.840	0.998	**4.228**	0	11.840	0.998	**4.228**	0	11.840	0.998	**4.228**	0	11.840	0.998	**4.228**	0	11.840	0.998	**4.228**	0
For speech signal Sp12																				
No attack	17.849	0.992	**4.220**	0	17.849	0.992	**4.220**	0	17.849	0.992	**4.220**	0	17.849	0.992	**4.220**	0	17.849	0.992	**4.220**	0
AWGN	11.608	0.968	**3.189**	0	12.350	0.973	2.963	0	10.012	0.956	2.757	1	11.137	0.966	2.922	0	7.105	0.928	2.584	0
LPF	13.573	0.978	3.114	0	15.415	0.986	**3.652**	0	14.135	0.981	3.091	0	10.938	0.959	2.929	0	11.210	0.962	2.822	0
RQ	17.077	0.990	3.835	0	17.052	0.990	3.816	0	17.463	0.991	**4.128**	0	16.489	0.989	3.872	0	16.033	0.988	3.713	0
RS	5.863	0.985	3.397	0	5.832	0.990	**3.893**	0	5.806	0.986	3.295	0	5.893	0.923	2.792	0	5.804	0.974	3.062	0
MP3	16.953	0.991	4.077	1	17.091	0.992	**4.219**	1	16.874	0.991	3.971	1	16.660	0.990	3.909	1	14.753	0.984	3.435	1
AS	11.086	0.992	**4.220**	0	11.086	0.992	**4.220**	0	11.086	0.992	**4.220**	0	11.086	0.992	**4.220**	0	11.086	0.992	**4.220**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp13																				
No attack	25.797	0.999	**4.292**	0	25.797	0.999	**4.292**	0	25.797	0.999	**4.292**	0	25.797	0.999	**4.292**	0	25.797	0.999	**4.292**	0
AWGN	12.453	0.974	**3.076**	0	13.905	0.982	2.908	0	12.321	0.974	2.973	0	12.433	0.975	2.852	0	7.894	0.935	2.517	0
LPF	12.652	0.973	3.014	0	18.759	0.994	**3.549**	0	15.431	0.986	3.142	0	13.185	0.976	2.983	0	12.290	0.971	2.917	0
RQ	21.806	0.997	3.856	0	22.033	0.997	3.880	0	23.963	0.998	**4.001**	0	20.297	0.995	3.643	0	20.191	0.995	3.627	0
RS	5.962	0.986	3.324	0	5.977	0.997	**3.880**	0	5.982	0.989	3.374	0	6.023	0.970	2.941	0	5.977	0.983	3.021	0
MP3	21.294	0.997	3.868	0	22.254	0.998	**4.262**	0	21.784	0.998	4.013	0	20.759	0.997	3.622	0	17.196	0.991	3.263	0
AS	11.873	0.999	**4.292**	0	11.873	0.999	**4.292**	0	11.873	0.999	**4.292**	0	11.873	0.999	**4.292**	0	11.873	0.999	**4.292**	0
For speech signal Sp14																				
No attack	19.804	0.995	**4.215**	0	19.804	0.995	**4.215**	0	19.804	0.995	**4.215**	0	19.804	0.995	**4.215**	0	19.804	0.995	**4.215**	0
AWGN	13.487	0.978	**3.227**	0	13.969	0.981	2.942	0	12.695	0.975	2.915	2	12.931	0.976	2.888	0	8.729	0.949	2.529	0
LPF	15.103	0.985	3.085	0	16.199	0.989	**3.539**	0	16.049	0.988	3.199	0	13.265	0.976	2.886	0	13.939	0.980	2.808	0
RQ	19.198	0.994	3.776	0	18.882	0.994	3.940	0	19.401	0.994	**4.087**	0	18.083	0.992	3.555	0	18.345	0.993	3.938	0
RS	5.893	0.989	3.272	0	5.888	0.994	**3.919**	0	5.907	0.991	3.369	0	5.898	0.950	2.778	0	5.823	0.984	2.996	0
MP3	18.583	0.994	4.007	0	18.551	0.994	**4.178**	0	18.548	0.994	3.955	0	18.066	0.993	3.760	0	16.218	0.989	3.552	0
AS	11.408	0.995	**4.215**	0	11.408	0.995	**4.215**	0	11.408	0.995	**4.215**	0	11.408	0.995	**4.215**	0	11.408	0.995	**4.215**	0
For speech signal Sp15																				
No attack	25.061	0.998	**4.241**	0	25.061	0.998	**4.241**	0	25.061	0.998	**4.241**	0	25.061	0.998	**4.241**	0	25.061	0.998	**4.241**	0
AWGN	11.479	0.969	2.765	0	13.606	0.981	**2.964**	0	11.445	0.970	2.758	0	12.372	0.975	2.797	0	8.648	0.948	2.324	0
LPF	15.201	0.985	3.008	0	18.393	0.993	**3.387**	0	16.290	0.988	3.119	0	12.548	0.972	2.741	0	11.535	0.966	2.570	0
RQ	20.476	0.996	3.646	0	21.560	0.997	3.754	0	23.460	0.998	**4.038**	0	20.055	0.995	3.523	0	20.090	0.995	3.542	0
RS	6.060	0.992	3.310	0	6.005	0.997	**3.811**	0	5.968	0.993	3.340	0	5.974	0.958	2.767	0	6.021	0.975	2.735	0
MP3	21.288	0.998	3.887	0	21.827	0.998	**4.156**	0	21.429	0.998	3.957	0	20.674	0.997	3.759	0	15.311	0.985	2.957	0
AS	11.843	0.998	**4.241**	0	11.843	0.998	**4.241**	0	11.843	0.998	**4.241**	0	11.843	0.998	**4.241**	0	11.843	0.998	**4.241**	0
For speech signal Sp16																				
No attack	16.659	0.989	**4.115**	0	16.659	0.989	**4.115**	0	16.659	0.989	**4.115**	0	16.659	0.989	**4.115**	0	16.659	0.989	**4.115**	0
AWGN	10.592	0.960	2.752	0	11.983	0.971	2.794	0	11.630	0.968	**2.924**	0	11.066	0.963	2.740	0	8.231	0.937	2.460	0
LPF	7.792	0.929	2.727	0	15.105	0.985	**3.426**	0	14.588	0.982	3.327	0	14.055	0.980	3.152	0	12.272	0.970	2.906	0
RQ	15.418	0.986	3.681	0	15.981	0.987	3.581	0	16.452	0.989	**3.895**	0	15.876	0.987	3.550	0	15.441	0.986	3.462	0
RS	5.947	0.962	2.915	0	5.773	0.988	**3.725**	0	5.757	0.985	3.433	0	5.725	0.977	3.334	0	5.849	0.971	2.905	0
MP3	15.103	0.986	3.615	0	15.975	0.989	3.856	0	15.916	0.988	**3.906**	0	15.843	0.988	3.878	0	14.082	0.981	3.215	0
AS	10.822	0.989	**4.115**	0	10.822	0.989	**4.115**	0	10.822	0.989	**4.115**	0	10.822	0.989	**4.115**	0	10.822	0.989	**4.115**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp17																				
No attack	16.552	0.989	**4.208**	0	16.552	0.989	**4.208**	0	16.552	0.989	**4.208**	0	16.552	0.989	**4.208**	0	16.552	0.989	**4.208**	0
AWGN	11.569	0.966	2.850	0	11.096	0.965	2.802	0	10.683	0.961	2.742	0	11.834	0.968	**2.879**	0	7.468	0.933	2.524	0
LPF	11.908	0.968	2.882	0	14.846	0.984	**3.415**	0	14.110	0.980	3.236	0	11.922	0.968	2.901	0	12.318	0.971	2.862	0
RQ	16.030	0.987	3.628	1	16.076	0.988	3.824	0	16.236	0.988	**3.885**	0	15.653	0.986	3.515	0	15.245	0.985	3.567	0
RS	5.744	0.978	3.126	0	5.743	0.988	**3.865**	0	5.734	0.985	3.469	0	5.924	0.949	2.730	0	5.777	0.979	3.026	0
MP3	15.780	0.988	3.910	0	15.954	0.989	**4.141**	0	15.886	0.988	3.889	0	15.279	0.986	3.669	0	13.799	0.979	3.150	0
AS	10.797	0.989	**4.208**	0	10.797	0.989	**4.208**	0	10.797	0.989	**4.208**	0	10.797	0.989	**4.208**	0	10.797	0.989	**4.208**	0
For speech signal Sp18																				
No attack	15.354	0.985	**4.118**	0	15.354	0.985	**4.118**	0	15.354	0.985	**4.118**	0	15.354	0.985	**4.118**	0	15.354	0.985	**4.118**	0
AWGN	11.041	0.962	**3.096**	0	11.336	0.965	2.839	0	9.323	0.951	2.731	0	10.172	0.956	2.646	0	6.414	0.914	2.244	0
LPF	10.523	0.956	2.971	0	13.518	0.978	**3.269**	0	12.914	0.974	3.183	0	11.690	0.966	2.883	0	10.272	0.954	2.767	0
RQ	15.058	0.984	**3.888**	0	14.959	0.984	3.838	0	15.152	0.985	**3.888**	0	14.500	0.982	3.413	0	14.435	0.982	3.511	0
RS	5.656	0.970	3.195	0	5.658	0.984	**3.639**	1	5.709	0.981	3.428	0	5.719	0.966	3.035	0	5.821	0.967	2.993	0
MP3	14.807	0.985	3.876	0	14.891	0.985	**4.031**	0	14.785	0.985	3.945	0	14.717	0.984	3.700	0	13.100	0.976	3.102	0
AS	10.466	0.985	**4.118**	0	10.466	0.985	**4.118**	0	10.466	0.985	**4.118**	0	10.466	0.985	**4.118**	0	10.466	0.985	**4.118**	0
For speech signal Sp19																				
No attack	21.462	0.996	**4.218**	0	21.462	0.996	**4.218**	0	21.462	0.996	**4.218**	0	21.462	0.996	**4.218**	0	21.462	0.996	**4.218**	0
AWGN	13.074	0.977	**2.956**	0	14.970	0.985	2.891	0	11.998	0.970	2.699	1	13.235	0.978	2.848	0	9.478	0.951	2.494	0
LPF	14.964	0.984	3.078	0	17.549	0.992	**3.440**	0	14.838	0.983	2.987	0	13.087	0.975	2.816	0	11.498	0.964	2.672	0
RQ	20.405	0.995	3.790	0	20.345	0.995	3.809	0	20.789	0.996	**3.936**	0	18.705	0.993	3.499	0	19.508	0.994	3.616	0
RS	6.004	0.990	3.284	0	5.937	0.995	**3.748**	0	5.986	0.989	3.178	0	5.990	0.973	2.963	0	5.961	0.976	2.856	0
MP3	19.420	0.996	3.742	0	19.434	0.996	**3.933**	0	19.326	0.996	3.788	0	19.035	0.995	3.699	0	15.501	0.986	3.034	0
AS	11.599	0.996	**4.218**	0	11.599	0.996	**4.218**	0	11.599	0.996	**4.218**	0	11.599	0.996	**4.218**	0	11.599	0.996	**4.218**	0
For speech signal Sp20																				
No attack	19.151	0.994	**4.222**	0	19.151	0.994	**4.222**	0	19.151	0.994	**4.222**	0	19.151	0.994	**4.222**	0	19.151	0.994	**4.222**	0
AWGN	11.525	0.968	3.107	1	13.070	0.977	3.054	1	11.943	0.971	**3.141**	0	12.744	0.976	3.005	1	7.756	0.941	2.689	0
LPF	8.447	0.939	2.937	0	16.438	0.989	**3.651**	0	15.062	0.984	3.304	0	13.034	0.975	3.226	0	11.761	0.967	2.946	0
RQ	17.680	0.991	3.781	0	18.257	0.993	4.003	0	18.755	0.993	**4.151**	0	17.660	0.991	4.003	0	17.461	0.991	3.626	0
RS	6.006	0.965	3.216	1	5.872	0.993	**4.029**	0	5.866	0.989	3.509	0	5.879	0.968	3.178	0	5.913	0.979	3.096	1
MP3	17.479	0.993	4.019	0	18.178	0.994	**4.133**	0	17.861	0.993	3.930	0	17.561	0.993	3.881	0	15.357	0.986	3.348	0
AS	11.314	0.994	**4.222**	0	11.314	0.994	**4.222**	0	11.314	0.994	**4.222**	0	11.314	0.994	**4.222**	0	11.314	0.994	**4.222**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp21																				
No attack	25.710	0.999	**4.315**	0	25.710	0.999	**4.315**	0	25.710	0.999	**4.315**	0	25.710	0.999	**4.315**	0	25.710	0.999	**4.315**	0
AWGN	11.803	0.970	**3.847**	0	13.682	0.981	3.605	0	11.419	0.968	3.497	0	12.638	0.976	3.556	0	8.575	0.945	3.206	0
LPF	12.142	0.971	3.562	0	19.151	0.994	**4.190**	0	17.204	0.990	3.927	0	13.417	0.977	3.598	0	13.696	0.979	3.643	0
RQ	21.752	0.997	4.152	0	22.618	0.997	4.227	0	23.879	0.998	**4.300**	0	20.806	0.996	4.171	0	19.511	0.994	3.979	0
RS	6.032	0.984	3.732	0	5.967	0.997	**4.277**	0	5.977	0.994	4.044	0	6.046	0.970	3.533	0	6.056	0.984	3.596	0
MP3	21.437	0.998	4.254	0	22.032	0.998	**4.300**	0	21.976	0.998	4.294	0	21.288	0.997	4.126	0	18.591	0.994	4.005	0
AS	11.870	0.999	**4.315**	0	11.870	0.999	**4.315**	0	11.870	0.999	**4.315**	0	11.870	0.999	**4.315**	0	11.870	0.999	**4.315**	0
For speech signal Sp22																				
No attack	24.736	0.998	**4.284**	0	24.736	0.998	**4.284**	0	24.736	0.998	**4.284**	0	24.736	0.998	**4.284**	0	24.736	0.998	**4.284**	0
AWGN	10.244	0.958	2.952	0	12.067	0.973	3.143	0	11.285	0.968	3.013	0	10.901	0.965	**3.237**	0	7.080	0.929	2.469	0
LPF	11.707	0.967	3.260	0	18.138	0.993	**3.807**	0	16.186	0.988	3.513	0	13.810	0.979	3.173	0	11.023	0.963	2.981	0
RQ	20.839	0.996	3.688	0	21.670	0.997	**4.181**	0	23.213	0.998	4.154	0	19.627	0.995	3.943	0	20.232	0.995	4.083	0
RS	5.903	0.981	3.521	0	5.990	0.997	**4.178**	0	6.003	0.991	3.593	0	6.013	0.959	2.892	0	6.186	0.974	3.123	0
MP3	20.915	0.997	4.077	0	21.670	0.998	**4.282**	0	20.923	0.997	4.165	0	20.534	0.997	4.051	0	16.458	0.989	3.199	0
AS	11.827	0.998	**4.284**	0	11.827	0.998	**4.284**	0	11.827	0.998	**4.284**	0	11.827	0.998	**4.284**	0	11.827	0.998	**4.284**	0
For speech signal Sp23																				
No attack	26.321	0.999	**4.318**	0	26.321	0.999	**4.318**	0	26.321	0.999	**4.318**	0	26.321	0.999	**4.318**	0	26.321	0.999	**4.318**	0
AWGN	9.572	0.955	**3.519**	0	14.195	0.983	3.379	0	11.186	0.966	3.339	0	12.883	0.977	3.478	0	8.544	0.943	2.910	0
LPF	12.906	0.975	3.566	0	19.268	0.994	**3.945**	0	18.291	0.993	3.830	0	12.801	0.974	3.434	0	13.399	0.977	3.274	0
RQ	19.890	0.995	4.115	0	22.856	0.997	4.192	0	24.275	0.998	**4.268**	0	21.708	0.997	3.976	0	19.859	0.995	3.813	0
RS	6.055	0.984	3.771	0	5.994	0.997	**4.203**	0	6.013	0.996	3.963	0	6.081	0.954	3.360	0	5.941	0.983	3.487	0
MP3	22.216	0.998	4.290	0	22.472	0.999	**4.299**	0	21.871	0.998	4.260	0	19.446	0.995	4.018	0	16.769	0.990	3.298	0
AS	11.891	0.999	**4.318**	0	11.891	0.999	**4.318**	0	11.891	0.999	**4.318**	0	11.891	0.999	**4.318**	0	11.891	0.999	**4.318**	0
For speech signal Sp24																				
No attack	36.535	1.000	**4.321**	0	36.535	1.000	**4.321**	0	36.535	1.000	**4.321**	0	36.535	1.000	**4.321**	0	36.535	1.000	**4.321**	0
AWGN	10.492	0.962	**3.274**	0	13.575	0.981	3.237	0	11.708	0.972	3.056	0	13.494	0.979	3.207	0	8.881	0.949	2.610	0
LPF	14.992	0.984	3.344	0	19.428	0.994	**3.905**	0	16.441	0.989	3.415	1	12.616	0.973	3.272	0	11.259	0.965	3.099	0
RQ	20.871	0.996	3.949	0	24.313	0.998	4.096	0	27.933	0.999	**4.296**	0	22.933	0.998	4.054	0	22.389	0.997	3.957	0
RS	6.096	0.991	3.551	0	6.028	0.998	**4.113**	0	6.014	0.995	3.711	0	6.104	0.964	3.216	0	6.082	0.975	3.227	0
MP3	23.928	0.999	4.201	0	24.402	1.000	**4.317**	0	23.525	0.999	4.161	0	20.948	0.997	3.915	0	17.105	0.991	3.130	0
AS	12.027	1.000	**4.321**	0	12.027	1.000	**4.321**	0	12.027	1.000	**4.321**	0	12.027	1.000	**4.321**	0	12.027	1.000	**4.321**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp25																				
No attack	31.617	1.000	**4.356**	0	31.617	1.000	**4.356**	0	31.617	1.000	**4.356**	0	31.617	1.000	**4.356**	0	31.617	1.000	**4.356**	0
AWGN	10.511	0.964	**3.487**	0	14.124	0.983	3.417	0	11.436	0.969	3.362	0	12.223	0.974	3.406	0	8.662	0.946	2.995	0
LPF	11.774	0.968	3.630	0	18.760	0.994	**3.956**	0	19.120	0.994	3.928	0	15.405	0.985	3.479	0	12.853	0.974	3.239	0
RQ	20.135	0.995	4.046	0	23.488	0.998	4.133	0	27.110	0.999	**4.323**	0	22.636	0.997	4.216	0	20.780	0.996	4.146	0
RS	6.033	0.979	3.878	0	6.009	0.998	**4.244**	0	6.041	0.996	4.062	0	5.998	0.977	3.488	0	5.984	0.983	3.422	0
MP3	22.911	0.999	4.302	0	23.788	0.999	**4.329**	0	23.383	0.999	4.308	0	22.944	0.999	4.282	0	18.639	0.994	3.377	0
AS	11.996	1.000	**4.356**	0	11.996	1.000	**4.356**	0	11.996	1.000	**4.356**	0	11.996	1.000	**4.356**	0	11.996	1.000	**4.356**	0
For speech signal Sp26																				
No attack	20.850	0.996	**4.295**	0	20.850	0.996	**4.295**	0	20.850	0.996	**4.295**	0	20.850	0.996	**4.295**	0	20.850	0.996	**4.295**	0
AWGN	10.592	0.961	3.075	0	12.790	0.975	**3.166**	0	10.285	0.960	3.086	0	11.198	0.968	3.143	0	7.578	0.936	2.831	0
LPF	12.659	0.973	3.114	0	17.002	0.990	**3.763**	0	15.370	0.985	3.410	0	14.126	0.980	3.304	0	11.679	0.967	3.133	0
RQ	18.628	0.993	3.878	0	19.528	0.994	4.058	0	20.182	0.995	**4.127**	0	17.940	0.992	4.048	0	17.886	0.992	3.619	0
RS	5.911	0.982	3.272	0	5.931	0.994	**4.101**	0	5.992	0.988	3.557	0	5.995	0.970	3.189	0	5.991	0.974	3.314	0
MP3	18.947	0.995	3.969	0	19.254	0.995	**4.207**	0	18.905	0.995	3.989	0	18.656	0.994	4.037	0	15.394	0.986	3.925	0
AS	11.536	0.996	**4.295**	0	11.536	0.996	**4.295**	0	11.536	0.996	**4.295**	0	11.536	0.996	**4.295**	0	11.536	0.996	**4.295**	0
For speech signal Sp27																				
No attack	25.389	0.999	**4.261**	0	25.389	0.999	**4.261**	0	25.389	0.999	**4.261**	0	25.389	0.999	**4.261**	0	25.389	0.999	**4.261**	0
AWGN	11.324	0.966	**3.346**	0	12.666	0.975	3.095	0	11.101	0.965	3.036	0	12.189	0.973	3.278	0	7.624	0.935	2.796	0
LPF	14.601	0.983	3.311	0	18.949	0.994	**3.922**	0	16.964	0.990	3.631	0	12.696	0.974	3.089	0	13.027	0.976	3.201	0
RQ	21.822	0.997	3.912	0	21.318	0.996	4.081	0	23.394	0.998	**4.153**	0	20.279	0.995	3.938	0	20.320	0.995	3.795	0
RS	5.985	0.992	3.551	0	6.003	0.997	**4.135**	0	5.970	0.994	3.830	0	6.062	0.940	2.984	0	6.156	0.983	3.384	0
MP3	21.926	0.998	4.220	0	22.136	0.998	**4.251**	0	21.675	0.998	4.162	0	19.751	0.996	3.931	0	16.288	0.988	3.500	0
AS	11.856	0.999	**4.261**	0	11.856	0.999	**4.261**	0	11.856	0.999	**4.261**	0	11.856	0.999	**4.261**	0	11.856	0.999	**4.261**	0
For speech signal Sp28																				
No attack	22.564	0.997	**4.270**	0	22.564	0.997	**4.270**	0	22.564	0.997	**4.270**	0	22.564	0.997	**4.270**	0	22.564	0.997	**4.270**	0
AWGN	11.771	0.969	**3.081**	1	13.235	0.979	3.047	0	11.135	0.966	2.970	0	12.145	0.973	3.047	0	6.770	0.926	2.443	1
LPF	11.961	0.969	3.049	1	17.992	0.992	**3.757**	0	15.487	0.986	3.412	0	14.177	0.981	3.203	0	12.150	0.970	2.865	0
RQ	20.207	0.995	3.770	0	20.598	0.996	3.991	0	21.546	0.996	**4.219**	0	19.994	0.995	3.827	0	19.050	0.994	3.825	0
RS	5.947	0.983	3.307	1	5.987	0.996	**3.932**	0	5.890	0.989	3.646	0	5.979	0.959	2.987	0	5.993	0.976	2.985	0
MP3	19.835	0.996	4.029	0	20.519	0.997	**4.268**	0	20.166	0.996	4.064	0	19.765	0.996	4.059	1	15.538	0.986	3.187	1
AS	11.694	0.997	**4.270**	0	11.694	0.997	**4.270**	0	11.694	0.997	**4.270**	0	11.694	0.997	**4.270**	0	11.694	0.997	**4.270**	0

Table 6 continued

Attack	Chorus				Classical				Jazz				Pop1				Pop2			
	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW	SNR_{rs}	NCC	PESQ	EW
For speech signal Sp29																				
No attack	20.162	0.995	**4.211**	0	20.162	0.995	**4.211**	0	20.162	0.995	**4.211**	0	20.162	0.995	**4.211**	0	20.162	0.995	**4.211**	0
AWGN	10.961	0.963	**3.143**	0	12.320	0.974	3.117	0	10.745	0.962	3.020	1	12.536	0.975	**3.143**	0	7.726	0.940	2.869	1
LPF	12.963	0.975	3.233	1	16.106	0.989	3.608	1	15.499	0.986	3.320	1	12.907	0.974	**3.245**	0	12.123	0.970	3.060	0
RQ	17.815	0.992	3.742	0	18.657	0.993	3.938	2	19.563	0.994	**4.132**	0	18.250	0.993	3.876	3	17.407	0.991	3.721	0
RS	5.952	0.985	3.433	0	5.913	0.994	3.923	2	5.893	0.989	**3.446**	0	5.915	0.965	3.160	0	5.973	0.981	3.232	0
MP3	18.272	0.994	4.090	1	18.854	0.995	**4.196**	1	18.690	0.994	4.001	1	18.150	0.993	3.857	1	15.681	0.987	3.690	1
AS	11.455	0.995	**4.211**	0	11.455	0.995	**4.211**	0	11.455	0.995	**4.211**	0	11.455	0.995	**4.211**	0	11.455	0.995	**4.211**	0
For speech signal Sp30																				
No attack	22.149	0.997	**4.211**	0	22.149	0.997	**4.211**	0	22.149	0.997	**4.211**	0	22.149	0.997	**4.211**	0	22.149	0.997	**4.211**	0
AWGN	9.586	0.954	3.147	2	13.490	0.980	**3.224**	2	11.328	0.968	3.271	3	12.926	0.977	3.345	3	7.569	0.934	2.695	3
LPF	13.065	0.976	3.302	1	17.857	0.992	**3.972**	0	16.747	0.989	3.587	0	12.390	0.972	3.228	0	13.135	0.976	3.124	3
RQ	18.795	0.993	4.056	0	20.333	0.995	4.138	0	21.311	0.996	**4.185**	0	19.415	0.994	3.837	0	19.023	0.994	3.844	0
RS	5.928	0.987	3.499	0	5.942	0.996	**4.098**	0	5.946	0.993	3.901	0	5.943	0.941	3.040	0	5.909	0.980	3.194	2
MP3	19.852	0.996	4.085	0	20.198	0.997	**4.198**	0	19.993	0.996	4.173	0	18.577	0.994	3.941	0	17.383	0.991	3.582	1
AS	11.661	0.997	**4.211**	0	11.661	0.997	**4.211**	0	11.661	0.997	**4.211**	0	11.661	0.997	**4.211**	0	11.661	0.997	**4.211**	0

Bold values represent the best PESQ scores obtained for the reconstructed speech signal under various signal processing attacks

Table 7 Robustness test results of proposed watermarking algorithm (Average of 150 tests)

Attack	Parameter		
	SNR_{rs} (dB)	NCC	PESQ
No attack	23.42	1.00	4.26
AWGN	11.24	0.96	3.08
LPF	14.48	0.98	3.37
RQ	19.97	0.99	3.96
RS	5.96	0.98	3.52
MP3	18.73	0.99	3.96
AS	11.59	0.99	4.25

Table 8 SNR and correlation coefficient of reconstructed secret speech under no attack condition

Method	SNR_{rs} (dB)	NCC	PESQ
[9]	14.50	1.00	–
[12]	30.74	0.99	–
[14]	31.85	0.97	4.40
[16]	34.20	0.99	–
Proposed	23.42	1.00	4.26

algorithms. It is observed that the SNR of the reconstructed speech is lesser when compared to the methods proposed in [12,14,16]. This is due to the fact that the number of DCT coefficients being embedded are limited by a factor CF $= 6/8$ as discussed in Sect. 4.1. Even though SNR_{rs} is lesser, the secret speech is reconstructed with a PESQ score of greater than 4.0 and NCC equal to unity.

The performance of proposed audio watermarking algorithm is compared with the techniques in [9,10,14,16]. Table 9 shows the comparison of robustness test results of the proposed algorithm with the relevant techniques. It is observed that the proposed algorithm shows better robustness toward AWGN and resampling attacks compared to the technique presented in [14]. From these experimental results, it is evident that the proposed audio watermarking technique shows good robustness to the signal processing attacks and is able to reconstruct the secret speech with the correlation closer to unity and an average PESQ score of 3.78.

4.3 Security Test

To evaluate the security of the watermarked audio, two approaches were adopted here:

1. False positive test: To ensure that the watermark cannot be extracted from the U and V matrices of other watermarked audio signals, the false positive test is performed as follows:

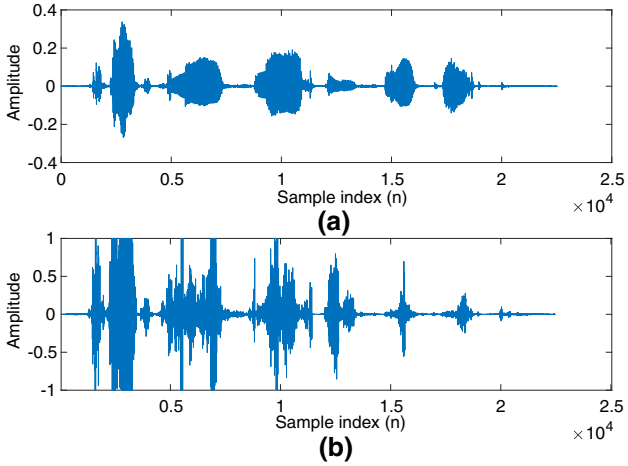

Fig. 11 Result of false positive test **a** original speech signal Sp01, **b** reconstructed speech signal with incorrect U and V matrices

The secret speech 'Sp01' is embedded in two different cover audios namely, chorus and classical. At extraction side, an attempt was made to reconstruct the secret speech 'Sp01' from chorus watermarked audio by using the U and V matrices of classical audio.

From Fig. 11, it is evident that the extraction of secret speech is not possible with incorrect U and V matrices. The reason is that the embedding is performed in SVD matrix of cover audio. Since, this decomposition is unique for each audio signal, it is not possible to extract the watermark from other cover audio signals.

2. Sensitivity to initial conditions: In this paper, the secret speech is embedded chaotically to increase the security of watermarking. So, a logistic chaotic map is chosen to generate random numbers with the initial conditions $y_0 = 0.052$ and $r = 3.95$.

Figure 12 shows the effect of sensitivity to y_0 and r values. It is observed that even if an intruder guesses value of r exactly and y_0 with an error of 10^{-10}, the extracted speech is not intelligible when compared to original speech.

From these results, it is evident that the proposed watermarking technique is secured against the intruder attacks as discussed above.

5 Conclusion

In this paper, a watermarking algorithm for chaotic embedding of DCT compressed speech signal using DWT and SVD is proposed. The DCT compression of secret speech signal is achieved by finding the suitable number of DCT coefficients that are required for embedding such that the PESQ score of the decompressed signal is greater than 4.0 to ensure the speech quality. To increase the security of watermarking

Table 9 Comparison of robustness test results of the proposed algorithm with the relevant techniques

(a) Comparison of reconstructed secret speech quality

Attack	SNR_{rs} (dB)		PESQ	
	In [14]	Proposed	In [14]	Proposed
AWGN	−8.5	11.24	1.24	3.08
RS	−9.46	5.96	1.42	3.52

(b) Comparison of correlation between original and reconstructed secret speech

Attack	NCC			
	In [9]	In [10]	In [16]	Proposed
AWGN	0.91	0.96	0.92	0.96
RS	–	–	–	0.98

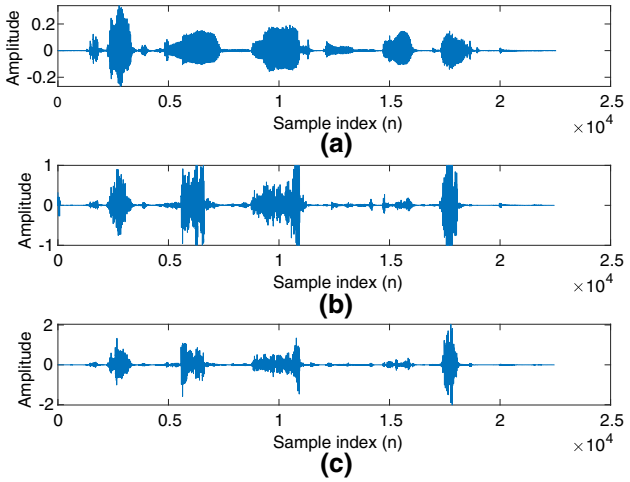

Fig. 12 Reconstructed speech signals with different initial conditions **a** original speech signal Sp01, **b** reconstructed speech signal with $y_0 = 0.0520000001$, **c** reconstructed speech signal with $r = 3.949999$

algorithm, logistic map is used to generate random numbers, and watermarking of the selected DCT coefficients is performed chaotically in cover audio by decomposing it using DWT followed by SVD. The experimental results show that the proposed watermarking algorithm achieves good imperceptibility with an average SNR and ODG of 46 dB and −1.07, respectively. The robustness test results show that the secret speech signal is reconstructed with an average NCC of 0.95 by preserving the perceptual quality of reconstructed speech signal under various signal processing attacks. In addition, it is found that the loss in the generality of the information of reconstructed speech signal is minimum when the watermarked audio is subjected to various signal processing attacks.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Sathiyamurthi, P.; Ramakrishnan, S.: Speech encryption using chaotic shift keying for secured speech communication. EURASIP J. Audio Speech Music Process. **1**, 20 (2017). https://doi.org/10.1186/s13636-017-0118-0
2. Lakshmi, C.; Ravi, V.M.; Thenmozhi, K.; Rayappan, J.B.B.; Amirtharajan, R.: Con (dif) fused voice to convey secret: a dual-domain approach. Multimed. Syst. **26**, 1–11 (2020)
3. Ehdaie, M.; Eghlidos, T.; Aref, M.R.: A novel secret sharing scheme from audio perspective. In: 2008 International Symposium on Telecommunications, IEEE, pp. 13–18 (2008)
4. Wang, J.Z.; Wu, T.X.; Sun, T.Y.: An audio secret sharing system based on fractal encoding. In: 2015 International Carnahan Conference on Security Technology (ICCST), IEEE, pp. 211–216 (2015)
5. Bharti, S.S.; Gupta, M.; Agarwal, S.: A novel approach for verifiable (n, n) audio secret sharing scheme. Multimed. Tools Appl. **77**(19), 25629–25657 (2018). https://doi.org/10.1007/s11042-018-5810-2
6. Djebbar, F.; Ayad, B.; Meraim, K.A.; Hamam, H.: Comparative study of digital audio steganography techniques. EURASIP J. Audio Speech Music Process. **1**, 25 (2012)
7. Hua, G.; Huang, J.; Shi, Y.Q.; Goh, J.; Thing, V.L.: Twenty years of digital audio watermarking—a comprehensive review. Signal Process. **128**, 222–242 (2016)
8. Mishra, S.; Yadav, V.K.; Trivedi, M.C.; Shrimali, T.: Audio steganography techniques: a survey. In: Advances in Computer and Computational Sciences, pp. 581–589. Springer, Berlin (2018)
9. Xu, T.; Yang, Z.; Shao, X.: Novel speech secure communication system based on information hiding and compressed sensing. In: 2009 Fourth International Conference on Systems and Networks Communications, IEEE, pp. 201–206 (2009). https://doi.org/10.1109/ICSNC.2009.71
10. Shahadi, H.I.; Jidin, R.; Way, W.H.: Lossless audio steganography based on lifting wavelet transform and dynamic stego key. Indian J. Sci. Technol. **7**(3), 323 (2014)
11. Ballesteros, L.D.M.; Moreno, A.J.M.: Highly transparent steganography model of speech signals using efficient wavelet masking. Expert Syst. Appl. **39**(10), 9141–9149 (2012). https://doi.org/10.1016/j.eswa.2012.02.066
12. Ali, A.H.; George, L.E.; Zaidan, A.; Mokhtar, M.R.: High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain. Multimed. Tools Appl. **77**(23), 31487–31516 (2018). https://doi.org/10.1007/s11042-018-6213-0

13. Ballesteros, D.M.; Renza, D.: Secure speech content based on scrambling and adaptive hiding. Symmetry **10**(12), 694 (2018). https://doi.org/10.3390/sym10120694
14. Bharti, S.S.; Gupta, M.; Agarwal, S.: A novel approach for audio steganography by processing of amplitudes and signs of secret audio separately. Multimed. Tools Appl. **78**(16), 23179–23201 (2019). https://doi.org/10.1007/s11042-019-7630-4
15. Alsabhany, A.A.; Ridzuan, F.; Azni, A.H.: The adaptive multi-level phase coding method in audio steganography. IEEE Access **7**, 129291–129306 (2019)
16. Ali, A.H.; George, L.E.; Mokhtar, M.R.: An adaptive high capacity model for secure audio communication based on fractal coding and uniform coefficient modulation. Circuits Syst. Signal Process. **39**, 1–28 (2020). https://doi.org/10.1007/s00034-020-01409-7
17. Kasetty, P.K.; Kanhe, A.: Covert speech communication through audio steganography using DWT and SVD. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp. 1–5 (2020). https://doi.org/10.1109/ICCCNT49239.2020.9225399
18. Hassan, T.A.; Al-Hashemy, R.H.; Ajel, R.I.: Speech signal compression algorithm based on the JPEG technique. J. Intell. Syst. **29**(1), 554–564 (2020). https://doi.org/10.1515/jisys-2018-0127
19. ITU-T R: Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Rec ITU-T P 862 (2001)
20. Al-Azawi, M.K.M.; Gaze, A.M.: Combined speech compression and encryption using chaotic compressive sensing with large key size. IET Signal Process. **12**(2), 214–218 (2017). https://doi.org/10.1049/iet-spr.2016.0708
21. Weisstein, E.W.: Logistic Map from MathWorld—A Wolfram Web Resource. https://mathworld.wolfram.com/LogisticMap.html. Accessed Feb 2021 (2021)
22. Peng, J.; Jiang, Y.; Tang, S.; Meziane, F.: Security of streaming media communications with logistic map and self-adaptive detection-based steganography. IEEE Trans. Depend. Secure Comput. (2019). https://doi.org/10.1109/tdsc.2019.2946138
23. Burrus, C.; Gopinath, R.; Guo, H.: Introduction to Wavelets and Wavelet Transforms—A Primer. Prentice-Hall, New Jersey (1998)
24. Chen, S.T.; Huang, H.N.: Optimization-based audio watermarking with integrated quantization embedding. Multimed. Tools Appl. **75**, 4735–4751 (2016). https://doi.org/10.1016/0003-4916(63)90068-X
25. Kaur, A.; Dutta, M.K.: An optimized high payload audio watermarking algorithm based on LU-factorization. Multimed. Syst. **24**(3), 341–353 (2018). https://doi.org/10.1007/s00530-017-0545-x
26. Kanhe, A.; Gnanasekaran, A.: A QIM-based energy modulation scheme for audio watermarking robust to synchronization attack. Arab. J. Sci. Eng. **44**(4), 3415–3423 (2019). https://doi.org/10.1007/s13369-018-3540-4
27. Hwang, M.; Lee, J.; Lee, M.; Kang, H.: 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
28. VoiceAge: Unified speech and audio database (USAC). http://www.voiceage.com/Audio-Samples-AMR-WB.html. Accessed Apr 2020 (2020)
29. Hu, Y.; Loizou, P.C.: Subjective comparison and evaluation of speech enhancement algorithms. Speech Commun. **49**(7), 588–601 (2007). https://doi.org/10.1016/j.specom.2006.12.006
30. Azure, M.: Microsoft™ Azure Speech Services API. https://azure.microsoft.com/en-in/services/cognitive-services/speech-to-text. Accessed Feb 2021 (2021)
31. Katzenbeisser, S.; Petitcolas, F.A.: Information Hiding Techniques for Steganography and Digital Watermarking, 1st edn Artech House, Inc., Norwood (2000)
32. Kabal, P., et al.: An examination and interpretation of ITU-R BS. 1387: perceptual evaluation of audio quality. TSP Lab Technical Report, Department Electrical & Computer Engineering, McGill University, pp. 1–89 (2002)
33. ITU-R R: Methods for objective measurements of perceived audio quality. ITU-R BS 13871 (2001)