International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 15, No.1 (Jan-24)

http://dx.doi.org/10.12785/ijcds/150123

An Efficient Tamper Detection and Recovery Scheme for Attacked Speech Signal

Younis M. Jalil , Haider I. Shahadi and Hameed R. Farhan

Electrical and Electronic Engineering, University of Kerbala, Kerbala 56001, Iraq

Received 22 Mar. 2023, Revised 2 Jan. 2024, Accepted 6 Jan. 2024, Published 15 Jan. 2024

Abstract: Data security is an essential communication issue, where the resistivity to data tampering, corruption, and attacks forms a critical problem. This paper proposes a robust approach for tamper detection in a speech signal and recovering the destroyed parts of the attacked signal from the embedded spare parts using a specific strategy. The given approach deals with the abnormalities in the quality of the message speech signal that occurs due to different attacks. To solve that issue, A watermarking method is suggested to embed self-extracted data of authentication codes, synchronization identifiers, and spare part data from the same speech signal. The authentication codes are generated using the Singular Value Decomposition (SVD) method, while G.723.1 speech CODEC generates the spare part data. The proposed strategy considers attacks such as muting and replacement. According to the experiments, the quality of the watermarked and retrieved signals has improved. The average signal-to-noise ratio is above 70 dB for the watermarked signal quality, and the normalized correlation for the retrieved signal is close to the original speech signal under different operations. The results state fully recovering rate for the attacked signal under different types and lengths of attacks (10%-80% of Tampering Rates). Additionally, this approach achieves less distortion in the watermarked signal and the recovered signal quality measured by log spectrum distortion, which is about 0.0072 and 0.225 (at high Tampering Rate), respectively. The proposed approach is compared to the related methods, where it achieves the best results, including high robustness against different types of attacks.

Keywords: Tampering Detection, Signal Recovery, Speech Watermarking, Singular Value Decomposition, Authentication Code, Speech Signal Attack.

1. Introduction

The rapid improvement in digital systems and technologies makes tampering and modification processes in speech signal content very easy and accessible by anyone. This risk affects the voice message reliability by elaborating the content information. To prevent this type of abnormality, an authentication watermark is embedded in short signal frames to improve tamper detection that is highly sensitive to malicious attacks to check signal authenticity. Authenticity-checking watermarks have been generated from the original signal as features [1] or external data such as logo images [2], [3]. For retrieving process, extracted coefficients or generated compressed data represent the recovery watermark. In general, tamper detection and recovery performances are estimated by measuring inaudibility, Recovery Rate (RR), recovery accuracy, and detection accuracy. According to the embedding mechanism, types, and attacked locations, the performance estimation parameters are affected and changed. So, improving one of them will affect the others. The frame-wise domain is used in most related schemes, where the features extracted from each speech signal frame are used as an identifier for tamper detection in such a framework [4].

The existing evaluation of the feature extraction mechanisms leads to issues in signal quality, embedding capacity, error detection, and RR of the retrieved corrupted data. The presented work has the ability for tamper detection and recovery. For quality degradation, the data concealing algorithm provides high inaudibility for watermark data at high embedding capacity. The secret data is encrypted at a high level of security to improve confidentiality through the transmission process. It is robust against noise production in the watermarked signal due to the data concealing algorithm. This work satisfies embedding capacity up to 7% from the original speech signal with at least 65 dB of quality in signal-to-noise ratio. Error detection is very sensitive to the attack, with recovery for the lost part reaching 100%. This RR is satisfied until 80% of the original signal length is lost from the received signal.

The rest of the paper is organized as follows: section two presents the related work with its drawbacks. Section three gives the methodology used with the details of the proposed tampering and recovery system. Section four discusses the results and the system performance. The last section summarizes the key findings and recommendations

younis.m@s.uokerbala.edu.iq, haider_almayaly@uokerbala.edu.iq, hameed.rf@uokerbala.edu.iq

for future research.

2. Related Work

In recent years, various methods have been proposed to locate tampering in speech signals. Some approaches also aim to recover the tampered parts from the signal. These methods typically involve extracting and utilizing internal or external features to authenticate the signal and achieve signal recovery [4], [5]. The previous studies that can be used for this purpose are classified based on the type of watermark data extracted for tamper detection and recovery.

The speech signal size is minimized to a specific number of bits by digitizing algorithms such as Most Significant Bits (MSBs) and Hash function, in which the selected data bits are used as recovery watermark data. The authors in [6] used Bessel Fourier moments extracted from speech signal frames at a specific location as an authentication code. The proposed method presented suitable imperceptibility and high robustness to common signal processing operations. On the other hand, the two segmented signals may lose the original data at unembedded signal frames, making authenticity-checking difficult. A scheme based on altering the least significant digits of the cover samples was proposed in [7] to embed authentication digits. The proposed scheme can recover the original signal after an attack with a highly embedded watermark to signal operations. The system has a drawback of high processing time, which is required for the embedding process because the embedding data overflows the number of rounds to conceal it. In [8], the integrity-checking data have been generated by performing a hash algorithm at encrypted data of the original signal. However, the checking results under common signal processing show high robustness; it needs high processing time (more than 30 seconds) for encryption and embedding processes. A tamper detection and recovery method using pixel-wise and block-wise mechanisms was introduced in [9]. The overlapping between groups of blocks at different locations to localize the tampered area was employed. The recovery and authentication bits were embedded in pixel location shared with other blocks for the embedding process. The proposed method provided high confidentiality for the watermarked signal but has low imperceptibility (maximum of 38 dB) and high required processing time due to the block-wise embedding mechanism. The researchers in [10] proposed a tamper detection and recovery method based on extracting the MSBs of the signal samples and combining them to produce a compressed signal. Based on the probability distribution and entropy calculations, the check bits were generated from the MSBs and combined with the frame number of the same frame. The three generated bit sequences were combined with a digitalized version of the compressed data and then embedded within the Least Significant Bits (LSBs) of the same frame. The tamper detection process was accomplished by comparing the extracted check bits with those extracted from the MSBs to identify the attacked region. The attacked signal frame was recovered depending on the digitalized MSBs extracted from the frame itself. This method achieved high security, inaudibility, recovered signal quality (Signal-to-Noise Ratio (SNR) = 53.33 dB) and low complexity. However, it has low resistance to additive noise, which causes loss in the embedded watermark data. An audio tampering detection and recovery scheme based on fragile watermarking was proposed in [11]. The MSBs intensities of the signal segments were summed to produce the watermark bits as feature extracted. The watermark data of each segment were represented in the binary form and embedded within the LSBs of the same segment, which were used for tamper detection and recovery. This method attained good quality and imperceptibility (SNR= 45.67) and low payload. However, it has low accuracy in tampering detection due to losing embedded watermark data under unintentional attacks such as noise. The Least Square QR factorization (LSQR) method was proposed in [12] to reconstruct an approximation version of the host signal by reducing the data size at a specific location. The tampered frames of the signal can be retrieved by extracting the linear equation of LSQR from the reserved embedding samples and solving it. The method provided less complexity and less recovered signal quality for long tampered signal length (more than 43% of the signal). The authors in [13] proposed a reference sharing mechanism and hyperchaotic system to generate and embed the compressed data for tamper recovery. The compressed data and check bits are generated using a reference sharing mechanism and hash function on encrypted speech signal frames. The result was embedded in the LSBs of every encrypted signal frame. Although the work provided a high tampering detecting mechanism, it has a weak point in the recovery process. The reference bits (recovery watermark data) can not be used for a self-embedded frame, so the tampered signal can not be recovered totally. For a tampering rate up to 20% of the signal size, the signal can be recovered at quality measure by SNR about 47.5 dB. Therefore, the method can deal only with small-length attacks at 50%, and the original signal cannot be recovered at a high tampering rate.

Several researchers proposed algorithms to produce the recovery watermark by applying frequency transform methods in the speech signal frames, such as Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Integer Wavelet Transform (IWT). Illustrations of some of these algorithms are presented. An authenticity verifying method based on DWT and hash function was proposed in [14]. For odd and even frame orders, the hash function was applied separately at the D-level DWT coefficients of each frame. The results were used as compressed data and combined with the frame number. The binary representation of the result data was used as watermark data and embedded within the detail coefficients of the same D-level DWT coefficients. For authenticity verification, two rows of watermark data were extracted for the odd and even frames and compared to each other to identify the attacked frames. The method offered high robustness to different attacks, acceptable detection accuracy, and inaudibility (SNR = 41 dB). However, the method's shortcomings are that it cannot

deal with multiple attack types, and a very small attack rate can be processed (less than 10%) concerning the overall original signal length. Speech recovery is a more valuable operation when compared to tamper detection, so several schemes are proposed in this field. An approach based on compressive sensing technique and DCT was proposed in [15]. The DCT coefficients extracted from the original signal were used to form the watermark data. The compressive sensing technique was used to recover the tampered DCT coefficients, and deep learning was employed to enhance the signal after retrieving. The approach submitted acceptable recovered signal quality at SNR = 41.5 dB and acceptable imperceptibility at SNR = 41.54 dB. Conversely, the method endured from the high impact of signal processing operations, which absorbs high processing time. In [16], an encrypted speech authentication and recovery scheme based on chaotic and block cyphers for concealing the statistical features of the signal was proposed. The IDWT is applied at each signal frame to produce the approximation and detail coefficients. The generated approximation coefficients were combined with the frame number to produce frame watermark data embedded within the detail coefficients. The method has high inaudibility (SNR=52.46 dB) of the embedded watermark data and a high ability to detect the desynchronization attack. However, it has low resistivity to other intentional attacks, which decreases the recovered data. A method based on DWT and DCT was proposed in [17] to get the compressed signals for tampered area recovery. The compressed signals are the approximation coefficients of D-level DWT embedded within the DCT coefficients, and the embedded data are used to reconstruct signals at specific tampered locations. The method achieved well for watermarked signal (Objective Difference Grade (ODG) = -0.652) and recovered signal (ODG = -1.087). The drawbacks are low accuracy at the watermark extraction step and low robustness. In [18], the DCT coefficients representing the compressed data were combined with the frame number and embedded in the selected location within the same frame. The recovered signal by this algorithm has good quality (ODG = -0.72) with high robustness and security for intentional attacks. The watermark data has a high sensitivity to signal processing operations. A self-recovery scheme-based integer DCT was introduced in [19] to extract and embed the watermark bits for recovering the attacked regions after the content replacement attack. Reference values were extracted as compressed data from each frame separately and blindly embedded. The watermark data was randomly permutated using a pseudo-random sequence to improve security. Although the scheme has acceptable imperceptibility (Peak SNR (PSNR) = 39.6 dB) and good robustness, it provides a low RR of the attacked signal under a high range of replacement attacks.Although the scheme has acceptable imperceptibility (Peak SNR (PSNR) = 39.6 dB) and good robustness, it provides a low RR of the attacked signal under a high range of replacement attacks.

Some researchers use lossy methods as source coding algorithms to produce the compressed signal. These works mainly aim to obtain a shortened version of the speech signal at a low data rate to provide low embedding capacity requirements. The authors in [20] proposed an algorithm based on source-channel coding, such that a compressed version of the input signal was generated using a lossy compression coder. The channel algorithm used the Reed Solomon (RS) algorithm to correct the attacked parts. The watermark data for each frame is generated by combining source and channel coding data and then embedding them within the same frame LSBs. The watermarking algorithm of this method offers high inaudibility (PSNR = 89.85 dB) and good quality for the recovered signal (PSNR = 40.7 dB). The RR decreases when the tampering rate increases over 50% of the total signal length and produces degradation of the recovered signal quality. The attacked signal can be recovered only at a tampering rate of less than 20% of the original signal length. The researchers in [21] suggested a tamper detection and recovery scheme based on a multipurpose watermarking algorithm. Two watermarks were embedded in the original signal for property protection and content authentication. The compressed signal was generated using the source coding scheme as a lossy compressor and hash function to generate the authenticity checking bits. For each frame, the related watermark data was embedded within each frame separately within DWT coefficients. The method has acceptable inaudibility (SNR = 19.17 dB) and robustness but produces a high degradation in the watermarked signal quality. A tampering recovery system for attacked signals based on a source-channel coding algorithm was presented in [22]. The system used the discrimination process at the sender side to decrease the input signal before compressing it by a source coder. On the receiver side, an interpolating process was used after the decoder to achieve a good quality of the recovered signal. The generated compressed data were represented by 8 bits/sample. Every two samples were concatenated in 16-bit representation and entered into the RS algorithm to generate error correction bits. Besides, there was a scrambling process between the two coding processes to perform more security depending on a secret key. The hash function was applied at each frame separately on the MSBs of the frame samples to generate the error-detecting watermark. The remaining LSBs were used to embed the watermark data in each frame. The tamper detection process was done by comparing the extracted hash values with the generated ones from the MSBs of each frame. The method was specified by the high quality of the recovered signal and high compression rate, but the attacked signal parts cannot be recovered totally.

This work proposes a tamper detection and recovery method based on a combination of strategies to perform critical issues for message recovery. It aims to establish some modifications, representing challenged points on retrieving lost parts from a modified speech message. Firstly, the embedding capacity is highly increased to satisfy multiple embedding for each compressed part from the original speech signal. Secondly, obtaining a high RR after tampering and

minimizing the distortion in the recovered speech at the receiver side. The watermark data provide high robustness to different attacks. Embedding watermark should applied at robust positions from the original speech samples to avoid noise. The proposed algorithm aims to cover all the mentioned specifications. Besides, it can estimate the RR's speech signal under continuous (muting) attacks at different tampering rates (TR). Therefore, the proposed watermarking system performs a pre-processing operation of boosting and resampling to achieve the requirements of the used compress codec. Two embedding strategies are employed to embed watermarks: self-sustained and mutual embedding. The mutual embedding is designed to deal with different tamper conditions. Its strategy is distributing copies of a recovery watermark that can effectively recover the tampered segments from untampered embedded locations. On the other hand, the self-sustained embedding strategy deals with integrity and desynchronization attacks, combining the frame's authentication and synchronization watermarks and embedding them in the same frame. Thus, the proposed system has a high TR and good recovery performance.

3. Methodology

The proposed approach solves several problems related to previous approaches, such as synchronization, signal quality, detection accuracy, and RR. The input speech signal passes through several steps to efficiently transmit and preserve speech data through a channel and a storage device. It can be summarized to compress a speech signal to decrease and remove the redundant data, making it proper for embedding. Simultaneously, the input speech signal of uncompressed version is decomposed into several fixed-sized frames. For each frame, authentication data and synchronization code are separately generated to be used later for integrity verification and authentication. The synchronization code is used not only for detecting frame boundaries but also to expand it to encompass the attack frame discovery. The produced authentication data, synchronization code, and compressed data are passed to a self-watermarking algorithm to perform the concealment process. The output speech signal is generated from transmitting the sender signal through a public channel to conserve it in a cloud server and storage device. On the receiver side, confirming that the correct recipient accepts the information in the received speech file must be subject to verification and a tampering detection system. This system can identify the sender's personality and determine whether speech frames are intact or attacked. In the attack frame, a recovery system retrieves attacked frames by extracting the intended frames' data from the embedded compressed data and recovering them. Finally, speech parts are combined and passed to the receiver based on the synchronization code.

The block diagram shown in Fig. 1 describes the main steps of the sider and receiver sides. The sender side comprises compressing and framing the input signal, generating frame IDs, authentication data generation, and the embedding algorithm to insert the required data into each frame of the original input signal. The receiver side engaged in attacked frame localization and correction.

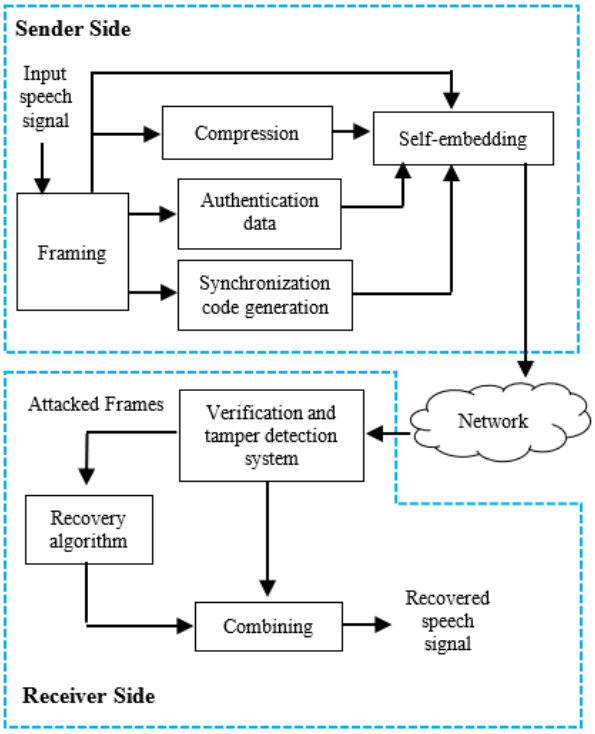

Figure 1. General block diagram of the proposed system

A. Sender Side

This section describes the operations and processes of the sender side for the proposed approach. Firstly, preprocess the input speech signal by boosting the signal energy and re-sampling it at an 8 kHz sampling frequency with a quantization of 16 bits/sample. Then, the preprocessed signal is framed and compressed by G.723.1 encoder to obtain spare parts of the original signal. The original boosted signal is also framed into non-overlapped frames, and then the frame ID for each frame is embedded for synchronization on the receiver side. Again, the frame ID is combined with each corresponding spare part to be embedded in the grouping frames of the signal. Subsequently, the authentication code for each resulting frame after embedding the spare parts is generated using the SVD algorithm to be embedded later in the same frame. Generally, the embedding process of the three watermarks data is done by dismantling the original signal samples into digits. The generated digits at selective locations within the digit sequence are modified according to the embedding process to hide the encrypted secret samples' encrypted digits. Within this operation, the size of the original signal does not change. Fig. 2 illustrates the procedure on the sender side of the proposed system.

The details of the proposed approach steps at the sender

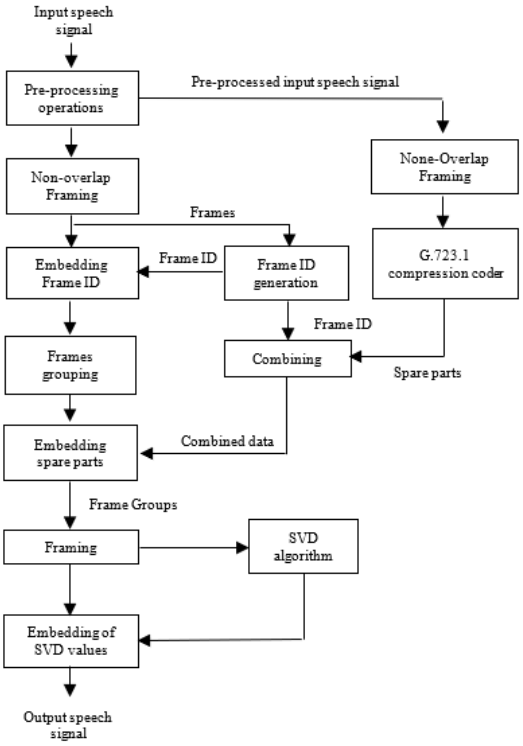

Figure 2. Procedure of the proposed method on the sender side

side are illustrated:

1) Preprocessing

The proposed system uses a speech file of 16 kHz sampling frequency and 32 bits/sample resolution as an input speech signal. It is marked as $S = a_i | 1 \le i \le M$, where i and M represent the sampling index and the total number of samples in S. The zero samples are deleted, and the remaining are boosted into an acceptable amplitude range. Then, another copy of the original input signal is made after resampling it into an 8 kHz sampling frequency and resolution of 16 bits/sample. In case of the input signal has a different sampling rate and resolution, it will be resampled into the 16 kHz with a resolution of 32 bits/sample and then complete the next procedure as mentioned above. So, the final result of this process is two signals, S_1 and S_2, representing the boosted input signal and the resampled input signal, respectively.

2) Framing

Each of S_1 and S_2 is framed into N non-overlapped frames. Each frame with size $P = M/N$ samples. The resulting frames are S_{1j} and S_{2j} , $1 \le j \le N$, where j is the frame index within the N frames sequence.

3) Signal Compression

The Compression aims to reduce the data size of the media file in sending and storing. The need to use compression is based on the assumption that media files, especially speech files, contain more information than humans can perceive. This additional information can be removed to minimize file data size. In this work, the input speech signal is compressed after preprocessing and framing (S_{2j}) to prepare the signal spare parts, which are embedded in the original signal frames to be used later as a recovery watermark at the receiver side. G.723.1 coder with lossy compression standard is utilized here because of its good compression rate, and it is the most widely used speech coder due to its mutable bit rates. Therefore, it can balance compression rate and reconstructed signal quality using algebraic code excited linear prediction and multi-pulse maximum likelihood quantization [23]. The resulted signal from this step S_3 has a $P/10$ sample per frame and a resolution of 8 bits/sample with the same frame number equal to N. So that the compression ratio is 20 times in terms of dividing the size of the S_3 over S_2.

4) Embedding Frame IDs

This step is used to obtain synchronization control on the receiver side. The index of each frame is converted into three decimal digits (Y_j) to identify the frame order in the N frames sequence. Some samples from each frame with length h_1 are preserved for embedding its index value digits. The embedding procedure is as follows:

- Y_j is mapped into a sequence of three integer digits as $Y_j = \{y_{j,2}, y_{j,1}, y_{j,0}\}$ using (1).

$$y_{j,k} = (Y_j / 10^k) \; mod \; 10 \; , \; k = 0, 1, 2 \qquad (1)$$

- Embedding Y_j the reserved cover samples from each frame depending on a threshold value (Thr) using (2) to make the value of V identical to the value of $y_{j,k}$.

$$V = ((x_1 + 2 \times sign) + 2 \times x_2) \; mod \; 10 \qquad (2)$$

where ($sign$) has a value of 1, if cover sample > 0, otherwise it is 0. The two digits x_1 and x_2 are the third and fourth decimal digits from the cover sample value. The values of these two digits are changed by increasing or decreasing (from 0-9) sequentially until $V = y_{j,k}$. The new values of x_1 and x_2 take the place of the old values in the cover sample digits. The same procedure is repeated to complete the embedding of the all three digits of Y_j for each frame.

5) Combining Frame ID with a Compressed Frame

As mentioned above, the compressed frame (S_{3j}) has the same frame number as the original frame number (N). In this step, the j index is also considered a sample and added in front of the compressed samples to obtain the frame called Spare parts (S_{3jID}), which has a length equal to $q = p/10 + 1$.

6) Grouping Frames

This step aims to obtain multiple groups carrying the same spare parts. It offers flexibility to extract the damaged data from more than one region. Because sometimes,

tampering includes several segments in the signal. Thus, the proposed approach suggests this procedure to overcome the above problem. In this step, each framing group is constructed from W frames, where the signal contains ($Q = N/W$) groups, and each group has ($Z = P \times W$) samples. The number of W frames is changed with increasing and decreasing the available embedding capacity required for embedding secret data. The resulted output from this step is D groups utilized for embedding the spare parts.

7) Embedding of Spare Parts

This step demonstrates the concept in which the spare parts are embedded into the frame groups. Each spare part frame with its ID (has $P/10+1$ samples) is embedded in D groups. To conserve the speech signal quality after the embedding process, only specific samples with high strength are selected for embedding from the length h_3 samples of each frame, which can be employed for embedding. These samples are selected based on a threshold value (Thr). So, there are some cover samples are not carried hidden data. The other cover samples for embedding are h_1 and h_2 samples, which are reserved from each S_{1j} for embedding synchronization and authentication codes, respectively. Fig. 3 shows the reserved samples within the S_{1j} frame used in embedding processes.

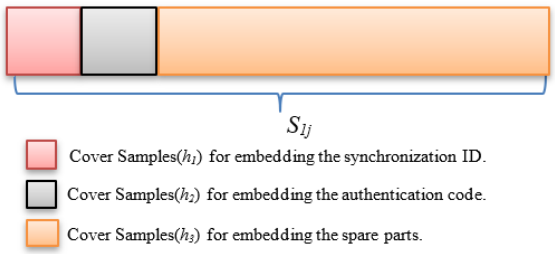

Figure 3. Reserved samples within each frame for embedding.

The following steps illustrate the spare parts embedding mechanism:

- Each sample (s_u) of the S_{3jID} frame is separated into decimal digits.

$$s_{u,k} = (s_u/10^k) \; mod \; 10 \tag{3}$$

where k : $digit\ index, u = 1, 2, P/10 + 1$, $s_{u,k}$ represents the k integer digit of s_u. Embedding s_u requires several frame samples from the cover depending on s_u digit length. The cover samples specified for embedding are used to embed each $s_{u,k}$. This is achieved by changing the values of x_1 and x_2 in (2) sequentially (between 0-9) until obtaining $V = s_{u,k}$).

- The new values of x_1 and x_2 that make the equality ($V = s_{u,k}$) are replaced with old ones from the cover sample. The same procedure is repeated to embed the other values of $s_{u,k}$ .Steps a and b are applied to other s_u to complete the spare part data embedding process in the intended frame groups. All embedding steps mentioned above are implemented to embed spare parts within the other $D-1$ remaining groups.

8) Generation of SVD

An authentication code is required to detect tampering or damage to the signal. The SVD algorithm is employed because of its high sensitivity to the input signal change. Before generating the SVD, the watermarked signal needs to be reframed into its N frames of the original S_1 signal. It is named $S_{1(watermarked)}$, referring to the cover contained within both frame ID and spare parts.

In general, SVD is a practical tool used to analyze numerical matrices.$S_{1j(watermarked)}$ matrix is a construction of three matrices such that $A = UER^T$, where R and U are orthogonal matrices, and $E = l_1^t, l_2^t, \ldots, (l_{hh})^t$ includes the singular values of $S_{1j(watermarked)}$ that are diagonally allocated [24]. Only the first singular value is used in the proposed method due to its resistivity and some properties that make it a good features extractor for a given matrix. These properties are:

- The SVD matrix size is adaptively changed with the transformation input matrix size.
- Any change in the SVD value can be easily detectable due to high sensitivity.

For all previously mentioned reasons, SVD is out sized as an extracted feature for detecting tampering in the speech signal. For $S_{1j(watermarked)}$, the SVD algorithm is applied on all $S_{1j(watermarked)}$ samples, except those samples (h_2), which are reserved for embedding the singular value. The SVD algorithm produces several singular values, one of which has the largest value and has been picked as the $S_{1j(watermarked)}$ authentication code (SVD_j value).

9) SVD Embedding

Before embedding, the $S_{1j(watermarked)}$ authentication code is permuted by a shifting mechanism to increase security. Each SVD_j value that includes L integer digits is decomposed using (1), in which $SVD_{j(value)} = SVD_{j0}, SVD_{j1}, \ldots.., SVD_{jL}$ circularly shifted to permutation its real value. Then, the $SVDs$ values of $S_{1j(watermarked)}$ frames are circularly shifted to make the operation more secure. The result is given by the symbol $SVD_{j(secured)}$ and embedded depending on the following steps:

- As mentioned above, each $SVD_{j(secured)}$ with L integer digits is embedded within h_2 samples within the $S_{1j(watermarked)}$ frame.
- For $SVD_{j(secured)} = SVD_{j0}, \ldots, SVD_{jk}, \ldots, SVD_{jL}$, where $k = 0, 1, \ldots, L$, each $SVD_{jk(secured)}$ requires at least two samples of h_2 to embed it according to (4).

$$V = (x_1 + v_1 + v_2 + 2 \times x_2) \; mod \; 10 \tag{4}$$

where $v1$ is the *sign* of the i frame sample and v_2 is the *sign* of the $i+1$ frame sample, x_1 and x_2 are digits at specific locations within the i and $i+1$ samples of h_2, respectively. The embedding is occurred by changing the values of x_1 and x_2 to satisfy the equality ($V = SVD_{jk(secured)}$). The new values of x_1 and $x_2$are pushed instead of the old values. Step (b) is repeated until all $SVD_{jk(secured)}$ values are embedded. Steps (a and b) are repeated to embed all $SVD_{j(secured)}$ values.

10) Frame Combining

Finally, the modified frames are combined according to their original order j to obtain the output speech signal (S_{out}).

B. Receiver Side

This part of the proposed system deals with the operations, including the integrity checking of the received data, tamper detection of attacked parts of the original signal, the method used to appoint the attacked locations, and the method to be used to retrieve these parts. Fig. 4 illustrates the general operations on the receiver side, which include the following steps:

1) Signal Framing

The received signal $S'(S + noise)$ of M samples is transformed to N non-overlapped frames, where each frame has a size of $P = (M/N)$ samples. The new framed signal is marked as S'_j, where $j = \{1, 2, \ldots, N\}$.

2) Generating SVD_j

The SVD algorithm is applied to each S'_j to generate the singular matrix. The highest value is selected from the resulting values to be the authentication data, marked as SVD_j.

3) Extracting SVD'_j

Every two continues samples of h_2 samples within S'_j are subjected to (4) to generate the L integer digits of $SVD_{j(secured)}$. So, x_1 of (i) sample and x_2 of $(i+1)$ sample are used to generate the V value that represents an integer digit of $SVD_{j(secured)}$ digits, which is marked as SVD_{jk}, where $k = 1, 2, \ldots, L$. Finding V value from every two continuous samples is repeated until all L integer digits are extracted. The extracted L digits are concatenated in a sequence order to produce $SVD_j = SVD_{j1}, SVD_{j2}, \ldots, SVD_{jL}$. An inverse shifting operation is occurred at all SVD_j values to select the correct arrangement. Additionally, inverse shifting operation is applied at L digits of each SVD_j to find the right ordering of these digits. The result is marked as SVD'_j.

4) Differnece Value (Δ)

After applying steps (2) and (3) at each S'_j, the difference value (Δ) is calculated for each S'_j as in (5).

$$\Delta = |SVD'_j - SVD_j| \tag{5}$$

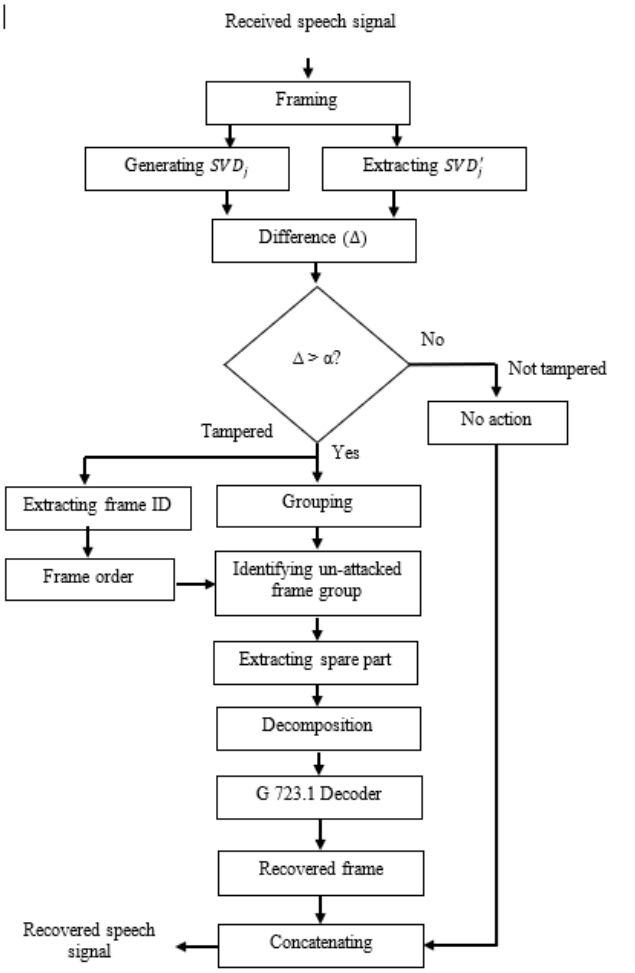

Figure 4. Receiver part of the tamper detection and recovery system

The calculated (Δ) value is utilized to identify whether the S'_j frame was attacked or not by comparing it with the threshold value (α).

5) No Action

When $\Delta < \alpha$, no operation must be performed, indicating that the existing frame is undamaged.

6) Action Required

When $\Delta > \alpha$, the current frame is attacked. Therefore, a decision must be taken, and some operations must be performed to retrieve this frame, as illustrated in the following steps:

a) *Extracting frame ID*

After identifying h_1 samples within S'_j depending on Thr value, they are subjected to (2) to find frame ID digits. The frame ID is composed of three digits, where each digit requires two digits of h_2 sample at specified locations by x_1 and x_2 to calculate the V value, which represents the value of the frame

ID digit. The other two digits are calculated by the same mechanism. After calculating all three digits, they are concatenated to produce the j frame ID, denoted by Y_j.

b) *Grouping*
The $S_j^{'}$ frames are grouped into Q frame groups, each having W frames. So, the number of samples in each group is represented by Z equal to (PW) samples.

c) *Frames order*
Y_j is utilized to find the order of j frame with the frames sequence and frame order having a big benefit to identifying attacked locations within the groups.

d) *Identifying unattacked Frame Groups*
The frame order is highly important for appointing the location of the attacked frame. This importance is clarified in any set of Q groups affiliated with the attacked frame. It is known that D groups of Q contain the spare parts. So, if the attacked frame belongs to one of these D groups, extracting the spare part from the attacked group will be difficult because some watermark data may be lost. So, the attacked frame cannot be retrieved at this state.

e) Extracting Spare Part
After appointing the intact D frame groups, the spare part can be extracted from one of these groups as explained in the following steps:
- For the intact group, h_2 samples of each frame of W frames are found according to the Thr value.
- Each sample of the spare part is composed of three integer digits, requiring three samples to extract it.
- Each digit of the spare apart sample required two digits of an h_2 sample at specified locations determined by x_1 and x_2 according to (2). The calculated V value represents the first digit of the spare part sample with the same step, and the other two digits are extracted. After extracting the three digits, they are combined to produce the first sample value of the spare part.
- The third step is repeated until all spare part samples are extracted. The total size of the spare part is given by $((p/10)+1)\times N$ samples.

f) *Decomposition*
In this step, the extracted spare part decayed into N non-overlapped frames, each with $(p/10)+1$ sample. This step has the benefit of retrieving only the attacked frame. S_{3jID} denotes each spare part frame, where it can be separated from the next spare part frame with the assistance of the first sample, which represents the frame ID. The frame ID is repeated every $(p/10)+1$ sample, which indicates the end of the previous spare part frame and the beginning of the current spare part frame. The decomposition process is highly interested because it makes the smoothest retrieve operation for attacked parts.

g) *Decompression*
After defining the attacked frame and decomposing the spare part into N frames equal to the number of $S_j^{'}$ frames, the spare part of the attacked frame is divided into two parts. The first is the frame ID, represented by the first sample, and the other is represented by the frame compressed data of length $p/10$ samples and denoted by $S_{3_j}^{'}$. To generate the original frame data, $S_{3_j}^{'}$ passes through the G 723.1 decoder.

h) *Recovered Frame*
The decompressed data are placed instead of the attacked frame to represent the recovered part and concatenated with the intact frames to produce the recovered signal.

4. Results and Discussion

In the proposed scheme, the speech files entered as input signals have limited specifications of a 16 kHz sampling rate with a resolution of 32 bits/sample. To accurately estimate the proposed scheme under different conditions, 100 speech signals are recorded by the MATLAB recording Toolbox and employed at different specifications (Gender: Male (40 files), Female (60 files), Language: English, Arabic, length of files from 10 sec to 60sec, mono channels). Additionally, 2000 (Male 800, Female 1200 files) speech samples of the LJ speech dataset [25] are also used for testing. A speech signal is down-sampled from 22 kHz to 16 kHz to fit the previously mentioned system conditions for speech file acceptance. To test and estimate the system performance in tamper detection and localization, a speech file of 11.7 sec is used under different attacks, such as muting and replacement (insertion) attacks. The embedding stage characteristics of using the speech file are: M=187200, N=26, P=7200, frame-index value={111, 112, ..., 136}, n=3, $j \in \{1,2,\ldots,26\}, W \in \{3,4,5,6,7\}$, p=360, q=361 , $Q = N/W$, $Z = P \times W$, h_1=3, h_2=10 , L=10, x_1= 3, x_2=4 , D = 3. Some parameters are tested to assess system capability in detecting, identifying, and retrieving attacked frames. These parameters are:

a) Log spectrum distortion (LSD) measures the distortion in both watermarked and recovered signals [26].
b) Signal-to-noise ratio (SNR) [27].
c) Objective Difference Grade (ODG) is used to evaluate the quality of the recovered signal [28].
d) Perceptual evolution of sound quality (PESQ) is used to measure the inaudibility of the embedded watermark data and the quality of the watermarked and recovered signals [29].
e) Normalized correlation (NC) is used to measure the similarity between the watermarked signal (WS) and cover signal (CS) [30].
f) Bit Error Rate (BER) estimates the amount of error bits that occurred after embedding watermark data in the original signal [31].

The LSD and SNR are calculated using (6) [26] and (7) [27], respectively.

$$LSD = \sqrt{\frac{1}{G} \times \sum_{z=1}^{G} \left(10 \times \log_{10} \frac{|A_z|^2}{|B_z|^2}\right)^2} \tag{6}$$

$$SNR = 10 \times \log_{10} \frac{\sum_{t=1}^{m} CS_t^2}{\sum_{t=1}^{m} (WS_t - CS_t)^2} \tag{7}$$

where z is the frame index, G is the number of frames, A_z and B_z are the Fourier spectra of z-th frame in the original signal and watermarked signal, and m is the samples of the cover (CS) and the watermarked (WS) signals. The limited value of LSD is chosen to be 1 dB for less distortion. BER is calculated using (8), representing the number of error bits after watermarking the cover signal. It has values between 0 and 1, and if the BER value gradually decreases to 0, it makes the watermarked signal less affected by error and distortion [31].

$$BER = \frac{1}{E} \times \sum_{t=1}^{m} \{ {}^{1,\quad WS_t \neq CS_t}_{0,\quad WS_t = CS_t} \} \times 100\% \tag{8}$$

Where E is the total number of bits of WS and CS. The NC is calculated as in (9) [30].

$$NC = \frac{\sum_{t=1}^{m} CS_t \times WS_t}{\sqrt{\sum_{t=1}^{m} CS_t^2} \times \sqrt{\sum_{t=1}^{m} WS_t^2}} \tag{9}$$

PESQ is one of the ITU_T standards used to test the quality of the recovered signal extracted from the tampered signal. Corresponding to mean opinion score values from 1 (very annoying) to 5 (imperceptible), choosing a value of 3 (slightly annoying) is a minimum or acceptable value of PESQ. The higher value of PESQ means that the sound quality is the best.

A. *Quality Test*

After embedding authentication data, synchronization code and recovery data, the quality of the watermarked and recovered signals with cover one are evaluated under different numbers of frame groups. The best quality is obtained for a speech file when PESQ = 5 and LSD < 0.5. Fig. 5 shows the performance of the watermarked signal under continuous attacks. In addition, the slight changes have been accurately demonstrated.

Fig. 5 (a) shows that the PESQ value used to measure the signal quality is changed slightly in a small value concerning the critical value for acceptable quality, illustrated in Fig. 5 (b). In Fig. 5 (c), there is a small variation in LSD values with increasing in N, which is not sensible, but it can be seen in Fig. 5 (d). So, the variation of PESQ and LSD can be neglected and considered a constant value. Tables I and II display the measured parameter for both recorded and speech dataset, demonstrating the effect of embedding the secret data and the number of frames within each frame group on the watermarked signal quality.

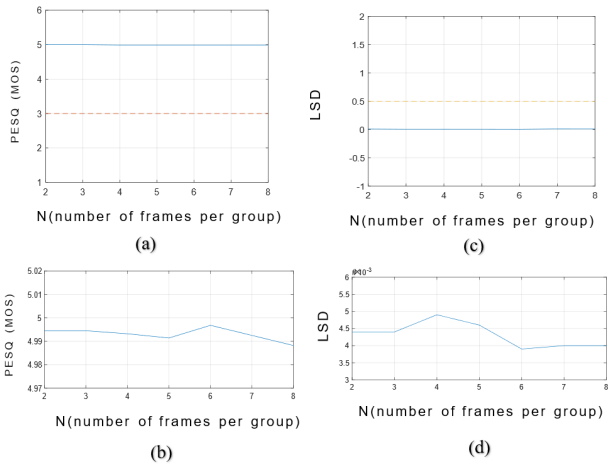

Figure 5. Inaudibility evaluating depending on critical PESQ and LSD values. (a) quality estimation by PESQ measurement, (b) clarification of the sensitive variation in PESQ values, (c) distortion estimating by LSD measurements, (d) clarification of the sensitive variation in LSD values

TABLE I. Quality test of the watermarked signal for the 2000 speech signal of LJ speech dataset

W	PESQ	LSD	SNR(dB)	NC	Embedding capacity(%)
2	4.9889	0.0027	65.0548	1	7
3	4.9889	0.0027	65.0548	1	7
4	4.9872	0.0034	65.4461	1	7
5	4.9882	0.003	64.7172	1	7
6	4.9978	0.0011	78.9595	1	7
7	4.9863	0.0008	78.9595	1	7
8	4.9863	0.0008	78.9595	1	7

TABLE II. Quality test of the watermarked signal for the 100 recorded signals

W	PESQ	LSD	SNR(dB)	NC	Embedding capacity(%)
2	5	0.0061	81.73	1	7
3	5	0.0061	81.74	1	7
4	4.9993	0.0063	81.74	1	7
5	4.9945	0.0063	66.77	1	7
6	4.9957	0.0068	66.76	1	7
7	4.9987	0.0072	66.79	1	7
8	4.99	0.0072	66.79	1	7

Tables I and II demonstrate several parameters for evaluating the quality and distortion produced in the original

speech signal under the proposed watermarking algorithm. The tabulated parameters clarify that a very low distortion occurred in the speech signal quality measured by LSD and SNR. PESQ and NC values under different frame groups indicate that the embedded secret data percentage is very low compared to the data of the watermarked signal. The calculated parameters under the available embedding capacity confirm that the embedded secret data is inaudible and imperceptible by an external entity.

B. Robustness test

Several signal-processing operations are applied separately to identify the effeteness and the amount of error that can occur under each attack to examine the robustness of the embedded watermark data. As a result, the amount of the error that occurred depends on the attack type and the detection accuracy, which can be measured to appoint the error bits after attacking. The obtained average BER and NC measurements are tabulated in Table III.

TABLE III. Quality test of the watermarked signal for the 2000 speech signal of LJ speech dataset

Attack type	Attack amount	Average BER%	NC
Additive White Gaussian Noise (AWGN)	15 dB	0.26	0.8
	20 dB	0.011	1
	50 dB	0.0015	1
Low Pass Filter	8 KHZ	0.0014	1
	6 KHZ	0.19	0.8225
Echo adding	0.1 sec	0.21	0.7853
Zero samples removing	—-	0	1

C. Performance of the tamper recovery system

This section demonstrates the performance evaluation of the recovered signal quality under different attacks, such as muting and discordant replacement attacks. The main aim is to estimate the recovered signal quality when the number of frames within the frame group is varied between 3 and 7 groups for the input signal at the receiver side.

1) Recovery performance under continuous muting attack

Many attacks are geared towards the secret information in the main speech file parts. So, manipulating the primary message data led to an abnormality within the data construction and losing secret data parts. The long-duration continuous attack has the same effect on the hidden important data in which large amounts of data containing the embedded watermark (more than 10% of the speech length has been lost or deleted) have been infected, and the contained data have been lost. The lost data cannot be returned due to the attack size concerning the original message length.

As previously explained, the recovery data are embedded at a location where other signal parts are embedded at different locations to provide a good extraction and restoration of the attacked parts. A long-duration continuous attack is applied to the watermarked signal to examine the proposed method of recovery performance testing. In our test, the TR of 10% to 80% of the watermarked speech is applied and generated at W= 3, 4, 5, 6, and 7. The recovery performance of the proposed system is estimated by measuring RR, PESQ, and LSD. Tables IV and V show the experimental results regarding the RR, PESQ, and LSD values of the recovered signal concerning the original speech for 10 and 100 speech files, respectively.

It is clear from Tables IV and V that the proposed system can fully retrieve the original signal (100%) at any TR with preserving the PESQ at a fixed value (PESQ = 3.94), which represents a good quality of the recovered signal. This value can be considered a drawback of our work because it stays at a constant value. The calculated PESQ value is above the limit representing the acceptable signal quality value (PESQ > 3). For the LSD values, it is evident that the degradation in the recovery signal is gradually increased when the TR rises above 30%. When the TR reaches 80% of the original signal size, the LSD value surpasses a maximum value of 0.2252. Therefore, the maximum measured LSD value is less than 0.5, which representing the critical LSD value. Thus, one can say that the recovered signal has minimal degradation from the original signal, so this degradation can be neglected in the tested range of TR and N values.

Fig. 6 shows an example of speech recovery performance, where 80% of the speech signal was tempered by continuous attack and W=7

Fig. 6 displays the signals about the processing operations to retrieve the original signal after 80% of the muting attack. Fig. 6 (a) represents the original signal entered the embedding algorithm as an input signal. The compressed signal is generated by digitizing the input signal by a lossy coder, as shown in Fig. 6 (b). The watermarked signal produced by embedding the compressed signal at different locations within the cover signal is shown in Fig. 6 (c). Fig. 6 (d) demonstrates the effect of the muting attack on the watermarked speech signal, which leads to losing some of the transmitted message data. Fig. 6 (e) exhibits the efficiency of the proposed algorithm in detecting the attacked regions of the tampered signal. Lastly, Fig. 6 (f) illustrates the original signal after the entire recovery.

2) Recovery performance under replacement attack

To assess the recovery performance of the proposed method, 32 bits/sample and 16 kHz sampled speech signals of a fixed size of 11.7 sec are employed. The speech files are subjected to a content replacement attack using the same dataset. The subjected attack is randomly generated with different lengths at different locations within the speech signal. A set of K replaced samples (group of samples from the watermarked signal) is superseded by m replacement

TABLE IV. Recovered signal quality of the proposed method under continuous attack of 100 recorded speech files

		TR%							
	W	10	20	30	40	50	60	70	80
RR%	3	100	100	100	100	100	100	100	100
	4	100	100	100	100	100	100	100	100
	5	100	100	100	100	100	100	100	100
	6	100	100	100	100	100	100	100	100
	7	100	100	100	100	100	100	100	100
PESQ	3	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	4	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	5	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	6	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	7	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
LSD	3	0.04	0.063	0.105	0.117	0.144	0.137	0.222	0.225
	4	0.04	0.063	0.105	0.117	0.144	0.137	0.222	0.225
	5	0.04	0.063	0.105	0.117	0.144	0.137	0.222	0.225
	6	0.039	0.063	0.104	0.117	0.143	0.137	0.222	0.225
	7	0.04	0.063	0.104	0.117	0.143	0.137	0.222	0.225

TABLE V. Recovered signal quality of the proposed method under continuous attack for 2000 speech signal of LJ speech dataset

		TR%							
	W	10	20	30	40	50	60	70	80
RR%	3	100	100	100	100	100	100	100	100
	4	100	100	100	100	100	100	100	100
	5	100	100	100	100	100	100	100	100
	6	100	100	100	100	100	100	100	100
	7	100	100	100	100	100	100	100	100
PESQ	3	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	4	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	5	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	6	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
	7	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
LSD	3	0.04	0.063	0.105	0.117	0.144	0.137	0.222	0.225
	4	0.04	0.063	0.105	0.117	0.144	0.137	0.222	0.225
	5	0.04	0.063	0.105	0.117	0.144	0.137	0.222	0.225
	6	0.039	0.063	0.104	0.117	0.143	0.137	0.222	0.225
	7	0.04	0.063	0.104	0.117	0.143	0.137	0.222	0.225

samples (external samples inserted instead of K-th the replaced samples). The length of the tampered speech signal is changed due to the length of m replacement samples, which has three lengths: smaller, equal, and larger concerning n replaced samples size. In the smaller size, the tampered signal length is decreased, whereas when the larger size of the replacement attack is inserted, it increases the original signal length. In both situations, the correct length of the original signal after tampering can be known by extracting the overall recovery data of all frames from the untampered frames group. Fig. 7 shows the recovery performance under replacement attack when W=7.

Fig. 7 displays the processing structure for retrieving an attacked speech signal by different lengths of replacement attack (discordant replacement attack). Fig. 7 (a), Fig. 7 (b) and Fig. 7 (c) represent the input speech signal, the watermarked signal after the embedding process and the compressed signal after minimizing the original signal, respectively. Fig. 7 (d) illustrates the attacked location of the watermarked signal. Fig. 7 (e) and Fig. 7 (f) demonstrate the replaced samples of the watermarked signal under small and large size replacement attacks, respectively.

D. Comparison with other work

A comparison with other related approaches is presented in Table VI for evaluating the proposed system performance under different measured parameters.

From Table VI, one can deduce that the methods proposed in [19][21][32] cannot retrieve the tampered location of the attacked signal of small and large sizes.

TABLE VI. Comparison between different approaches regarding the RR and the resistivity

parameter	Attack type	Attack size	[19]	[20]	[21]	[22]	[32]	This Work
RR%	Replacement attack. Replacement samples size with respect to replaced samples	Small size	NA	20	NA	73	NA	100
		Equal size	60	20	15	90	20	100
		Large size	NA	20	NA	93	NA	100
Resistivity	Deletion (cropping),insertion, substitution attacks on beginning,middle, end	15%	NA	NA	Pass	NA	NA	Pass
		25%	NA	NA	Fail	NA	NA	Pass

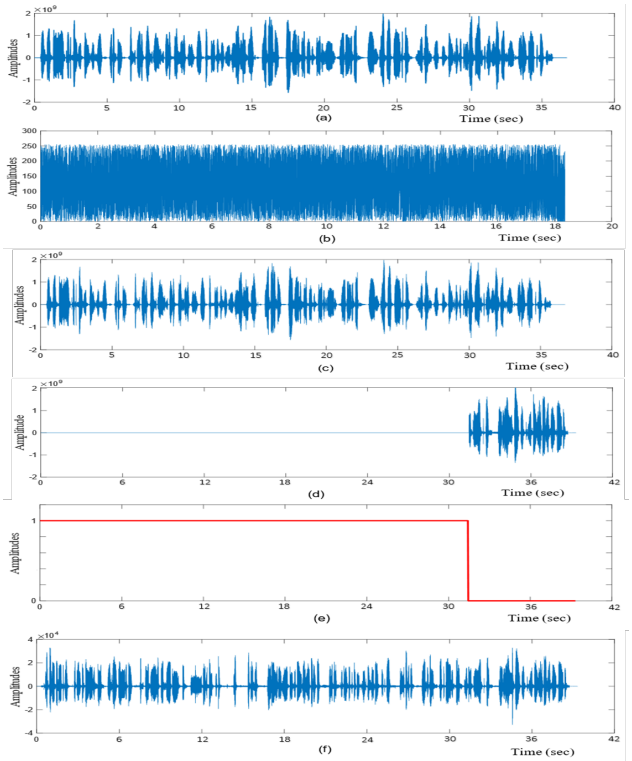

Figure 6. Signals of retrieving process under muting attack (a) original signal, (b) compressed speech, (c) watermarked signal, (d) tampered signal, (e) tampering detection, and (f) the recovered signal

Additionally, the RR of the substituted parts is relatively low, where the non-overriding is not exceeding 20%. The work introduced in [20] can retrieve the attacked speech signal under different length replacement attacks, but it has an RR not exceeding 20%. In contrast, the research presented in [22] shows a maximum RR of 93% for a large attack size. However, the proposed method can recover the tampered speech signal by replacement attack, which means high speech quality with low distortion occurrence of the retrieved speech signal. In comparison, Table VI shows that the proposed method has superior properties to other approaches due to its RR, up to 100% of the attacked parts.

Table VII compares the proposed system to other work regarding the watermark data inaudibility and the ability to detect attacks.

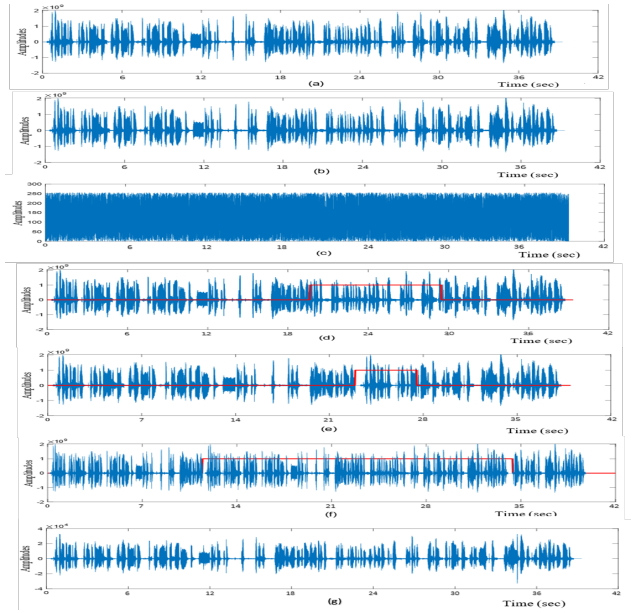

Figure 7. Signals of retrieving process under replacement attack (a) original signal, (b) watermarked speech, (c) compressed signal, (d) detected equal length replacement attack, (e) detected small length replacement attack, (f) detected large length replacement attack, and (g) the recovered speech signal.

It is clear from Table VII that the proposed work achieves the highest value of inaudibility of the embedded watermark data, which is about 72.53 dB for SNR. Besides, the ODG value of the recovered signal is the best among the others, which is about -0.01, such that the quality of the recovered signal is acceptable, with low distortion occurring in the watermarked signal. Therefore, the proposed system has high imperceptibility and undetectable embedded watermark data over the others. Additionally, Table VII states the type of attacks detected by each approach, where the related work can resist the replacement attack with a length equal to the length of the attacked signal parts. This means that when the inserted or substituted parts exceed the replaced samples, they show low resistance to this attack. Unlike the others, the proposed method can resist replacement attacks of different sizes.

TABLE VII. Comparison between different approaches regarding the inaudibility and attacks detection

Ref. No.	Inaudibility SNR (dB)	Inaudibility ODG	Ability to detect a typical attack
[14]	45.36	NA	Mute; Substitution; Insertion; Deletion
[18]	NA	-0.63	Substitution; Insertion; Deletion
[10]	66.43	-0.05	Substitution; Insertion; Deletion
[16]	53.27	0	Mute; Substitution; Insertion; Deletion
[11]	60.83	NA	Substitution; Insertion; Deletion
This Work	72.53	-0.01	Mute; discordant replacement (insertion and Substitution)

5. Conclusion

An efficient scheme has been introduced to detect speech tampers and recover the original signal using data embedding. The SVD generates the authentication codes due to its sensitivity to subtle changes in input data. Besides, G.723.1 speech CODEC is used to generate the spare parts due to its low bit rate output and is suitable for real-time coding systems. The proposed embedding method in this work is simple and very fast. Moreover, it is robust and based on the energy of the cover sample to maintain both the watermarked signal quality and the immunity of the embedded data to different types of attacks. This study also takes into consideration the synchronization issues. Therefore, the proposed approach generates a unique synchronization identifier for each frame to enable the receiver to recover the correct frame order for the attacked speech signals. The experimental results prove that the proposed scheme can retrieve the original speech signal exposed to replacement and insertion attacks at TR up to 80%, achieving about 100% RR under different attack cases. The embedding system can balance the robustness against AWGN and available embedding capacity by preserving the signal quality at an acceptable measured value by PESQ = 3.94 and imperceptibility estimated by average SNR of 72.53 dB. The embedding capacity reaches up to 3% of the signal size under the above conditions. It should be noted that the proposed method has a limitation when it comes to achieving a high PESQ score. To get a high-quality output with a PESQ score greater than 3.94, it may be worth considering combining different compression methods, which can enhance the recovered signal. For future work, the proposed system could be developed using high-speed hardware components to minimize processing time and enable real-time operation.

References

[1] M. Swain and D. Swain, "An effective watermarking technique using btc and svd for image authentication and quality recovery," *Integration*, vol. 83, pp. 12–23, 2022.

[2] R. Sripradha and K. Deepa, "A new fragile image-in-audio watermarking scheme for tamper detection," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, 2020, pp. 767–773.

[3] X. Chen, W. Yuan, S. Wang, C. Wang, and L. Wang, "Speech watermarking for tampering detection based on modifications to lsfs," *Mathematical Problems in Engineering*, vol. 2019, pp. 1–10, 2019.

[4] Z. Lv, Y. Hu, C.-T. Li, and B.-b. Liu, "Audio forensic authentication based on mocc between enf and reference signals," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*. IEEE, 2013, pp. 427–431.

[5] B. B. Haghighi, A. H. Taherinia, and A. Harati, "Trlh: Fragile and blind dual watermarking for image tamper detection and self-recovery based on lifting wavelet transform and halftoning technique," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 49–64, 2018.

[6] Z. Liu and H. Wang, "A novel speech content authentication algorithm based on bessel–fourier moments," *Digital Signal Processing*, vol. 24, pp. 197–208, 2014.

[7] Q. Qian and Y. Cui, "A fragile watermarking algorithm for speech authentication by modifying least significant digits," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 2020, pp. 680–684.

[8] C. Shi, H. Wang, Y. Hu, and X. Li, "A novel nmf-based authentication scheme for encrypted speech in cloud computing," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 25 773–25 798, 2021.

[9] C. Qin, P. Ji, X. Zhang, J. Dong, and J. Wang, "Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy," *Signal processing*, vol. 138, pp. 280–293, 2017.

[10] Q. Qian, Y. Cui, H. Wang, and M. Deng, "Repair: fragile watermarking for encrypted speech authentication with recovery ability," *Telecommunication Systems*, vol. 75, pp. 273–289, 2020.

[11] F. Chen, H. He, and H. Wang, "A fragile watermarking scheme for audio detection and recovery," in *2008 Congress on Image and Signal Processing*, vol. 5. IEEE, 2008, pp. 135–138.

[12] J. Li, W. Lu, L. Du, J. Wei, X. Cao, and J. Dang, "A study on detection and recovery of speech signal tampering," in *2016 IEEE Trustcom/BigDataSE/ISPA*. IEEE, 2016, pp. 678–682.

[13] Q. Qian, H. Wang, S. M. Abdullahi, H. Wang, and C. Shi, "Speech authentication and recovery scheme in encrypted domain," in *Digital Forensics and Watermarking: 15th International Workshop, IWDW 2016, Beijing, China, September 17-19, 2016, Revised Selected Papers 15*. Springer, 2017, pp. 46–60.

[14] Q. Qian, H.-X. Wang, Y. Hu, L.-N. Zhou, and J.-F. Li, "A dual

fragile watermarking scheme for speech authentication," *Multimedia Tools and Applications*, vol. 75, pp. 13 431–13 450, 2016.

[15] W. Lu, Z. Chen, L. Li, X. Cao, J. Wei, N. Xiong, J. Li, and J. Dang, "Watermarking based on compressive sensing for digital speech detection and recovery," *Sensors*, vol. 18, no. 7, p. 2390, 2018.

[16] Q. Qian, H. Wang, X. Sun, Y. Cui, H. Wang, and C. Shi, "Speech authentication and content recovery scheme for security communication and storage," *Telecommunication systems*, vol. 67, pp. 635–649, 2018.

[17] Z. Liu, D. Luo, J. Huang, J. Wang, and C. Qi, "Tamper recovery algorithm for digital speech signal based on dwt and dct," *Multimedia Tools and Applications*, vol. 76, pp. 12 481–12 504, 2017.

[18] Z. Liu, F. Zhang, J. Wang, H. Wang, and J. Huang, "Authentication and recovery algorithm for speech signal based on digital watermarking," *Signal Processing*, vol. 123, pp. 157–166, 2016.

[19] A. Menendez-Ortiz, C. Feregrino-Uribe, J. J. Garcia-Hernandez, and Z. J. Guzman-Zavaleta, "Self-recovery scheme for audio restoration after a content replacement attack," *Multimedia Tools and Applications*, vol. 76, pp. 14 197–14 224, 2017.

[20] J. J. Gomez-Ricardez and J. J. Garcia-Hernandez, "An audio self-recovery scheme that is robust to discordant size content replacement attack," in *2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2018, pp. 825–828.

[21] H.-T. Hu and T.-T. Lee, "Hybrid blind audio watermarking for proprietary protection, tamper proofing, and self-recovery," *IEEE Access*, vol. 7, pp. 180 395–180 408, 2019.

[22] J. J. Gomez-Ricardez and J. J. Garcia-Hernandez, "A low distortion audio self-recovery algorithm robust to discordant size content replacement attack," *Computers*, vol. 10, no. 7, p. 87, 2021.

[23] S.-K. Jung, K.-T. Kim, Y.-C. Park, and H.-G. Kang, "A fast adaptive-codebook search algorithm for g. 723.1 speech coder," *IEEE Signal processing letters*, vol. 12, no. 1, pp. 75–78, 2004.

[24] Q. Zhang, Y. Di, Z. Liu, Y. Huang *et al.*, "A robust audio watermarking scheme based on singular value decomposition," 2013.

[25] K. Ito and L. Johnson, "The lj speech dataset," https://keithito.com/LJ-Speech-Dataset/, 2017.

[26] A. Prodeus, "On some features of log-spectral distortion as speech quality measure," *Autom. Softw. Dev. Eng. J*, vol. 1, 2016.

[27] M. Y. Nejad, M. Mosleh, and S. R. Heikalabad, "An lsb-based quantum audio watermarking using msb as arbiter," *International Journal of Theoretical Physics*, vol. 58, no. 11, pp. 3828–3851, 2019.

[28] M. Arnold, "Subjective and objective quality evaluation of watermarked audio tracks," in *Second International Conference on Web Delivering of Music, 2002. WEDELMUSIC 2002. Proceedings.* IEEE, 2002, pp. 161–167.

[29] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)*, vol. 2. IEEE, 2001, pp. 749–752.

[30] N. Chen and J. Zhu, "A multipurpose audio watermarking scheme for copyright protection and content authentication," in *2008 IEEE International Conference on Multimedia and Expo*. IEEE, 2008, pp. 221–224.

[31] G. Breed, "Bit error rate: Fundamental concepts and measurement issues," *High Frequency Electronics*, vol. 2, no. 1, pp. 46–47, 2003.

[32] H.-T. Hu and Y.-H. Lu, "Frame-synchronous blind audio watermarking for tamper proofing and self-recovery," *Adv. Technol. Innov*, vol. 5, pp. 18–32, 2020.

Younis Mohammed Jalil received his B.Sc. and, M.Sc. degrees in Electrical and Electronic Engineering from the University of Kerbala, Iraq, in 2017, and 2022, respectively. He is currently work at the Electrical and Electronic Engineering Department, College of Engineering, University of Kerbala. His significant interests include digital signal and multimedia processing, data security.

Haider Ismael Shahadi received his B.ESc degree in information engineering from the University of Baghdad, Iraq in 2001, his master's degree in Electronic and Communication Engineering from the University of Baghdad-Iraq in 2004, and his Ph.D. in Electronic and Communication Engineering from the Tenaga National University, Malaysia in 2014. Currently, he is a professor at the University of Kerbala, Iraq. His research interests include digital signal and multimedia processing, data security, FPGA design and implementation and embedded systems, IOT systems, and smart systems.

Hameed Rasool Farhan received his B.Sc., M.Sc. and Ph.D. degrees in Electronic Engineering from the University of Technology, Baghdad, Iraq, in 1986, 2011, and 2018, respectively. He is currently an Assist. Prof. at the Electrical and Electronic Engineering Department, College of Engineering, University of Kerbala. His significant interests include Digital Electronics, DSP, Image Processing, Pattern Recognition, and Computer Vision..