Indian Journal of Engineering & Materials Sciences
Vol. 31, April 2024, pp. 258-269
DOI: 10.56042/ijems.v31i2.942

Epoch event based speech watermarking for tamper detection

Rajeev Kumar[a] & Jainath Yadav[b]*

[a]Faculty of Information Technology, Gopal Narayan Singh University, Jamuhar, Bihar 821 305, India
[b]Deapartment of computer science, Central University of South Bihar, Gaya, Bihar 824 236, India

Received: 27 April 2023; Accepted: 15 February 2024

There are big challenges to handle the content authentication and tamper detection in speech watermarking techniques. These challenges can be addressed using the detection of epoch locations, logistic mapping, verification of the tamper locations, and the Min algorithm. Epoch events provide the variable length epoch intervals, and it has been used to find the variable lengths of frame segments in this work. The logistic mapping provides strong security for original signals. They can provide a robust and secure method that addresses copyright issues, illegal intentional, or unintentional modifications, and tampered locations. In this paper, we have implemented an epoch event based speech watermarking technique for tamper detection using the logistic map method. It has been observed that the simulation results demonstrated that the PRE value is approximately more than 82% and the PESQ value lies between excellent and good. It means that there are no significant differences between original and watermarked signals and we have obtained a highly imperceptible, robust, and secure system. Moreover, the overall simulation results of the proposed method provide the better solution as well as more security than existing methods.

Keywords: LP analysis, Speech watermarking, Logistic map, Authentication, Tamper detection, Epoch detection

1 Introduction

In recent decades, the content sharing/ communicating through the Internet has been increased rapidly. In this case, our contents must need to be secure from the outside world. Digital speech watermarking is one of the most popular schemes to protect our content from the real world[1-4]. Traditionally, the digital speech watermarking scheme is partitioned into two parts: the first one is watermark embedding and the second part is the extracting process. The first procedure should protect our materials from unauthorized replication while preserving the integrity of the original signal. The second process is used to recover our watermark data from watermarked signals and later, it is used for evidence purposes.

In the recent trends, watermarking techniques also include validation phase along with the traditional process[5].The validation is related to identifying the original contents. Content authentication and tampered detection are the main reasons to use the validation phase. The watermarking techniques should meet basic requirements such as imperceptibility robustness, capacity,and security[6-10].

The speech watermarking techniques require the algorithm for embedding, extricating and validating the watermark, sampling rate/frequency, selection of watermark location, and parameters for finding the imperceptibility and robustness. In this paper, we have focused on invisible speech watermarking technique for tamper detection and authentication. The working method decides how to embed, where we can insert, how much we can embed, and in which form, *i.e*., time or frequency or in both domains or encrypt the watermark before insertion. The watermarking extraction process is determined based on three different algorithms, i.e., blind, semi-blind, and non-blind methods[6]. These algorithms navigate us in which way we can extract our watermark.

In this paper, we have explored the epoch events for embedding the watermark in the proposed method. An epoch occurs at the time when the vocal-tract system undergoes considerable excitement[11-13]. The prior purpose of epoch event is to provide high security using the variable length of frames instead of fixed length. Secondly, the logistic mapping has been considered for random number generation instead of an image as a watermark. The main benefits of the logistic map over the image are: (i) to reduce the time complexity, (ii) require less memory, (iii) single time processing is required for embedding the watermark to reduce processing time, (iv) flexible in length/size of watermark, due to this reason we can control the

*Corresponding author (E-mail: jainath@cub.ac.in)

capacity of the watermark, and (v) simple process and it provides strong security. The third important contribution is related to the verify_tamper_loc method. This method is very useful to match the tampered epoch intervals of original watermark and extracted watermark sequences. Verify_tamper_loc method is useful to detect the tamper location where some samples are modified after performing the attacks. The fourth important contribution is the Min algorithm. Using this algorithm, we can easily embed and extract the watermark in a systematic way. Also, it improves the security without degrading the quality of watermarked signals. The fifth important role is percentage relative efficiency (PRE). The PRE has been introduced as a quality evaluation parameter for the first time in the watermarking field and it indicates the efficiency of the watermarked signal.

We have discussed the digital watermarking technique based on the authentication process. Numerous models based on spatial, transformation, modulation, and feature extraction techniques have been studied in the literature. Tamper detection using deep neural network for recovery[14].

Yamni[15] propose an efficient speech watermarking technique based on discrete Tchebichef moment transform, the chaotic system coupled map lattices, and DWT. Pavlovic[16] use two trained adversarial deep neural networks (embedder and detector). Baziyad[17] provided copy protection to protect deep neural networks using a robust DCT-based watermarking method. Xiang[18] use a keyword interchange method and a error expansion scheme to protect copyright and sensitive information by developing a watermarking scheme. Kim[19] detect the audio signal's transient's points and protect audio copyright in an online streaming environment. Chen[20] provide a survey of a detailed analysis of attacks and defenses in a deep learning based watermarking scheme. Patil[21] propose a deep convolution neural network-based watermarking method to overcome the low reliability under various attacks. Amrit[22]provide a detailed survey of various watermarking schemes based on emerging technology such as artificial intelligence, and deep learning. It also provides challenges and future research directions for these areas.

Kumar[23] have proposed vowel onset and offset events based speech watermarking. Authentication issues were addressed by the speech watermarking schemes presented[24]. They have also analyzed the performance of various audio and speech watermarking systems. The transformation method and chaotic-based speech watermarking techniques were proposed[25]. They created a secure watermarking system using these techniques, which also allowed them to address the false positive issue. SNR, PESQ, and correlation coefficients were utilized to gauge how well the developed approach worked. Yadav[26] has presented vowel events based method for the detection of vowel transition regions. Kumar[27] have proposed compression and decompression concept based on DWT and DCT methods. The significant modifications in LSB bit approach have implemented[28]. There is an integrity authentication algorithm that uses the perceptual hash function and learnt dictionary method proposed[29]. On the basis of this gamma tone filter model, this procedure was carried out.

Liu[30] described the recapture and de-synchronization attacks that can be prevented using transformation methods, together with a modified patchwork method. The SNR, BER, and ODG metrics were used to assess the suggested method's robustness. For the copyright protection and forensics track they also performed several attacks, which deliver a superior result. Based on Band pass Filter, the work has focused on protecting copyrights and confidential communications[31]. BPF and MP3-compression assaults were found to be resistant to simulation results. They haven't discussed how to make themselves invisible or how to defend against various attacks. The DCT and SVD transformation methods were proposed[32]. When it comes to inserting the watermark, they have also utilised the concept of voiced and unvoiced frames. The watermark is embedded in the low-frequency voiced and unvoiced components of the audio signal. The intended outcome is extremely inconspicuous and stable. When it comes to BER and average information loss in re-sampling attacks, the imperceptible and robust approach is found to be robust and imperceptible.

Revathi[33] have described enhancing the security of speaker authentication using a biometric system. For this, they used the DWT method for the watermark embedding area and the feature selection process is used for authentication. Simulation parameters are based on PSNR, BER, and perceptual evaluation speech quality (PESQ). Agradriya[34] developed a watermarking method to protect the integrity of the audio signal. According to the findings of the simulation, it is able to evaluate all types of music files in different iterations. Despite its low 0.01 intensity, it's incredibly durable. Weina[35] proposed an audio watermarking technique based on the scrambling encryption method. The method is based on

Block Truncation Coding (BTC) and Arnold transform method to embed the watermark information in the appropriate place. Simulation results are based on the PSNR, BER, and NC parameters. Results represents that the proposed method has good imperceptibility, robustness, and security. Watermark embedding techniques have been proposed[36]. Segmentation and the DCT technique are used to identify the low-frequency area for the embedded process. BER, ODG, and SDG are the simulation parameters (SDG). Effective robustness was demonstrated, confirming the integrity and security of the content.

DCT-based frame number and the compressed signal, the watermark information is formed. BER, SDG, and ODG are used to meet the digital requirements of the watermarking techniques[37]. Liu[38] have proposed a dual image watermarking system for authentication and copyright protection. They achieved this technique using arobust and fragile watermarking method. The DWT and quantization methods are used for watermark insertion in YCbCr color space. The simulation results are evaluated based on PSNR and SSIM methods and it shows effective results on various attacks. Deokar[39] have described the transformation based watermarking scheme. Simulation results demonstrate that the original audio signal and the embedded signal are perpetually indistinguishable. It means that there is no more degradation on the original content after embedding. When they apply different types of attacks, the result exhibits a slight degradation in the quality.

Sarreshtedari[40] havedescribed a self recovery watermarking scheme. The self-recovery feature and digital self- embedded voice signals were introduced by them. According to the Tolerable Tampering Rate (TTR), the approach is secure and robust. A speech watermarking methodology using DCT, SVD, and DWT methodologies has been proposed[41]. They assessed the quality of the speech using BER and SNR metrics. DWT-SVD and DCT-SVD procedures are used to calculate and compare the experimental data. Combining all three strategies yielded greater results than a single method. DWT and SVD watermarking concepts have been described[42]. As a result, the LL sub-band was subjected to the SVD technique. A digital picture watermarking approach employing the self-adaptive differential evolution (SDE) algorithm has been proposed[43]. In order to create sub-bands of different frequencies, they used DWT up to the second level on the cover image. Second-level DWT uses SVD for each subb and. Scaling factors have certain limitations in this method.

The dynamic DWT levelling and error correction algorithm for audio signals has been explained[44]. They used hash methods that are more resistant to watermarking attacks on the watermark signal. It was found that putting the watermark at low frequencies made it more resistant to attacks, while embedding it at high frequencies made the watermark more undetectable.

After reviewing the literature, it has been found that speech content authentication and tampered detection are required. The attacker can modify original content and impose his/her contents. The integrity of speech content is also known as speech content authentication. If the mark is not present in the speech signal, then we can say that the speech signals are altered. In this paper, we present the epoch event based speech watermarking for tamper detection using logistic mapping and verify_tamper _loc methods.

2 Materials and Methods

In this section, we have presented the novel idea based on epoch intervals to embed the watermark in the speech signal. The speech signals can be considered as a sequence of events. An event is defined as any significant changes during the production of the speech. The sequence of changes in the shape of the vocal tract and the nature of excitation are reflected as events in the speech signals. The epoch is the moment of the vocal-tract system's considerable stimulation. At the precise moment that the glottis contracts, the most intense stimulation occurs. Therefore, in voiced speech, epoch coincides with the moment of glottal closure, but in unvoiced speech, it coincides with a moment of random excitation[11-13]. Glottal closure (GC)/vocal tract events are useful for accurate estimation of pitch period because it works at the signal level. It estimates the accurate frequency response of the vocal tract system and it is used as a pitch marker.

For the embedding process, the epoch detection, LPA and logistic mapping have been explored. Zero-time windowing (ZTW) and the numerator of the group delay (NGD) function are used to properly locate the epochs in voice signals. Spectral information can be found at each sample location. It is possible to reduce the redundancy in the short-term correlation of nearby data using the LP analysis. It decomposes the speech signal into two independent

components: LP estimated/predicted signal, and the LP residual signal. Based on a linearly weighted collection of preceding samples, the LPA predicts the time-domain speech sample. The main benefits of the LPA method are as follows:

- It removes the redundancy from adjacent samples.
- LPA is based on the all-pole filter.
- To minimize the computational complexity, its components are divided into the time-domain itself.
- It minimizes the prediction error while choosing optimal values for all LP coefficients using the autocorrelation method.
- It provides the stability of a system and it is computationally more efficient.

A logistic map is a non-linear deterministic dynamical system that demonstrates pseudo-random behaviour. The logistic map method is used for the generation of encrypted random numbers for the watermark. The advantages of the logistic map instead of the greyscale image as a watermark are as follows:

- It provides high security due to the random behavior and depending on the settings of specific parameters and initial conditions.
- To reduce the time and space.
- Reduced processing time.
- It is flexible to maintain the length/size of the watermark.

The logistic map is the part of chaotic maps that are used to generate pseudo-random numbers using a mathematical function. It is represented as:

$$I_{n+1} = r.I_n(1 - (I_n)) \quad \ldots (1)$$

Where, I_n lies between 0 and 1. As a constant, r accepts values from 1 to 4. The behaviour of the logistic map is determined and explored by the value of a parameter in themap. The parameter r (> 3.7) is set to a higher value in this study in order to produce a chaotic yet deterministic discrete-time signal. The subsections for embedding, extracting, and verifying have been created within this section.

2.1 Proposed method for embedding the watermark

The proposed method takes input as a speech signal in a '.wav' format. The logistic map method generates pseudo-random numbers as the watermark. These random numbers are generated in encrypted form to make them more robust. In the next step, the embedding process has been used to detect epochs from the speech along with the help of the minimum value selection method. The proposed embedding flow process for speech watermarking technique using the logistic map and LPA-based methods has been depicted in Fig. 1. The steps have been performed for the embedding process and they are discussed in the Algorithm 1.

Algorithm 1: Embedding procedure for speech watermarking

Step 1: Take input as speech signal.
Step 2: Detect epoch location from the speech signal (See Algorithm 2).
Step 3: The LPC method is applied to separate the speech signal into estimated and error signals.
Step 4: Consider only the estimated signal for finding minimum value from each epoch interval.
Step 5: Generate random number using logistic map method for watermark purpose.
Step 6: Apply the Min algorithm (see in section 2-C).
Step 7: After embedding the watermark, recombine the errorestimation part from LPC and reconstruct the new signal, i.e., watermarked speech signal.

The proposed method utilizes NGD function and ZT windowing to detect epochs[11]. Figure 2 shows the steps of epoch detection from the speech signal. Figure 2(a) shows the voiced segments of speech signal, whereas Fig. 2(b) shows the differenced EGG signal. Figure 2(c) shows the dominating spectral energy profile derived from the sum of three significant spectral peaks in HNGD spectra. The epoch evidence signal is represented in Fig. 2(d) after convolving the dominating spectral energy profile with a Gaussian window of 2 ms. The reference epochs in Fig. 2(b) and the detected epochs in Fig. 2(e) are observed to be identical and similar to each other. The detection steps of epoch location are described in Algorithm 2.

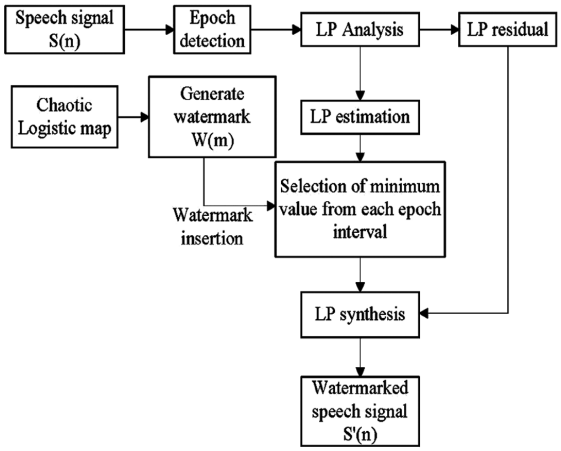

Fig. 1 — The proposed watermark embedding block diagram.

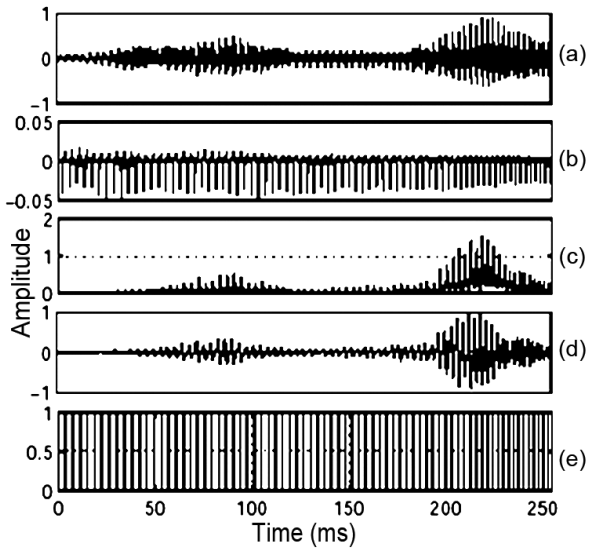

Fig. 2 — The detection of epoch positions in a speech signal (a) Speech segment, (b) Differenced EGG signal, (c) Dominant spectral energy profile generated from HNGD spectra, (d) Epoch evidence signal and (e) Detected epoch locations.

Algorithm 2: Epoch location detection procedure for speech watermarking

Step 1: Input the speech signal.

Step 2: Determine voiced regions using combined evidences from zero crossing rate and short-time energy of the speechsignal.

Step 3: Perform ZTW operation using h_1 [n].

$$h_1[n] = \begin{cases} 0, n = 0 \\ \frac{1}{4\sin^2\frac{\pi n}{M}}, n = 1, 2, .. M-1 \end{cases} \quad ...(2)$$

Step 4: Perform ripple reduction using h2 [n].

$$h_2[n] = \left\{ 4\cos^2 \frac{\pi n}{2M}, n = 0, 1, 2, ... M-1 \right. \quad ...(3)$$

Step 5: Compute numerator of group delay (NGD).

Step 6: Compute double differencing of numerator of group delay (DNGD).

Step 7: Obtain the DNGD Hilbert envelope (HNGD spectrum).

Step 8: The HNGD spectrum has three distinct peaks, the sum of which needs to be calculated.

Step 9: Convolve spectral energy profile with Gaussian filter G(n).

$$G[n] = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-n^2}{2\sigma^2}} \quad \ldots (4)$$

Step 10: Select positive peaks from convolved output and obtain epoch locations.

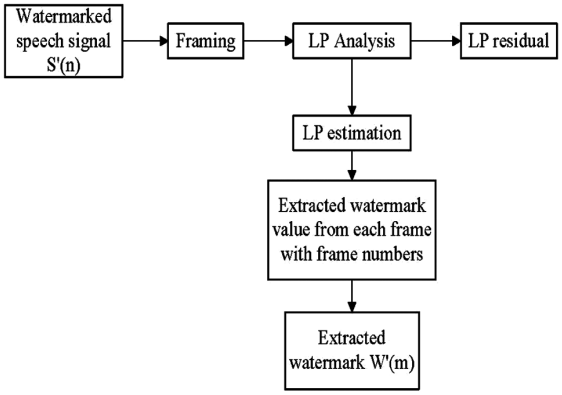

Fig. 3 — Extraction procedure for the proposed watermarking scheme.

2.2 Proposed methods for extracting the watermark

The proposed extraction process of speech watermarking scheme based on the logistic map and LPA methods has been depicted in Fig. 3. The steps of extracting process are defined in Algorithm 3.

Algorithm 3: Extraction procedure for speech watermarking

Step 1: Take input as watermarked speech signal.

Step 2: Detect epoch location from watermarked speech signal (See Algorithm 2).

Step 3: The LPC method is applied for separating the signal into estimated and error signals.

Step 4: Consider only estimated signal for finding minimum value from each epoch interval.

Step 5: Apply extract-Min algorithm (see in section 2-D) and extract watermark.

2.3 Min algorithm for watermark embedding process

Minimum values have been selected from epoch intervals of the speech signal. We have been performed logistic map method to generate the random number for watermark. The steps of Min algorithm for embedding process are depicted in the Algorithm 4.

Algorithm 4: Min algorithm for watermark embedding process.

Step 1: Consider the LPA method to obtain the estimated signal of the original speech signal asgiven below: Z = LPA(epoch intervals of original speech);
mag = estimated signal(Z);

Step 2: Compute minimum values from each epoch interval having estimated signal using LPA, and store them in a variable 'p' with the location as indices.

(p, indices) = min(estimated signal, [], 2);

Step 3: Generate the random numbers using logistic map method. Random numbers are defined as

$x_{n+1} = r.x_n .(1 - (x_n))$ (See the Eq. 1) and they are stored in variable 1.
Step 4: variable 2 = variable 1/mean(mean(variable 1));
Step 5: Embed the minimum value of original matrix from encrypted random numbers using following sub-steps.
Step 5.1: for i = 1 to length of variable
Step 5.2: Z(i, indices(i)) = variable 2(i);
Step 5.3: for loop end
Step 6: Finally, the random numbers are inserted into each epoch interval of the speech signal.

2.4 Min algorithm for watermark extraction process

After insertion of the watermark, the minimum value of epoch locations may be changed in the watermarked speech signal. To handle this problem, we must save the new indices with values of watermarked speech for the extraction process. We have performed an extract-Min algorithm to extract the watermark values. The steps of the extract-Min algorithm for the extraction process are defined in the Algorithm 5.

Algorithm 5: extract-Min algorithm

Step 1: Compute the actual minimum value indices from the watermark data, and then consider the estimated signal from the LPA of watermarked speech.
wd = LPA(epoch intervals of watermarked speech);
wd mag = wd estimated signal(wd);
(wd variable, indices 3) = min(wd mag, [], 2);
Step 2: Find the actual watermark values in the variable 'wd variable'.
Step 3: variable 2 = wd variable/mean(mean(wd variable));

Step 4: Finally, we get the watermark values and store in variable 'variable 2'.

2.5 Proposed method for identifying the tamper location in speech watermarking

The tamper detection process in speech watermarking scheme based on the verify tamper loc method has been depicted in Fig. 4. This method is very useful for comparing the two different equal length sequences and to verify the tamper locations. It is useful to verify or to find thetamperlocations. This method is used to identify the modified epoch intervals between the original (O_w) and extracted watermarks (E_w).

The steps of the verify tamper loc method for the detection of tampered locations in the watermarked speech signal are defined in the Algorithm 6.

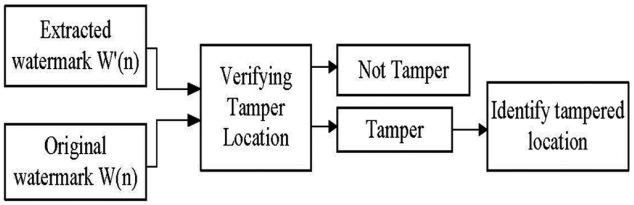

Fig. 4 — Identifying tamper location procedure in the proposed watermarking scheme.

Algorithm 6: Identifying tamper locations

Input: Original watermark values in array Ow and extracted watermark values in array E_w. 'n' is the number of epoch intervals.
Output: Identified tamper locations.

```
verify_tamper_loc(Ow , Ew , n)
do
Declare variable i;
fori ← 0 to n-1 do
ifOw [i] == Ew [i] then
Print(Epoch intervals are not tampered);
End if
End for loop else
fori ← 0 to n-1 do
ifOw [i] != Ew [i] then
{
fori ← 0 to n-1 do
ifOw [i] != Ew[i] then
Print(Tamper location with epoch index interval i,
Original watermark value: Ow[i], and extracted watermark value: Ew [i]);
End inner if
End for loop
End outer if
}
Print (Epoch intervals are tampered);
End of outer for loop
End function
```

2.6 Quality evaluation parameters for speech watermarking technique

In this study, various parameters have been used to simulate the proposed method. The BER, PSNR, MSE, SNR, and NC values are determined by comparing the watermarked signals to the original signals.

- **Peak signal to noise ratio (PSNR)**

The imperceptibility factor measures quality between original (Os) and watermarked speech (Ws) signals using PSNR method in terms of decibels[23,28].

$$PSNR = 10 * \log_{10} \frac{maximum^2}{MSE} \quad \ldots (5)$$

- **Signal to noise ratio (SNR)**

The metric of quality according to the formula Eq. 6, SNR is calculated by dividing the difference between the watermarked and original speech signals by the square of the watermarked signal.

$$SNR(dB) = 10\log10\frac{\sum_{i=1}^{N}[W_i]^2}{\sum_{i=1}^{N}[W_i - W_i]^2} \quad \ldots (6)$$

Normalizedcorrelations (NC) have been evaluated to verify the robustness of the signal[45].

- **Bit Error Rate (BER)**

The BER is defined as in Eq. 7.

$$BER = \frac{1}{N_1}\sum_{i_1}^{N_1}[X_1(i_1) \oplus Y_1(i_1)] \quad \ldots (7)$$

Where $\oplus$ is the XOR operator, X_1 (i_1) represents the original signal, Y_1 (i_1) is the extracted watermark signal, and N_1 is the number of embedded watermark values.

- **Percentage relative efficiency (PRE)**

The PRE calculation is introduced to check the efficiency in the watermarking techniques. The PRE is based on mean and variance of original and watermarked speech signals and defined in Eqs 8-12.

$$M_O = \frac{\sum_{i=1}^{N}(O)_i}{N} \quad \ldots (8)$$

$$M = \frac{\sum_{i=1}^{N}(W)_i}{N} \quad \ldots (9)$$

Where, N represents the length of original speech. The original and watermarked speech signals are denoted as O and W, respectively.

$$V_O = \frac{\sum_{i=1}^{N}(O - M_O)^2}{N} \quad \ldots (10)$$

$$V_W = \frac{\sum_{i=1}^{N}(W - M_W)^2}{N} \quad \ldots (11)$$

$$PRE = \frac{V_W}{V_O} \quad \ldots (12)$$

Where, M_O and V_O represent the mean and variance of the original speech, respectively. M_W and V_W represent the mean and variance for watermarked speech signals, respectively.

- **PESQ measure**

The PESQ is computed between original and reconstructed signals on the basis of mean opinion score (MOS). The PESQ grade scale is decided by 25 expert listeners based on signal quality. In this study, we have considered 25 expert listeners to judge the signal quality using the PESQ grade scale as given in Table 1.

Table 1 —The PESQ measure between original and watermarked speech signals

PESQ scale	Range of quality	PESQ analysis
5	Excellent	Speech which is audible and similar
4	Good	Audible but with slightly lesser quality
3	Fair	Audible, but of varying quality
2	Poor	Although audible, the quality is vastly different
1	Bad	The voice is indistinct and of extremely different quality

3 Results & Discussion

In this study, dataset having 215 male and 185 female speakers from the TIMIT database have been considered for the evaluation of the proposed and the existing methods. The sampling frequency is 16 kHz, and the resolution is 16 bits for each of the 400 speech signals that are recorded. Simulation of the dataset for watermarking is carried out in MATLAB. The details of the simulation results are given in the following two sub-sections.

3.1 Performance evaluation

We have considered the 400 speech signals from the TIMIT database and watermark as pseudo-random numbers. The watermark data is generated by the logistic map method for the watermarking procedure. In this work, we have developed the proposed watermarking scheme based on the epoch detection, verify_tamper_loc, and LPA methods.

For the evaluation of the proposed method, we have computed PSNR, MSE, Percentage relative efficiency (PRE), and NC values between original speech and watermarked speech signals, and watermark and extracted watermark signals, respectively as shown in Table 2.

According to Eq. 5, the bond between PSNR and MSE values is inversely related to each other, as demonstrated in Eq. 13.

$$\text{PSNR} \propto \frac{1}{\text{MSE}} \quad \ldots (13)$$

Similarly, the average percentage PRE values are evaluated to find the efficiency of the proposed technique. Table 2 shows that the PRE values of the proposed procedure are approximately more than 82 % efficient. From Table 2, we observed that the proposed method based on epoch detection and the LPA-based techniques gives better results with the help of the logistic mapping method. The proposed method uses epoch detection to find the variable size of frames which provides more security. Secondly, the LPA method is used to reduce the prediction error and it also uses the logistic mapping method to generate pseudo-random behavior of the watermark. Due to this reason, it provides effective results compared to existing methods.

Table 2 compares the average PSNR and NC values from the TIMIT database to original and watermarked speech signals. The PSNR readings represent the signal imperceptibility in decibels. Table 2 shows that the quality of the watermarked speech is higher than 82 percent. The NC values for each dataset are equal to 100 %, which means both signals are same. On the basis of both PSNR and NC measurements, we have found that the imperceptibility, resilience, and security of watermarked speech signals are high.

Table 3 shows the average comparison between the proposed and different existing watermarking methods in terms of SNR and PESQ for watermarked signals. Table 3 is divided into three columns. The first column consists of existing and the proposed methods. The second column shows the average SNR values. The last column represents the PESQ values, which are measured by the 25 expert listeners. We have observed that the average PESQ values of the proposed method lie between excellent and good (see Table 1) range of the quality. Average SNR and PESQ values of the proposed method are higher than existing methods.

Table 4 represents the results of various attacks on watermarked speech signals in terms of BER. Table 4 consists of 6 columns. Columns 2-5 represent the result of different existing methods. The last column describes the result of the proposed method. Zarean[48] implemented speech watermarking method in the dark domain to find the best band for watermark insertion, due to this reason it has more bit errors compared to other methods. Wang[46] evaluated the BER values based on FFT, miu-law and quantisation methods. Approximately, for all the cases the bit error is high compared with the other existing methods. Shi[29] successfully detected and localised the content of voice signals, as well as embedded the watermark in each segment of speech. In the instance of echo addition, we observed that the high sampling frequency had an effect on voice quality. Hu[46]have chosen low frequency region for watermark insertion, due to this reason the BER values are '0' in all the cases except echo addition. In this case, it may occur due to the delay of signals. We can concluded that the

Table 2 — Quality evaluation using average of PSNR, PRE, and NC values.

Classification of the signals	Epoch intervals, chaotic logistic map and LPA methods based average values			
	PSNR (dB)	MSE	PRE (%)	NC
Watermarked speech	45.2135	0.000029	82.53	1.0000
Extracted watermark	84.1023	0.000012	92.31	1.0000

Table 3 — Result comparison between existing and proposed methods based on Avg. values of SNR and PESQ for watermarked speech.

Watermarking Method	Avg. SNR(dB)	Avg. PESQ
FFT-PVNM[46]	19.224	3.713
DCT-QIM-SWT[29]	24.51	4.4
DWT-MFCC[33]	25	4.47
DCT[36]	25.83	4.5
Self correlation method[47]	22.3	4.2
FFT-DCT[49]	15.7542	3.2
DCT[38]	22.18	4.1
G.723 and MSD5[40]	24.74	4.45
FFT and μ-law QIM[48]	20	4
Proposed	28.2	4.58

Table 4 — Result comparison between existing and proposed methods using BER(%) after performing various attacks on watermarked speech signals.

Attack	BER(%)				
	Based on FFT and miu-law QIM[48]	Based on self correlation method[47]	Based on DCT-QIM and SWT[29]	Based on FFT and PVNM[46]	Proposed
Lowpass (14 kHz)	0.01	0.2651	0	0	0
Highpass (50 Hz)	0.04	0.2432	0	0	0
Requantization (16/8/16 bits)	0.17	0.3435	0	0	0
Resampling (16/8/16 kHz)	0.00	0.3954	0	0	0
Echo addition	9.12	-	1	0.6	0.48

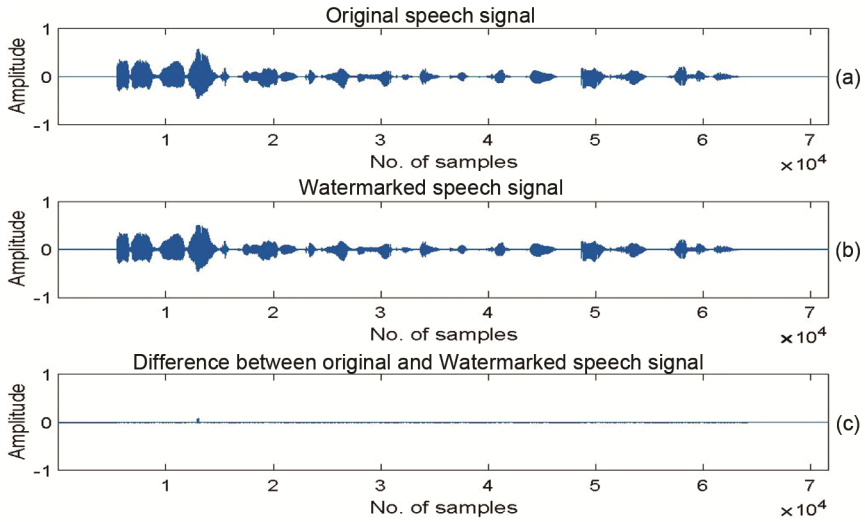

Fig. 5 — Waveforms of (a) Original speech, (b) Watermarked speech and (c) the differences between amplitudes of the original and watermarked speech signals.

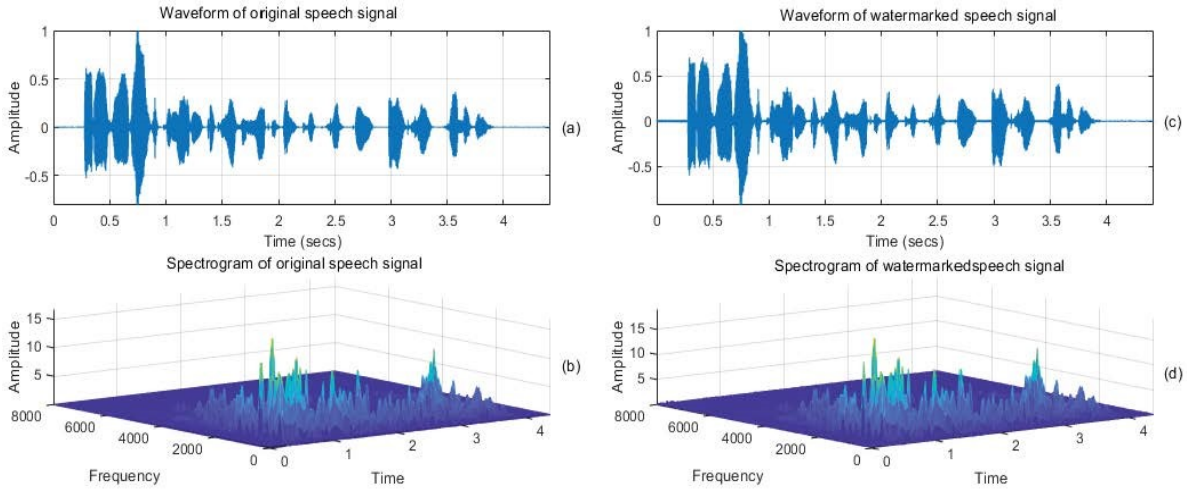

Fig. 6 — Waveform and spectrogram of (a) cover speech signal, (b) cover speech signal's spectrogram, (c) watermarked speech signals and (d) watermarked speech signal's spectrogram.

proposed method provides better results compare to the existing methods.

Figure 5 shows the waveform of original [Fig. 5(a)], and watermarked signals [Fig. 5(b)]. The difference between the original and watermarked versions of the speech signal is depicted in Fig. 5(c). Generally, similarity between cover and watermarked signals is depicted through waveform and spectrograms. Figure 6 (a) shows waveforms of original speech signal, and Fig. 6 (c) represents waveforms of watermarked speech signal. Similarly, Fig. 6 (b) represent the spectrogram of the original speech signal, and Fig. 6 (d) shows the spectrogram of the watermarked speech signal. It is observed that the waveform and spectrograms are similar to each other.

3.2 The detection of tampered location using proposed method

The illegitimate users perform the integrity attacks on watermarked signals for modifying the location of watermark. Due to this reason, incorrect watermark signal is detected. Detection of tampered locations increases the ability for verifying the authenticity. It improves the embedding capacity and security of speech watermarking techniques. The verify tamper loc method is used to apply for identifying the tampered epoch intervals. In this paper, we have randomly selected one watermarked signal to implement four different types of integrity attacks such as the mute, substitution, insertion, and deletion. In this sub-section, we have performed mute,

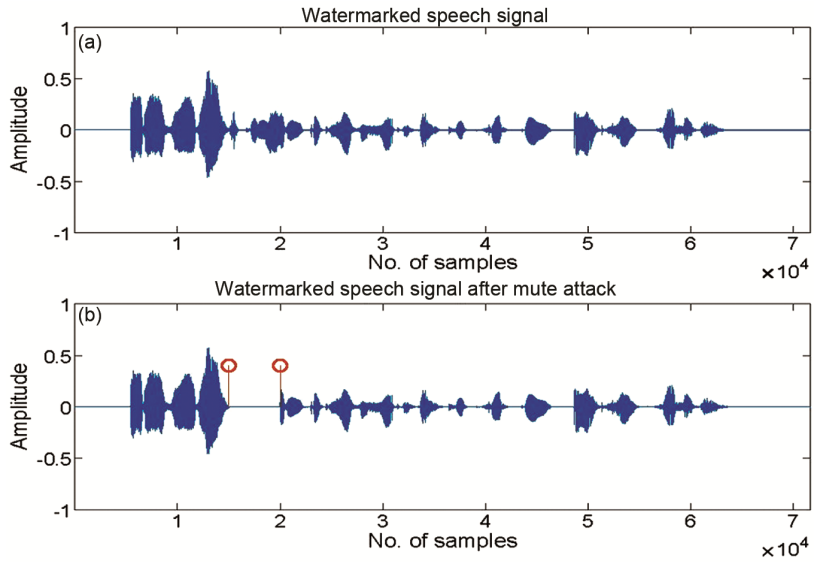

Fig. 7 — Tampered location in mute attack (a) watermarked speech signal and (b) the muted speech samples between marked portions.

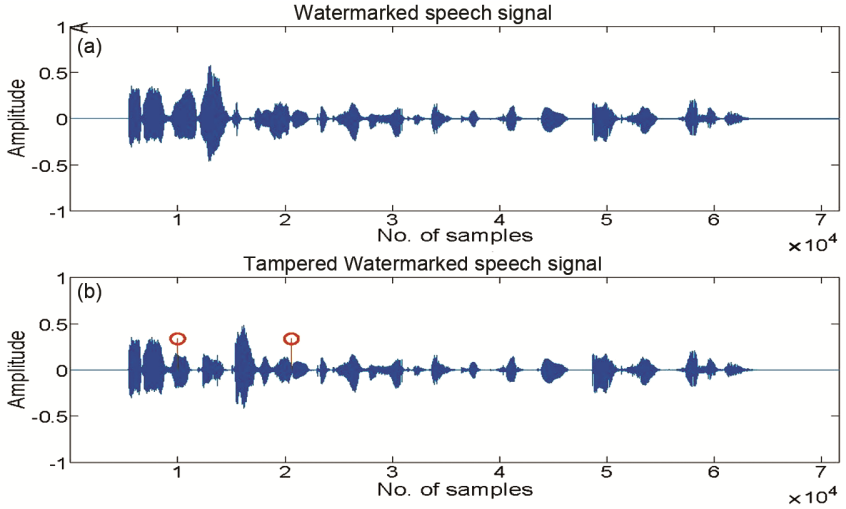

Fig. 8 — Tampered location in substitution attack (a) watermarked speech signal and (b) the substituted speech sample between marked portions.

substitution, insertion, and deletion attacks on the watermarked speech signals for demonstrating the detection of tamper location. Various attacks and detected tampered locations are discussed in the following subsections.

(1) Mute attack: In this section, we have selected 5000 speech samples from watermarked speech signals for performing the mute attack on the speech signal as shown in Fig. 7(a). The speech sample starting from 15000 to 20000 sample locations are marked by red line markers as shown in Fig. 7 (b). The main purpose of this attack is to evaluate the tamper location and recover them after performing mute attacks by the illegitimate users.

(2) Substitution attack: In this section, we have performed a substitution attack on watermarked speech signals as shown in Fig. 8. This type of attack is also called an integrity attack. Using substitution attack, the actual watermarked speech signals of sample location range from 10000^{th} to20600^{th} are substituted by the new 10600 samples. There is a need to detect the substituted place and to recover the actual samples.

(3) Insertion attack: This attack is known as an integrity attack. The insertion of dummy/duplicate data could be possible in any place of the watermarked speech signals. Here, the insertion attacks (10000 new samples from other signals) have

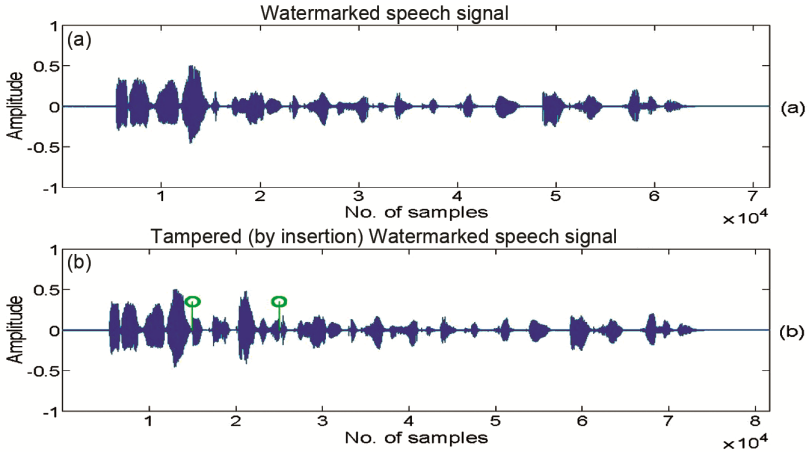

Fig. 9 — Tampered location in insertion attack (a) watermarked speech signal and (b) the inserted samples from anywhere.

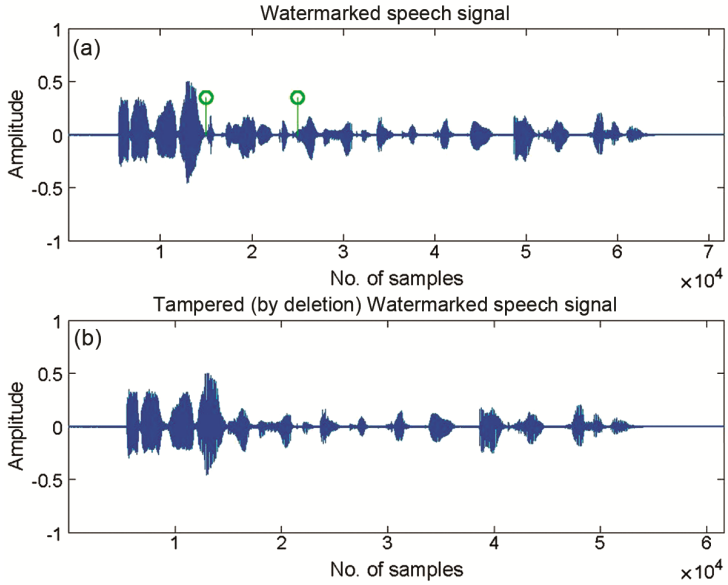

Fig. 10 — Tampered location for deletion attack (a) watermarked speech signal and (b) the deleted samples from anywhere.

been performed at random place of watermarked speech signals as shown in Fig. 9. The attackers can insert new samples at random place of watermarked speech signals as shown in Fig. 9(b).

(4) Deletion attack: This attack is also called an integrity attack. Using deletion attack, the 10000-15000 samples are deleted from any random places of watermarked speech signals as shown in Fig. 10. The actual watermarked samples are deleted by the attackers from the random place (let us consider sample 15000th to 25000th) of watermarked speech signals as shown in Fig. 10(a).

Table 5 shows the computed PESQ values obtained from recovered signals after performing different attacks. We have concluded that the recovered signals are similar to original signals.

Table 5 — PESQ values for tamper recovery of watermarked speech signals on various attacks.

Integrity attacks	Avg. PESQ
Mute attack	4.38
Sustitution attack	4.3
Insertion attack	4.2
Deletion attack	4.35

4 Conclusion

In this paper, we have proposed epoch event based tamper detection algorithm for digital speech watermarking technique. The proposed method has been developed using epoch interval, LP analysis, logistic mapping, and verify tamper location. The detection of epoch location has been applied for the purpose of watermark insertion. The length of each epoch intervals is different. Therefore, it will provide high security. The

primary focus of our method is to provide copyright protection, authentication and tampered detection. The simulation results demonstrated that the PRE value is approximately more than 82% and PESQ value lies between excellent and good. It means that there are no significance differences between original and watermarked signals and we have obtained highly imperceptible, robust, and secure system. We also detected tampered locations after performing the mute, substitution, insertion, and deletion attacks.

References

1 Kumar R, & Yadav J, *Test Engmanag,* 83 (2020)22661.
2 Hartung F & Kutter M, *Proc IEEE*, 87(7) (1999)1079.
3 Brassil J T, Low S, Maxemchuk NF & O'Gorman L, *J Selec Areas Comm*, 13(8) (1995) 1495.
4 Bhattacharya S, Chattopadhyay T & Pal A, *Int Symposium Consum Electron IEEE,* (2006)1.
5 Kumar R, Kumar S & Brar S S, *IEEE,* 11 (2016)1.
6 Singh V, *Digital watermarking: a tutorial, JSAT*, January Edition, (2011).
7 Kankanhalli MS, Ramakrishnan K R, & Rajmohan, *Content based watermarking of images, In Proceedings of the sixth ACM international conference on Multimedia*, (1998)61.
8 Nematollahi M A & Al-Haddad S A, *Int J Speech Technol*, 141 (2013)471.
9 Jadhav A & Kolhekar M, *Signal Proc Comput Technol, IEEE,1* (2014)140.
10 Langelaar G C, Lagendijk R L & Biemond J, *J Visual Comm Image Repres*, 9(4)(1998)256.
11 Yadav J, Fahad M S & Rao K S, *Speech Comm*, 96 (2018)142.
12 Vijayan K, & Murty K S R, *Speech Signal Proc (ICASSP) IEEE*, (2014) 1493.
13 Vijayan K & Murty K S, *CircSyst Signal Proc,* 35(2016)2584.
14 Zhao G, Qin C, Yao H & Han Y, *Pat Recog Let,* 164(2022)16.
15 Yamni M, Karmouni H, Sayyouri M & Qjidaa H, *Expert Syst Appl*, 203(2022)117325.
16 Pavlović K, Kovačević S, Djurović I &Wojciechowski A*, Digital Signal Proc*, 122(2022)103381.
17 Baziyad M, Kamel I, Rabie T & Kabatyansky G, *Proc Comput Sci,* 231(2024)397.
18 Xiang L, Liu Y, & Yang Z, *Info Proc Managt,*61(3) (2024)103661.
19 Kim DW, Kim JK, Piao Z & Seo Y H, *Digital Signal Proc*, 146 (2024)104352.
20 Chen H, Liu C, Zhu T& Zhou W, *Comput Stand Inter*,(2024)103830.
21 Patil A J & Shelke R, *High-Confi Comput*, 3(4)(2023) 100153.
22 Amrit P, & Singh A K,*Comput Comm*, 188(2022) 52.
23 Kumar R & Yadav J, *J Info Sec Appl*, 68 (2022)103218.
24 Faundez-Zanuy M, Lucena-Molina J J & Hagmüller M, *J forensci*, 55(4) (2010)1080.
25 Kumar K P, & Kanhe A, *Arabian J Sci Eng*, 47(8)(2022)10003.
26 Yadav J, *Comp Speech Language*, 70(2021)101231.
27 Kumar R & Yadav J, *Int J Signal Imag Sys Eng*, 12(3)(2021)71.
28 Kumar R, Kumar M & Yadav J, *Int J Innov Technol Explor Eng (IJITEE)*, 9(6)(2020) 126131.
29 Shi C, Li X & Wang H, *IEEE Access*, 8, (2020), 22249.
30 Liu Z, Huang Y & Huang J, *IEEE transactions on information forensics and security*, 14(5), (2018), 1171.
31 Sakai H & Iwaki M, *IEEE,*(2018)343.
32 Kanhe A & Gnanasekaran A, *J Aud Speech Mus Proc*, (2018), 1.
33 Revathi A, Sasikaladevi N & Jeyalakshmi C, *Int J Speech Technol*, 21, (2018), 1021.
34 Agradriya BA, Perdana F K, Safitri I & Novamizanti L, *(ICACOMIT),*(2017), 17.
35 Weina W, *Digital audio blind watermarking algorithm based on audio characteristic and scrambling encryption, IEEE, 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC),*(2017), 1195.
36 Sun F, Liu Z & Qi C, *Int J Dig Cri Foren (IJDCF)*,9(2), (2017), 1.
37 Liu Z, Zhang F, Wang J, Wang H& Huang J*,, Signal Proc*, 123, (2016), 157.
38 Liu XL, Lin C C & Yuan SM, *IEEE Transcircsysts vid technol*, 28(5), (2016), 1047.
39 Deokar S M & Dhaigude B, *Blind audio watermarking based on discrete wavelet and cosine transform, international conference on industrial instrumentation and control (ICIC), IEEE,* (2015), 264.
40 Sarreshtedari S, Akhaee M A & Abbasfar A, *IEEE/ACM*, 23(11), (2015), 1917.
41 Ambika D & Radha V, *Int. J. Comput. Sci. Eng. Technol*, 5 (11), (2014), 1089.
42 Jane O & Elbaşi E, *J applres technol*,12(4)(2014)750.
43 Ali M & Ahn C W, *Signal proc,* 94, (2014), 545.
44 Alshammas H A, *IEEE*,(2013), 1.
45 Kaur R & Jindal S, *IEEE,* (2013), 19.
46 Hu H T, Chou H H & Lee T T, *IEEE Access*, 9 (2021), 9916.
47 Wang J & He J, *Multimedia Tools Appl,*76 (2017) 14799.
48 Zareian M, Zargarchi S M & Sarsarshahi A*, IEEE,11* (2012) 27.
49 Talbi M, Ftima S B & Cherif A, *IEEE,2* (2017) 522.