Multimedia Tools and Applications (2023) 82:327–346
https://doi.org/10.1007/s11042-022-13301-4

Hybrid scheme for safe speech transmission based on multiple chaotic maps, watermarking and Arnold scrambling algorithm

Hadda Ouguissi[1] · Slami Saadi[2] · Ahmed Merrad[2] · Mecheri Kious[1]

Received: 15 April 2021 / Revised: 21 January 2022 / Accepted: 30 May 2022 /
Published online: 6 June 2022

Abstract
In this paper, we present a novel scheme for enhancing the security of speech information in communication systems. We build a hybridization of three approaches: Chaotic logistic and tent maps for generating an arbitrary vector by some primarily initiated values to be joined in the original speech signal, an integrated watermark image within the encrypted signal in order to verify, through decryption process, that the encrypted signal is authentic as well as does not suffer from eventual attacks, and the third approach is using an Arnold scrambling key (cat map) to spread signal samples by means of a secret key, then recuperate the original signal from samples which is not possible without this key. Obtained correlation value in the proposed scheme is closer to null which proves that original and encrypted signals are completely dissimilar. Moreover, we recovered the original speech without disturbing the quality. Numerical results of the Signal to Noise Ratio (SNR) and Correlation Coefficient (CC) reported below, and the comparison between the proposed approach to seven recently published works, also reported, reveal the superiority of the proposed scheme and validate our design to be considered amongst the best methods compared to other recently existing strong approaches.

Keywords Speech security · Hybrid · Chaotic · Watermarking · Arnold

1 Introduction

Speech security systems have been broadly exploited in many applications. Currently, it is very important to protect speeches over communication systems with rapid and safe cryptosystems. As

✉ Slami Saadi
saadisdz@gmail.com

[1] Faculty of Technology, Materials, Energetic Systems, Renewable Energies, and Energy Management Laboratory (LMSEERGE), Ammar Thelidji University of Laghouat, Laghouat, Algeria

[2] Faculty of Exact Sciences & Informatics, Ziane Achour University of Djelfa (UZAD), Djelfa, Algeria

speech communications become further broadly used and yet more delicate, the significance of offering a superior rank of security is of great significance. Up to date, various speech encryption methods have been suggested. Speech watermarking is a strong means to secrete and hence protect information from several intended or unintended utilization during communication. Speech watermarking types and applications as well as topics of robustness, capacity and imperceptibility are detailed in [21]. Authors of [5] presents an efficient safe communication system based on a speech watermarking approach in the purpose to permit an automatic recognition of the speaker, with an optimization of the whole system to boost its performance.

We proposed in [24] a new design for blind watermarking of speech and audio signals, in which we introduced the discrete wavelet transform (DWT) and the discrete cosine transform (DCT) after segmenting the signal. For protection reason, we applied Arnold transform on the watermark to save recognition security. In addition, we presented in [18, 19] a robust blind speech watermarking method using DCT and DWT inside signal sub-sampling. To get high-quality imperceptibility, fusion is realized against various attacks such as: re-quantization, cropping, echo amplification and additive white Gaussian noise (AWGN).

Chaos is a characteristic action of nonlinear dynamic systems. It is described by its large sensitivity to factors and first conditions, which operates as the encryption keys. Mathematically identified as uncertainty governed through deterministic laws. This behavior of chaotic signals offers the possibility to handle several applications. Amongst, the use of Chaos within safe communication has a big consideration. The interesting chapter in [11] explains chaotic systems and illustrates their aptness for use to protect communications information. One of the major motivations for the improved protection of communication given by Chaotic is its broadband signal property that permits efficient spectral cover up of the communication by the chaotic transporter. Authors of [29]discuss the problem of the chaotic safe communication. New double channel diffusion system is given and used in protected communication design, next the channel-switching procedures are assumed to more boost the safety of messages transmission. Paper [20] introduces a low complexity, small delay, and high degree of secured speech encryption method based on change of speech pieces by chaotic Baker map and replacement by masks in together time and transform domains to fill the unvoiced periods inside speech conversation. Chaotic shift keying-based speech encryption and decryption approaches has been presented in [25], where the input speech signals are sampled and its values are segmented into four levels which are permuted using four chaotic generators. A novel speech encryption using fractional chaotic systems is given in [28], where two-channel transmission process is used. The original speech is encoded by a nonlinear function of the chaotic states. The work in [23] intends to show modules for improving the security of speaker authentication by inserting the watermark in the detail coefficients of the speech signal after applying wavelet transform and basing on the energy computation. Speaker is identified by speech and the removed watermark from the watermarked speech. Authors of [4] study and implement the effect of using different floating-point representations on the chaotic system's performance, for speech security, with numerical simulations for all discussed chaotic systems showing good results in terms of MSE, entropy, correlation coefficient, and pass the NIST test.

In this work, we combined the watermarking with a chaotic approach in a hybrid scheme based on Arnold scrambling Algorithm, in the scope to enhance the security of speeches in communication systems. Our contribution in this work is the substitution of the discrete wavelet transform (DWT) and the discrete cosine transform (DCT) and the segmentation of the speech signal by a novel chaotic representation after spreading and Arnold scrambling by a secret key. The superiority of the proposed scheme is revealed in numerical comparisons with other published works and our design is validated as well in order to be considered amongst the best methods compared to other recently existing strong approaches. In addition, this work highlights the efficiency of the designed schemes in furnishing strength for copyright protection to possession of the data and validating people by using speech as a biometric tool.

2 Chaotic generator (tent map, logistic map)

2.1 Logistic map

The simplest discrete chaotic systems functions that have been used recently for cryptography applications is the logistic map. The logic map function is expressed as:

$$\mathrm{x}_{\mathrm{n}+1} = r.\mathrm{x}_{\mathrm{n}}.(1-\mathrm{x}_{\mathrm{n}})$$

Where x_{n} takes value in the interval (0, 1), the parameter r is a positive constant and takes values up to four. Its value establishes and investigates the manner of the logistic map. From r = 3.57 the iterations become completely chaotic and start to provide themselves to the aim of encryption. So a superior value of parameter r is selected to get an extremely chaotic so far deterministic discrete-time signal [20, 25, 28]. The preliminary value x_0 and the parameter r areconsidered as the secret key.

2.2 Tent map

The chaotic manners of the tent map (a piecewise linear, constant map with a single maximum) has been considered analytically all over its chaotic region in terms of the invariant density and the power spectrum. As the elevation of the highest point is lowered, consecutive band-splitting changes take place in the chaotic area and gather to the change point into the non-chaotic area. The time-correlation function of non-periodic paths and their power spectrum are computed precisely at the band-splitting points and in the neighborhood to these points. The tent map is topologically conjugate, and hence the performances of the map are in this sense equal below iteration. The chaotic tent map is defined by:

$$\begin{gathered}\mathrm{x}_{\mathrm{i}+1} = f(\mathrm{x}_{\mathrm{i}}, u)\\ f(\mathrm{x}_{\mathrm{i}}, u) = u\mathrm{x}_{\mathrm{i}} \quad \text{if } \mathrm{x}_{\mathrm{i}} < 0.5\\ f(\mathrm{x}_{\mathrm{i}}, u) = u(1-\mathrm{x}_{\mathrm{i}}) \quad \text{otherwise}\end{gathered}$$

Where:$\mathrm{x}_{\mathrm{i}} \in [0,\ 1]$ for $\mathrm{i} \geq 0$.

This map converts an interval [0, 1] onto itself and includes merely one control parameter u, correspondingly, where u∈[0, 2], x_0 is the initial value of the system. The set of real values x_0, x_1, $\ldots\mathrm{x}_{\mathrm{n}}$ is named the orbit of the system. Depending on the control parameter u, the system illustrates a variety of dynamical actions varying from expected to chaotic [7, 13, 26].

3 Watermarking

Digital watermarking retrieve is stronger if the original un-watermarked information are available. However, access to the original main signal cannot be acceptable on the entire real-world circumstances [16].In many applications, the identification algorithm is capable of using the original audio signal to extract the watermark from the watermarked signal [3]. It, often significantly, obtains superiorly the detector performance; because the watermark information is extracted throughout subtract the original signal from the watermarked signal. However, if the identification algorithm does not have access to the original signal and this inability considerably decreases the amount of information that could be masked in the original signal. The full process of the watermark insertion and removal is modeled as a communication canal where the watermark is distorted due to the presence of strong intrusive in addition to canal properties.

4 Arnold scrambling algorithm

The KxK matrix W is altered into W′ by Arnold transformation to decrease the autocorrelation coefficient of the image and subsequently the confidentiality of watermark is strengthened [14]. Arnold transformation is cyclical and whereas it is iterated, rarely the original signal will be achieved. The Arnold scrambling algorithm [10] has the features of simplicity and periodicity. So it is generally used to provide an extra stage of protection. Arnold Transform is well recognized as cat seem transforms and is merely suitable for scrambling speech signals by dividing the signal into some vectors which can be converted to N × N dimension matrices used then to mix up the signal.

Arnold Transform is cyclic in nature. The signal decryption depends on the scrambling key, which can be used as secret key and identifies the amount of times that has been scrambled.

5 The proposed hybrid chaotic watermarking architecture

5.1 Emitter side

We try to construct a random chaotic signal using (Tent map, Logistic map). We carry out the fusion of the used watermark with the original speech signal. The produced watermarked signal is combined with the random chaotic signal using a chaotic key to generate an encrypted signal, this signal is then transmitted, Figs. 1, 2 and 3.

The encryption process starting with reading a speech signal stocked in Hard disk using a Matlab function, where, the Matlab recommends to represent the speech file in the range [−1,1]. Also, read the watermark file. Then steps are as follows:

1) In this step, the user inputs a key (key1) to embed the watermark securely within the original speech signal. The scheme considered to embed the watermark is presented in [24]. In this method, we embed the watermark in DCT and DWT domain and employ sub-sampling technique. This method offers the embedding control from side transparency and robustness of the watermark with a shifting value (Δ). This step results a speech signal marked by a secret information (watermark) named Wtr_Sp.
2) Logistic map and tent map create two chaotic signals, depending on the initial values input by the user, those chaotic signals are generated, and named as Lg_S and Tn_S, respectively.

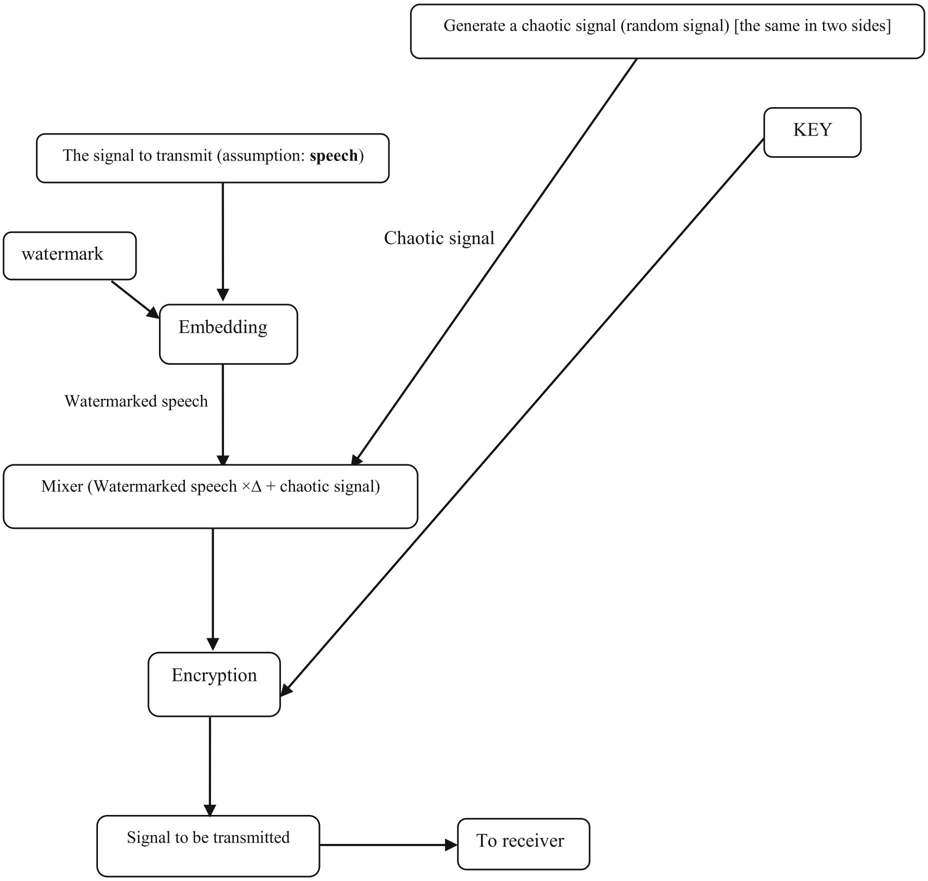

Generate a chaotic signal (random signal) [the same in two sides]
KEY
The signal to transmit (assumption: **speech**)
Chaotic signal
watermark
Embedding
Watermarked speech
Mixer (Watermarked speech ×Δ + chaotic signal)
Encryption
Signal to be transmitted
To receiver

Fig. 1 Flowchart of the encryption scheme

3) Using the formulas below, the three signals Lg_S, Tn_S and Wtr_Sp are mixed to produce a new signal named Mx_Sg:

$$\begin{cases} Mx\_Sg_i = (\mathrm{Tn\_S_i} \times \mathrm{Wtr\_Sp_i} + (1-\mathrm{Tn\_S_i})\mathrm{Lg\_S_i})-1; & \mathrm{Wtr\_Sp_i} => 0 \\ Mx\_Sg_i = (\mathrm{Tn\_S_i} \times \mathrm{Wtr\_Sp_i} + (1-\mathrm{Tn\_S_i})\mathrm{Lg\_S_i}) + 1; & \mathrm{Wtr\_Sp_i} < 0 \end{cases} \quad (1)$$

Where i represents samples index.

4) Decomposing the Mx_Sg into segments, where each segment length is a square number.
5) Before applying Arnold transform, each segment is reshaped into 2D matrix (N × N elements)
6) The user inserts another key (key 2), then the encryption process employs that key on each matrix to scramble its elements with Arnold transform.
7) Reshape each scrambled matrix into 1D vector with length N^2.
8) To obtain the final encrypted speech signal, the process of encryption collects the segment with each other.

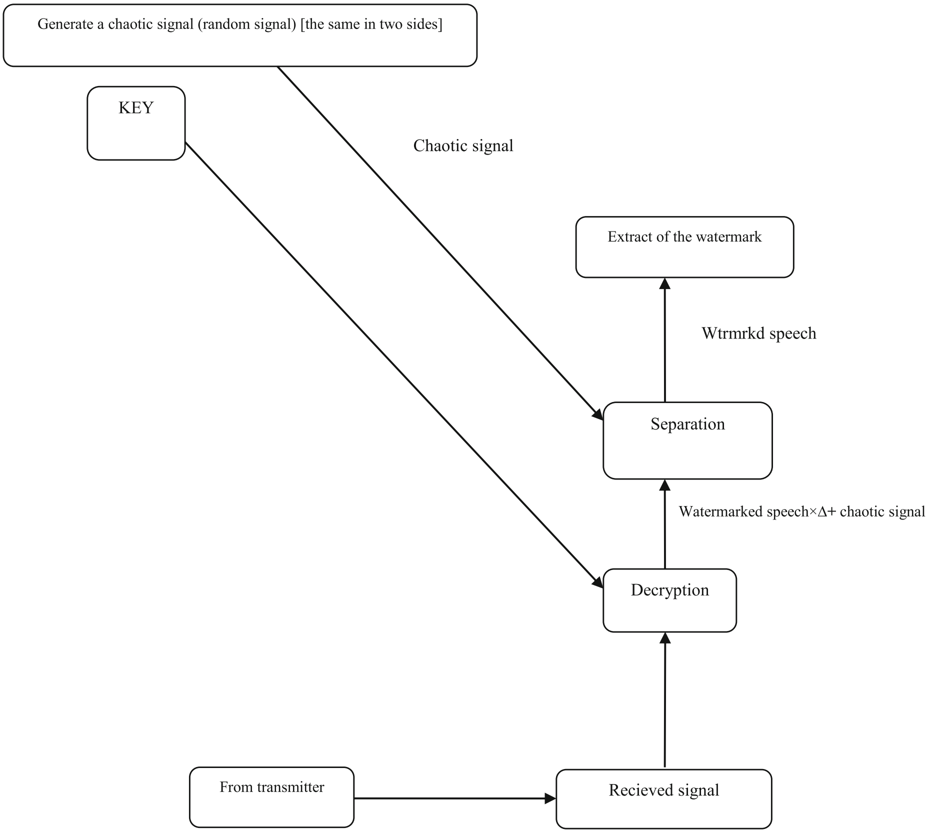


Fig. 2 Flowchart of the decryption scheme

5.2 Receiver side

The inverse process is performed here. Using the previous chaotic key, the received signal is decrypted by removing the same random chaotic signal generated in the transmitter side. We extract the watermark from the decrypted signal and verify the obtained signal with the original to ensure its originality without degradation, Figs. 2 and 3.

1) The steps 4 and 5 in the encryption process, are applied on the encrypted speech signal.
2) Inverse Arnold transform is then applied on each 2D matrix using the same key(key2) employed previously.
3) Reshape each retrieved matrix to 1D vector with length N^2.
4) Collect the retrieved segments with each other to produce$Mx_Sg_i^{'}$.
5) The same second step in the encryption process is applied without changing the initial value.
6) Decrypted speech signal samples separation is accomplished respecting the following:

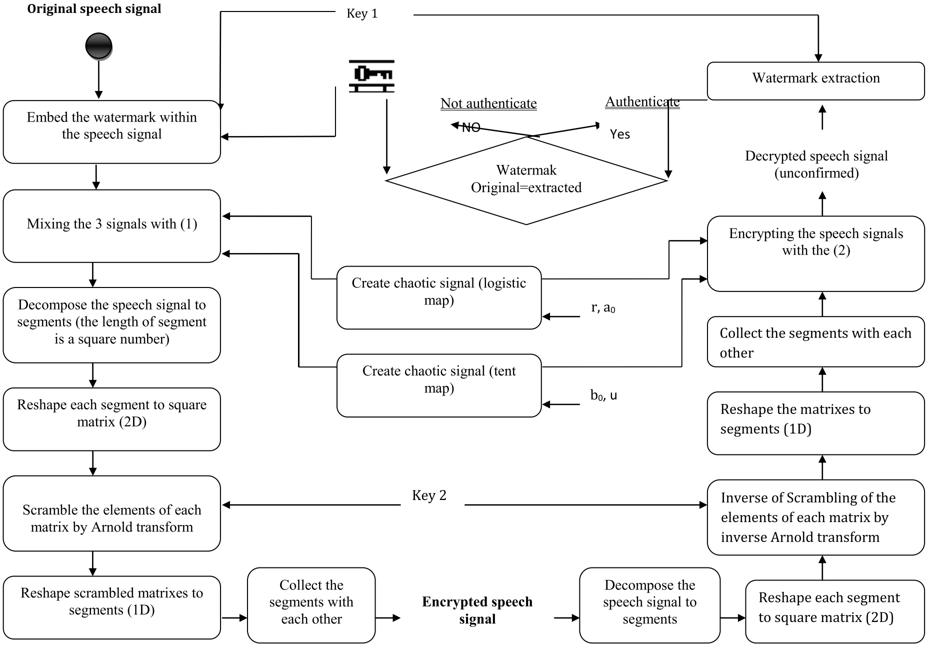

Original speech signal
Key 1
Watermark extraction
Embed the watermark within the speech signal
Not authenticate
Authenticate
NO
Yes
Watermak Original=extracted
Decrypted speech signal (unconfirmed)
Mixing the 3 signals with (1)
Encrypting the speech signals with the (2)
Create chaotic signal (logistic map)
r, a0
Decompose the speech signal to segments (the length of segment is a square number)
Collect the segments with each other
Create chaotic signal (tent map)
b0, u
Reshape each segment to square matrix (2D)
Reshape the matrixes to segments (1D)
Scramble the elements of each matrix by Arnold transform
Key 2
Inverse of Scrambling of the elements of each matrix by inverse Arnold transform
Reshape scrambled matrixes to segments (1D)
Collect the segments with each other
Encrypted speech signal
Decompose the speech signal to segments
Reshape each segment to square matrix (2D)

Fig. 3 Encryption and decryption speech processes cycle

$$\begin{cases} \mathrm{Wtr\_Sp}_{\mathrm{i}}^{'} = \dfrac{\left(Mx\_Sg_i^{'} + 1 - (1 - \mathrm{Tn\_S_i})\mathrm{Lg\_S_i}\right)}{\mathrm{Tn\_S_i}} & Mx\_Sg_i < 0 \\ \mathrm{Wtr\_Sp}_{\mathrm{i}}^{'} = \dfrac{\left(Mx\_Sg_i^{'} - 1 + (1 - \mathrm{Tn\_S_i})\mathrm{Lg\_S_i}\right)}{\mathrm{Tn\_S_i}} & Mx\_Sg_i => 0 \end{cases} \tag{2}$$

Where: $\mathrm{Wtr\_Sp}_{\mathrm{i}}^{'}$ is the decrypted speech signal and $Mx\_Sg_i^{'}$ results from the fourth step.

Until this step the speech signal is decrypted, but not confirmed. To verify that the speech signal is safe and sent from authenticate side, the decryption process maintain with these steps:

1) Extraction of the watermark included within the encrypted speech signal, and that by employing the same key(key1) in extraction process presented in [24].
2) Authentication of decrypted speech signal controlled by verification of the similarity between extracted and original watermark. So, more similarity between the two means decrypted speech more authenticate.

6 Performance evaluation metrics

6.1 Correlation coefficient

The correlation coefficient, usually denoted by ‘*r*’, is a measure of the strength of the straight-line or linear relationship between two variables [22].In our case the variables are original, encrypted and decrypted speech signal. If two variables are closely related with stronger association, the correlation coefficient is close to the value 1. On the other hand, if the coefficient is close to 0, two variables are not related and cannot predict each other.

The correlation coefficient '*r*' can be calculated respecting the following formulas [6]:

$$r_{S_1S_2}=\frac{\frac{1}{L}\sum_{i=1}^{L}\left(S_{1,i}-E(S_1)\right)\left(\left(S_{2,i}-E(S_2)\right)\right)}{\surd\left(\frac{1}{L}\sum_{1,i}^{L}\left(S_{1,i}-E(S_1)\right)^2\right)\times\surd\left(\frac{1}{L}\sum_{1,i}^{L}\left(S_{2,i}-E(S_2)\right)^2\right)}\text{ Where }E(S)=\frac{1}{L}\sum_{i=1}^{L}S_{i.}$$

L is the length of speech signals (number of samples), S_1andS_2 are the duality of the two signals (original,encrypted) or (original, decrypted).

6.2 Signal to noise ratio (SNR)

To confirm the performance of digital speech encryption schemes the SNR is calculated, where the SNR measures the noise content in the encrypted speech signals. Cryptanalyst always try to increase the noise content in the encrypted signal so as to minimize the information content in the encrypted data. Also the decipher tries to reduce the noise content in the decrypted signal. Signal to noise ratio is a factor employed to identify the amount by which the signal is stained with noise. The Signal to noise ratio can be calculated by the equation below [8]:

$$\mathrm{SNR}=10\times\log 10\frac{\sum_{\mathrm{i}=1}^{\mathrm{L}}\mathrm{S}_{1,\mathrm{i}}^{2}}{\sum_{\mathrm{i}=1}^{\mathrm{L}}\left(\mathrm{S}_{1,\mathrm{i}}-\mathrm{S}_{2,\mathrm{i}}\right)^{2}}$$

$S_{1,\ i}$and $S_{2,\ i}$represent the ith samples of the (original, ciphered) or (original,deciphered) speech signals, respectively, and L represents the length of speech signals,

6.3 Bit error rate (BER)

To authenticate that the received encrypted speech signal was sent from trusty side, we make examination of the BER. The BER is employed to verify the similarity between the two watermarks, the original and the extracted watermarking image. In addition, BER equals zero means that there is no effect on the watermark and the extraction is successful which means that the received speech signal is sent from authenticated side. BER is expressed by the following formula [24].

$$\mathrm{BER}=\frac{\mathrm{B_{ERR}}}{\mathrm{N}}\times 100\%$$

Where: $\mathbf{B_{ERR}}$: The quantity of erroneous bits

N: The number of all bits(size of the watermark)

6.4 NSCR and UACI

In our proposal, computing the unified average changing intensity (UACI) and number of sample change rate (NSCR)between the two encrypted speech signals is to look for the degree of variation when the key is modified slowly. In other words to evaluate the sensitivity of the key. The NSCR and UACI of the two encrypted speech signals are calculated using equations below [26]:

$$\mathrm{NSCR} = \sum_{i=1}^{L}\frac{d_i}{L} \times 100\% \;\; \text{Where}: d_i = \begin{cases} 1, & S'_{i,1} = S'_{i,2} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{UACI} = \frac{1}{L}\left[\sum_{i=1}^{L}\frac{S'_{i,1} - S'_{i,2}}{\mathrm{Max}}\right]$$

$S'_{i,1}$and $S'_{i,2}$are the two speech signals with a slow difference on the key in the i^{th} sample.

L: represents the length of the speech vector.

Max: depends on [2, 15], each sample of speech and audio signals assuming an integer value in the range [0–65,535]and in that situation Max = 65,535,so when using Matlab environment, the digital speeches are normalized in the range [−1–1], subsequently, the Max is 2.

7 Experimental results

In this part, we will assess the proposed scheme by experimental tests using two computing PC's with windows7, 32bits, CPU dual core. The first with 2gb RAM and on MATLAB7.1 environment and the second with 4gb RAM and on MATLAB8.1 environment. We used the second computing machine for experimenting elapsed time for execution.

All experiments are made using 20 speech files including male and female voices with different periods. This mono voice samples are selected randomly with 16bits for each sample. Table 1 illustrates the used speech signals with duration taken from the famous voices database TIMIT with gender identification. In Table 2, we present the initial statics data values of the two Chaotic maps (Tent and Logistic) on which all results are obtained. The watermark used image (16x16bit) is given in Fig. 4.

Table 1 The used speech signals with duration taken from the famous voices database TIMIT

Speech signal	Duration (second)	Gender (Male or Female)	Speech signal	Duration (second)	Gender (Male or Female)
SI560	**4.378**	**M**	SI1303	**4.294**	**M**
SI734	**4.525**	**M**	SI1308	**5.779**	**F**
SI770	**4.762**	**F**	SI1390	**5.094**	**F**
SI839	**4.653**	**M**	SI1460	**5.069**	**M**
SI860	**4.365**	**F**	SI1715	**4.512**	**M**
SI863	**4.474**	**F**	SI1992	**3.974**	**M**
SI943	**3.757**	**F**	SI2194	**4.723**	**F**
SI1103	**6.086**	**M**	SI2303	**4.058**	**F**
SI1109	**4.544**	**F**	SX29	**7.571**	**M**
SI1217	**5.197**	**M**	SX364	**4.339**	**M**

Bold entries indicate that these entries are obtained results.

Table 2 The initial statics data values of the two Chaotic maps

Δ	Logistic map		Tent map		Max_val
	a_0	R	b_0	U	
0.002	0.5	3.85	0.5	1.8	1

7.1 Key space

The total number of different keys that used in the encryption system called briefly as key space [13]. In addition, the good encryption system needs to offer a great key space and that for compensating the degradation dynamics in PC, and thus prevents invaders to decrypt original data even after they invest large amounts of resources and time [9].With omitting logistic and tent map, only Arnold scrambling can give a wide key space, where, available permutation positions of an M × M matrix are (M × M)!, for example if we consider a size of matrix are $(8 \times 8)! \approx 1,26 \times 10^{89}$, so what will be if the size of the matrix with hundreds.

Depending on [17] the designing of a cryptosystem resists against brute force attack, the size of the key space should be larger than $2^{128}(\approx 3,24 \times 10^{38})$, based on this point we can conclude that the proposed cryptosystem can resists brute-force attack sufficient for reliable practical employ.

7.2 Keys sensitivity analysis

All the initial values (**a_0** and **r**) from logistic map, (**b_0** and **u**) from tent map and **K** from Arnold scrambling are a keys, so we tried to test and examine the sensitivity of the encryption algorithm by changing one key or multiple keys. Table 3 shows the values of correlation coefficient, NSCR and UACI, where those values obtained with encryption one of the selected speech signal with a series of keys, then tried to detect the encrypted speech signal using different keys series. To look at the difference and importantly of the keys, firstly the speech signal is encrypted then the metrics mentioned previously calculated using the decrypted speeches signals with the true keys and with wrong keys. The NCSR demonstrates that the two decrypted speech signal hold usually a differ samples with a percentages near 100%. The UACI confirms that the intensities of the samples

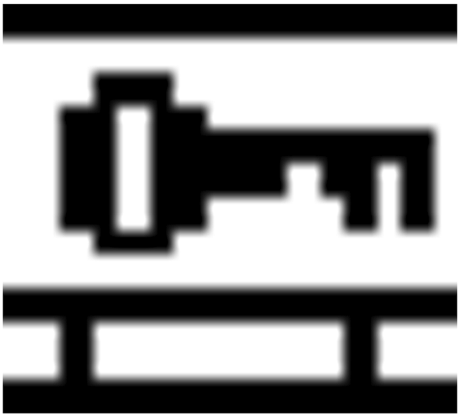


Fig. 4 The watermark used image

Table 3 Keys Sensitivity

Speech's name	Keys(a_0,r,b_0,u,k)	(3.87,0.48,1.82,0.54,205)			(3.85,0.5,1.8,0.5387)		
		NSCR (%)	UACI(%)	Corr_Coef	NSCR(%)	UACI(%)	Corr_Coef
SI770	**(3.85,0.5,1.8,0.5,205)**	99.9974	19.2024	−0.0921	99.9934	23.3060	0.0041
SI1390		99.9939	19.2098	−0.0767	99.9816	23.3290	-1.2662×10^{-4}
SI863		99.9930	19.2134	−0.0909	98.6923	29.4809	0.0023
SI1715		99.9931	19.1898	−0.0369	99.9848	18.6973	0.0053
SI1217		99.9988	19.2191	−0.0838	99.9964	23.3279	6.8499×10^{-4}

Bold entries indicate that these entries are obtained results.

between the decrypted speeches signals are divergent. Finally correlation coefficient affirms the relationship it to be underprivileged, specifically from Arnold key changing. From the obtained results we can conclude that even changes on the encryption keys values during the decryption process leads to wrong decryption results.

7.3 SNR and correlation coefficient

7.3.1 Encryption process effect

The encryption is considered more acceptable when the correlation coefficient value is close to zero. In addition, the encryption process is better when the SNR value decreased. Based on this and from data gathered in Table 4,we observe that the SNR values look too small and the correlation coefficient values are close to zero, and become negative which show that the encrypted signal is very far from the original speech signal and this indicates that the characteristics of the original signal are completely segregated.

7.3.2 Decryption process effect

The quality of the speech signal extracted from the encrypted signal is an essential characteristic. Otherwise, the encryption process is not significant. For this, we will discuss the quality of the

Table 4 Numerical results of the Signal to Noise Ratio (SNR) and Correlation Coefficient (CC) between the original and the encrypted signals

Speech signal	SNR (origin, encrypted)	Corr_coef (origin, encrypted)	Speech signal	SNR (origin, encrypted)	Corr_coef (origin, encrypted)
SI560	**−29.0691**	**−0.0019**	SI1303	**−35.4358**	**0.0073**
SI734	**−39.5201**	**0.0017**	SI1308	**−35.2122**	**0.0020**
SI770	**−42.6570**	**1.3989×10^{-4}**	SI1390	**−36.4082**	**0.0047**
SI839	**−32.5074**	**0.0038**	SI1460	**−37.8784**	**-5.5553×10^{-4}**
SI860	**−34.9999**	**0.0078**	SI1715	**−31.4766**	**0.0029**
SI863	**−39.7264**	**8.9170×10^{-4}**	SI1992	**−31.3417**	**−0.0047**
SI943	**−31.5172**	**−0.0030**	SI2194	**−29.6634**	**−0.0028**
SI1103	**−42.3124**	**0.0056**	SI2303	**−40.4777**	**−0.0015**
SI1109	**−37.1985**	**−0.0061**	SX29	**−35.2280**	**-7.4206×10^{-4}**
SI1217	**−36.9070**	**−0.0068**	SX364	**−30.7479**	**0.0043**

Bold entries indicate that these entries are obtained results.

Table 5 Numerical results of the Signal to Noise Ratio (SNR) and Correlation Coefficient (CC) between the original and the decrypted signals

Speech signal	SNR (original, decrypted)	Corr coef (original, decrypted)	Speech signal	SNR (original, decrypted)	Corr coef (original, decrypted)
SI560	**35.3820**	**0.99985**	SI1303	**34.6068**	**0.99982**
SI734	**32.5493**	**0.9997**	SI1308	**35.8761**	**0.99987**
SI770	**27.0283**	**0.9990**	SI1390	**32.8723**	**0.99974**
SI839	**35.8473**	**0.99985**	SI1460	**33.4023**	**0.99977**
SI860	**31.2040**	**0.9996**	SI1715	**38.4230**	**0.99992**
SI863	**31.5794**	**0.9997**	SI1992	**36.1811**	**0.99987**
SI943	**34.6754**	**0.9998**	SI2194	**40.1429**	**1.0000**
SI1103	**31.0270**	**0.9996**	SI2303	**29.4997**	**0.99943**
SI1109	**32.0621**	**0.99968**	SX29	**39.0748**	**0.99993**
SI1217	**32.5694**	**0.99972**	SX364	**37.0662**	**0.99990**

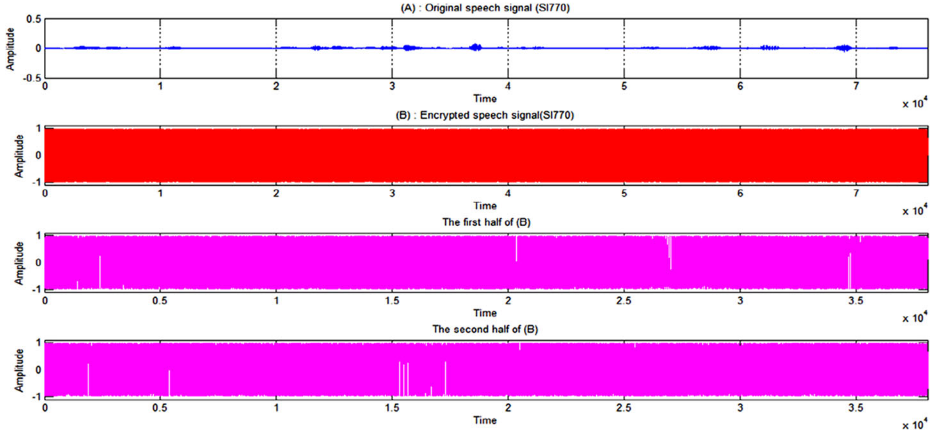
Bold entries indicate that these entries are obtained results.

decrypted signal from the encrypted one, using the two previous coefficient (correlation and SNR). Table 5 gives all statistics data for these coefficients for all speech signals. From these values, we can easily observe that the obtained values are excellent. The correlation coefficient reaches the smallest value of 0.99943 and close to 1, which signifies that there is no difference between the original and the decrypted speech signals and the encryption process is very good. The SNR values are also almost significant. The variations in SNR values are due to speech signal interval and energy, see Table 5. From all this discussion, we can conclude that the proposed scheme conserves greatly the quality of the speech signal when it is decrypted.

7.4 Waveforms review

7.4.1 Original and encrypted speech signals

The waveform A in Figs. 5, 6, 7 and 8 shows the original speech signal of: SI770,SI839,SI943 and SI1217 respectively. The waveform B shows the encrypted signal, and for more


(A) : Original speech signal (SI770)

(B) : Encrypted speech signal(SI770)

The first half of (B)

The second half of (B)

Fig. 5 SI770 waveforms (**A**): original, B:encrypted and the first/second half of the encrypted (**B**)

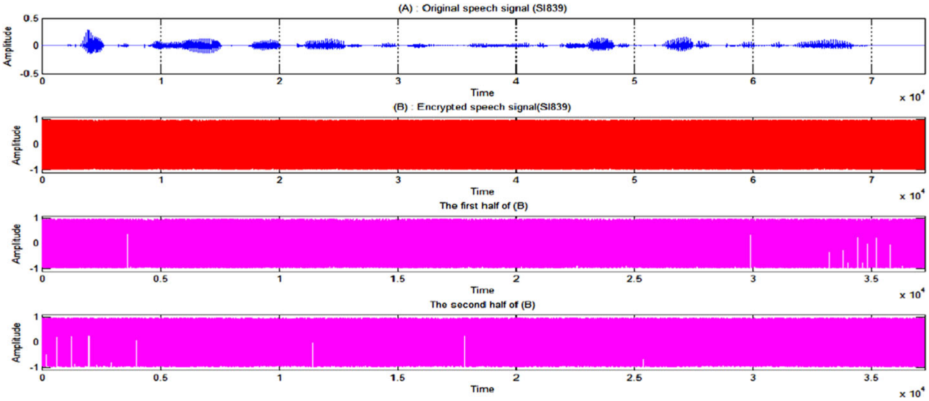


Fig. 6 SI839 waveforms (**A**): original, B: encrypted and the first/second half of the encrypted (**B**)

clarification, the last waveform B is illustrated in two parts. By observing these figures, we can clearly mention that there is no similarity between the original speech signal (A) and its encrypted version (B) which is regularly uniform and it has no relation with the variations of the original waveform (A).

7.4.2 Original and decrypted speech signals

The speech signals SI1715, SI2194, SI2303 and SX29 are showed in the first waveform of Figs. 9, 10, 11 and 12 respectively, and the decrypted speech signals are presented in the second waveform of the same figures. The third waveform illustrates the difference between the original speech signal end the decrypted one. Even if we focus well on the waveforms, we cannot distinguish between the original speech signal and the extracted decrypted signal, and we can only see the difference when we make the difference waveform. This difference waveform is showed with very small amplitude (0.01–0.01), and based on this very tiny difference, we can conclude that the two speech signals: the original and the decrypted one are similar and too close to each other.

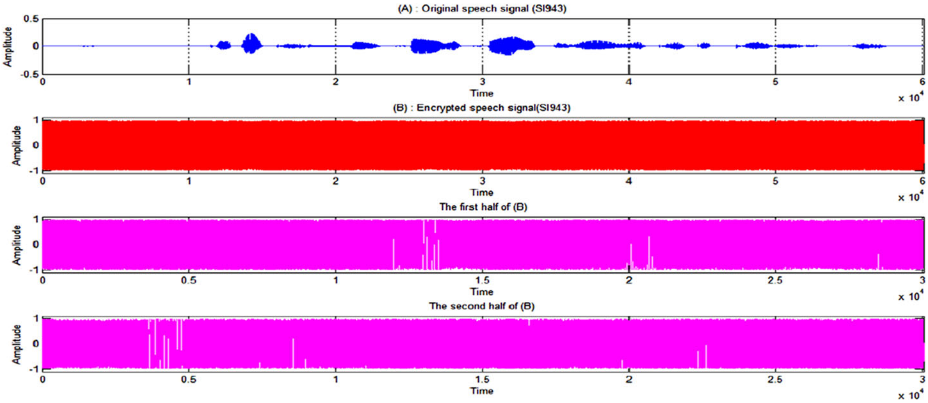


Fig. 7 SI943 waveforms (**A**): original, B: encrypted and the first/second half of the encrypted (**B**)

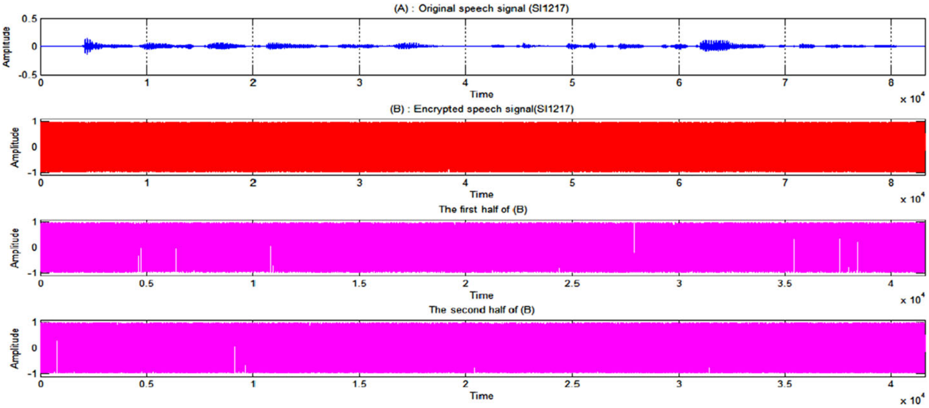


Fig. 8 SI1217 waveforms (**A**): original, B:encrypted and the first/second half of the encrypted (**B**)

7.5 Watermark control and authentication

The proposed scheme is based on adding a watermark to the original speech signal during encryption process and extracting this watermark during decryption process. The purpose of this operation is enhancing more the security and further credibility, so that extracting successfully the watermark during decryption confirms that the received signal is well authenticated and it is transmitted from and authenticated original signal without any transformations in the transmission media. Table 6 provides results when the speech signal is attacked by some AWGN additive white Gaussian noises. In the presented data in this table, we mention that the watermark is extracted successfully in the presence of small noise, but when the noise increases considerably, it affects the watermark. Which indicates that the speech signal is affected. We can observe this in the BER values implying that the transmitted encrypted signal is suffering from some attacks. We cite that we can control the strength of the watermark introduction so that it is possible increasing or decreasing the watermark sensitivity during undergoing the attacks by only varying the Δ values.

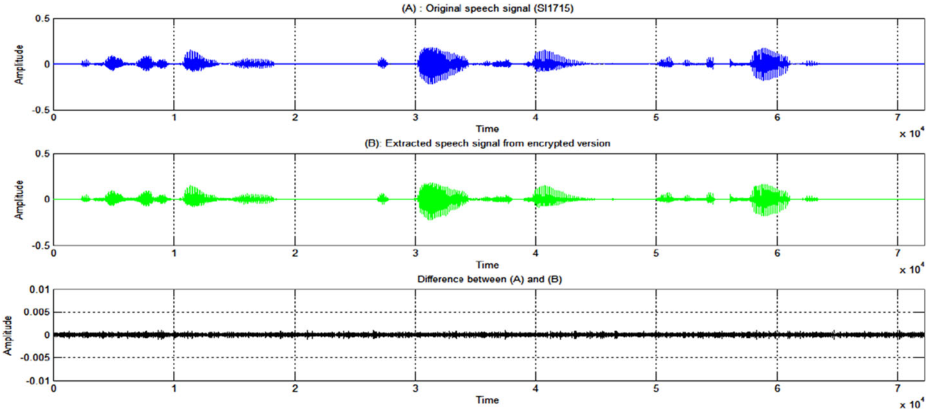


Fig. 9 SI1715 waveforms (**A**): original, (**B**): decrypted, the difference between the original and the decrypted speech)

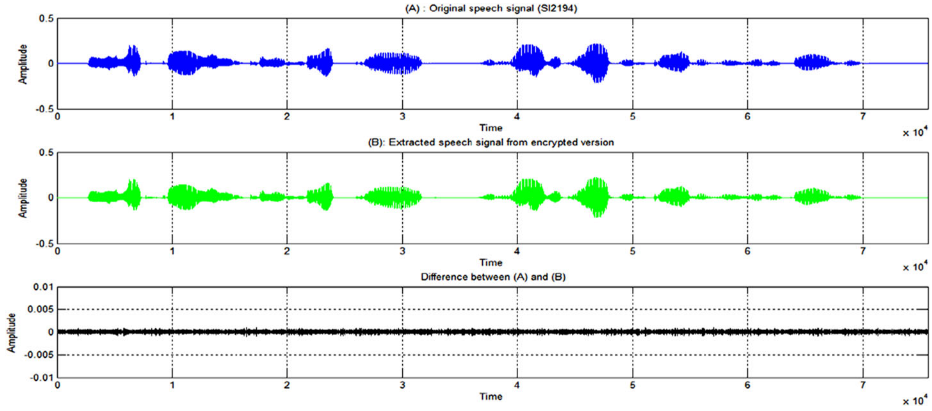


Fig. 10 SI2194 waveforms (**A**): original, (**B**): decrypted, the difference between the original and the decrypted speech)

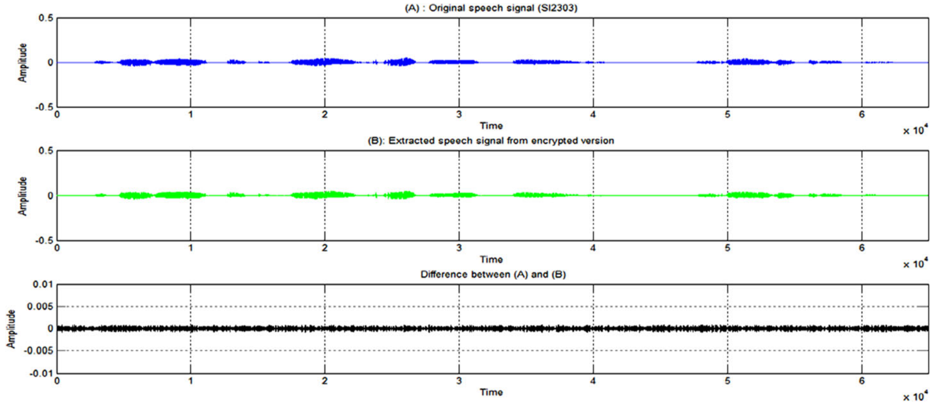


Fig. 11 SI2303 waveforms (**A**): original, (**B**): decrypted, the difference between the original and the decrypted speech)

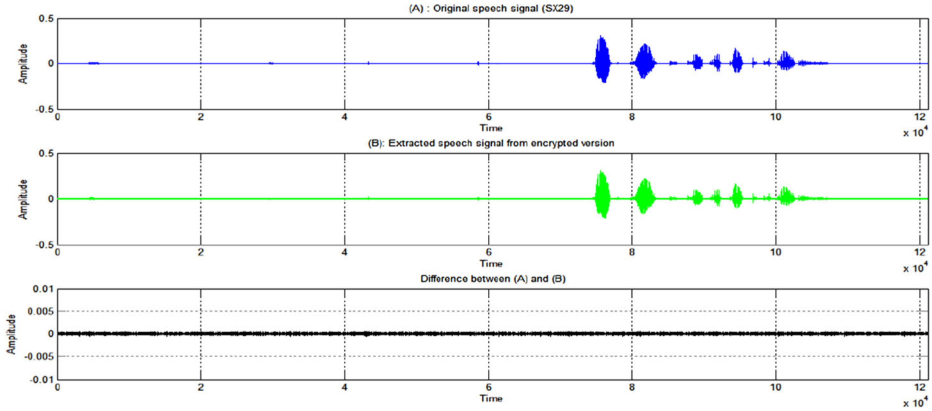


Fig. 12 SX29 waveforms (**A**): original, (**B**): decrypted, the difference between the original and the decrypted speech)

Table 6 BER values variation after speech Signals AWGN attacks

Speech signal	AWGN (dB)	BER	Extracted watermark	Speech signal	AWGN (dB)	BER	Extracted wtermark
SI560	70	**0**		SI1303	70	**0**	
	60	**0.0430**			60	**0.0352**	
SI734	70	**0**		SI1308	70	**0**	
	60	**0.0781**			60	**0.0547**	
SI770	70	**0**		SI1390	70	**0**	
	60	**0.0508**			60	**0.0430**	
SI839	70	**0**		SI1460	70	**0**	
	60	**0.0430**			60	**0.0625**	
SI860	70	**0**		SI1715	70	**0**	
	60	**0.0234**			60	**0.0352**	
SI863	70	**0**		SI1992	70	**0**	
	60	**0.0742**			60	**0.0664**	
SI943	70	**0**		SI2194	70	**0**	
	60	**0.0547**			60	**0.0391**	
SI1103	70	**0**		SI2303	70	**0**	
	60	**0.0508**			60	**0.0547**	
SI1109	70	**0**		SX29	70	**0**	
	60	**0.0938**			60	**0.0547**	
SI1217	70	**0**		SX364	70	**0**	
	60	**0.0938**			60	**0.0508**	

Reversible watermarking is based on the process of watermark insertion into a medical image, transmission of the watermarked image, where the complete removal of the watermark from the image on the recipient's side is important and after watermark removal, the original image is completely restored and unchanged. In our case, since the quality of the decrypted speech signal is accepted and the SNR is greater than the requested value (20 dB), the removal of the watermark is not necessary to be reversible. The only condition on the watermark is that it does not affect the encrypted speech signal.

7.6 Time complexity analysis

Table 7 presents the elapsed time to accomplish the encryption/decryption operations using the proposed scheme on some speech signals.

We observe from Table 7 that the number of seconds taken by the proposed Algorithm to complete the encryption/decryption process is less than the time duration of the speech signal (see Table 1). So, we can judge that the proposed scheme works in real time. This can be explained by the well exploitation, and not costly, of the computing machine performances. Figure 13 illustrates the speech signal durations in addition to the two graphs with different colors represents the time variation of the two operations (encryption and decryption). We can deduce from this figure observation, that the length of speech signal can slightly affects the Algorithm execution time with a proportional relation, when the speech signal length increases the needed time for its processing increases with a real time treatment.

7.7 Comparisons

From previous results, we confirmed that the proposed scheme offers excellent results and we can stand on them. For more substantiation that our design merits further interest and may be considered among the best methods, we try to compare it with other recently published strong approaches.

Basing on results illustrated in Table 8, the proposed method seems well again in many records than other methods used in the comparison and too close in other records. The correlation coefficient between the original and the encrypted speech signal in the proposed

Table 7 Elapsed times for encryption/decryption operations

Speech signal	Elapsed time (sec)		Speech signal	Elapsed time(sec)	
	Encryption	Decryption		Encryption	Decryption
SI560	**1.547782**	**1.442497**	SI1303	**1.496006**	**1.362281**
SI734	**1.618274**	**1.385894**	SI1308	**1.851431**	**1.690632**
SI770	**1.647364**	**1.467160**	SI1390	**1.651050**	**1.543154**
SI839	**1.604325**	**1.479764**	SI1460	**1.605215**	**1.500492**
SI860	**1.496299**	**1.410839**	SI1715	**1.468997**	**1.370247**
SI863	**1.598295**	**1.417768**	SI1992	**1.382709**	**1.300392**
SI943	**1.554790**	**1.318709**	SI2194	**1.557797**	**1.476601**
SI1103	**1.851177**	**1.709405**	SI2303	**1.502707**	**1.381959**
SI1109	**1.461102**	**1.456605**	SX29	**2.420631**	**1.911237**
SI1217	**1.794836**	**1.519344**	SX364	**1.642267**	**1.395422**

Bold entries indicate that these entries are obtained results.

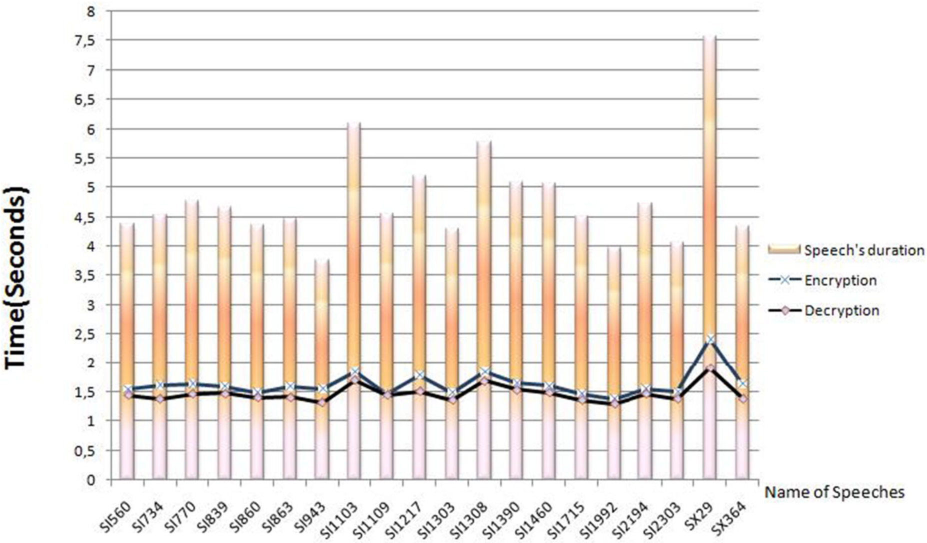


Fig. 13 Speech signal durations and the execution time variation of the two operations (encryption and decryption)

approach is classified second for its neighboring to zero just following the method proposed in [26]. The rest of values are also close to zero indicating good quality encryption.

But the correlation coefficient in the proposed scheme between the original and the decrypted speech signal extracted at the receiver is observed the best with the value of one, which means that there is no difference between the original and the decrypted speech signal. In addition, the proposed approach has the preference of the farthest value from one compared to other methods. The very robust methods presented in [12, 15], show significant SNR values between the original and the decrypted speech signal. The proposed scheme comes following giving a SNR value of 34.08 dB; But the SNR between the original and the encrypted signal is the smallest in the proposed scheme which demonstrates that the encrypted signal is very far from the original speech signal compared to other methods.

Table 8 Comparison between the proposed approach and seven published methods

	Corr coef (Original, Encrypted)		Corr coef (Original, Decrypted)		SNR dB (average)	
	Nearest to 0	Farthest to 0	Nearest to ±1	Farthest to ±1	Original, Decrypted	Original, Encrypted
Proposed scheme	**-5.555×10^{-4}**	**0,0073**	**1.00**	**0.9990**	**34,081**	**−35,514**
Method in [09] (speech)	0.0119	0.0384	1.00	0.999	33,523	/
Method in [27] (audio)	0.0010	−0.0060	0.9962	0.7317	/	/
Method in [1] (audio)	5.30×10^{-5}	5.7×10^{-4}	/	/	/	−33,2875
Method in [26] (speech)	0.0312	−0.0974	0.9958	0.9899	123.57	/
Method in [15] (audio)	−0.0018	0.0087	0.9989	0.6405	/	/
Method in [7] (speech)	0.00136	0.00992	/	/	/	−17,485
Method in [12] (speech)	0.0002	0.0209	1	1	50.01	/

Bold entries indicate that these entries are obtained results.

8 Conclusion

In this work, a novel scheme for securing speech signals using three approaches: Chaotic generator (tent and logistic maps) for producing a random vector by some initially introduced values to be merged with the original speech signal values, secondly the watermarking is included inside the encrypted signal for the purpose of verification during decryption process that the encrypted signal is authenticated and does not undergo external attacks; The third process Arnold scrambling key (cat map) is used to disperse signal samples by a secret key, and recovering the original signal from samples is not achievable without this key. As a result, we can say that the larger key space is a measure of better encryption and the obtained correlation value in the proposed scheme is nearer to zero which shows that original and encrypted signals are totally uncorrelated. Also, we recovered the original speech without affecting the quality.

Declarations

Conflict of interest/Competing interests no conflicts of interest.

Participants no other research participants except authors that are all consent.

References

1. Abdelfatah RI (2020) Audio encryption scheme using self-adaptive bit scrambling and two multi chaotic-based dynamic DNA computations. IEEE Access 8:69894–69907. https://doi.org/10.1109/ACCESS.2020.2987197
2. Al-Hooti M, Ahmad T, Djanali S (2019) “Developing audio data hiding scheme using random sample bits with logical operators”, Indonesian J Electr Eng Comput Sci. https://doi.org/10.11591/ijeecs.v13.i1.pp147-154.
3. Dhar PK, Shimamura T (2015) “Advances in Audio Watermarking Based on Singular Value Decomposition”, Springer briefs in electrical and computer engineering. https://doi.org/10.1007/978-3-319-14800-7.
4. Elsafty AH, Tolba MF, Said LA, Madian AH, Radwan AG (2020) Enhanced hardware implementation of a mixed-order nonlinear chaotic system and speech encryption application. Int J Electron Commun 125: 153347. https://doi.org/10.1016/j.aeue.2020.153347
5. Fantacci R, Menci S, Micciullo L, Pierucci L (2009) A secure radio communication system based on an efficient speech watermarking approach. Security and Communication Networks 2:305–314. https://doi.org/10.1002/sec.70
6. Farsana FJ, Gopakumar AK (2016) “A Novel Approach for Speech Encryption: Zaslavsky Map as Pseudo Random Number Generator”, 6th International Conference on Advances InComputing& Communications (Procedia Computer Science 2016) https://doi.org/10.1016/j.procs.2016.07.302.
7. Farsana FJ, Gopakumar K (2020) “Speech Encryption Algorithm Based on Nonorthogonal Quantum State with Hyperchaotic Keystreams”, Hindawi Advances in Mathematical Physics https://doi.org/10.1155/2020/8050934.
8. Farsana FJ, Devi VR, Gopakumar K (2020) An audio encryption scheme based on fast Walsh Hadamard transform and mixed chaotic keystreams. Appl Comput Inf. https://doi.org/10.1016/j.aci.2019.10.001
9. Huang CK (2009) H.H. Nien “multi chaotic systems based pixel shuffle for image encryption”. Opt Commun 282:2123–2127. https://doi.org/10.1016/j.optcom.2009.02.044
10. Joshi Subir Dr Amit M (2016) “DWT-DCT based Blind Audio Watermarking using Arnold Scrambling and Cyclic Codes”, 3rd International Conference on Signal Processing and Integrated Networks. https://doi.org/10.1109/SPIN.2016.7566666.
11. Jovic B (2011) Chaotic Signals and Their Use in Secure Communications. In: Synchronization Techniques for Chaotic Communication Systems. Signals and Communication Technology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21849-1_2

12. Kaur G,Singh K, Gill HS “Chaos-based joint speech encryption scheme using SHA-1”, Multimedia tools and applications (2021). https://doi.org/10.1007/s11042-020-10223-x.
13. Khanzadi H, Eshghi M, Borujeni SE (2014) Image encryption using random bit sequence based on chaotic maps. Arab J Sci Eng. https://doi.org/10.1007/s13369-013-0713-z
14. LalithaCh NV, JayaSree SRPVY (2013) "DWT-Arnold Transform Based Audio Watermarking", IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics. https://doi.org/10.1109/PrimeAsia.2013.6731204.
15. Lima JB, da Silva Neto EF (2015) Audio encryption based on the cosine number transform. Multimedia Tools Appl. https://doi.org/10.1007/s11042-015-2755-6
16. Lin Y, Abdulla WH, “Audio watermark: a Comprehensive Foundation using MATLAB”, springer Cham (2015). https://doi.org/10.1007/978-3-319-07974-5.
17. Liu H, Zhao B, Huang L (2019) Quantum image encryption scheme using Arnold transform and S-box scrambling. Entropy. https://doi.org/10.3390/e21040343
18. Merrad A, Saadi S (2018) Blind speech watermarking using hybrid scheme based on DWT/DCT and sub-sampling. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-5939-z
19. Merrad A, Saadi S, Benziane A, Hafaifa A (2018) Robust Blind Approach for Digital SpeechWatermarking. In 2nd International Conference on Natural Language and Speech Processing. https://doi.org/10.1109/ICNLSP.2018.8374366.
20. Mosa E, Messiha N, Zahran O, El-Samie FEAbd (2011) “Chaotic encryption of speech signals”, Int J Speech Technol https://doi.org/10.1007/s10772-011-9103-7.
21. Nematollahi SAR Al-Haddad (2013) “An overview of digital speech watermarking”. Int JSpeech Technol. https://doi.org/10.1007/s10772-013-9192-6.
22. Ratner B (2009) The correlation coefficient: its values range between + 1 / − 1, or do they ? J Target Meas Anal Mark 17:139–142. https://doi.org/10.1057/jt.2009.5
23. Revathi A, Sasikaladevi N, Jeyalakshmi C (2018) Digital speech watermarking to enhance the security using speech as a biometric for person authentication. International Journal of Speech Technology 21:1021–1031. https://doi.org/10.1007/s10772-018-09563-9
24. Saadi S, Merrad A, Benziane A (2019) Novel secured scheme for blind audio/speech norm-space watermarking by Arnold algorithm. Signal Proc. https://doi.org/10.1016/j.sigpro.2018.08.011
25. Sathiyamurthi P, Ramakrishnan S (2017) “Speech encryption using chaotic shift keying for secured speech communication”, EURASIP Journal on Audio, Speech, and Music Processing https://doi.org/10.1186/s13636-017-0118-0.
26. Sathiyamurthi P, Ramakrishnan (2020) “Speech encryption algorithm using FFT and 3D-Lorenz–logistic chaotic map”, Multimed Tools Appl https://doi.org/10.1007/s11042-020-08729-5.
27. Shah D, Shah T, Ahamad I, Haider MI, Khalid I (2021) A three-dimensional chaotic map and their applications to digital audio security. Multimed Tools Appl. https://doi.org/10.1007/s11042-021-10697-3
28. Sheu LJ (2011) “A speech encryption using fractional chaotic systems”, Nonlinear Dyn https://doi.org/10.1007/s11071-010-9877-1.
29. Wangab B, Dong XC (2016) On the novel chaotic secure communication scheme design. Commun Nonlinear Sci Numer Simul 39:108–117. https://doi.org/10.1016/j.cnsns.2016.02.035

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.