



Speech compression and encryption based on discrete wavelet transform and chaotic signals

Abbas Salman Hameed¹ 

Received: 6 April 2020 / Revised: 6 October 2020 / Accepted: 22 December 2020 /
Published online: 17 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

To increase transfer and storage efficiencies of the information, data compression has emerged as a significant issue in the communication environments. This paper introduces compression and encryption of speech signals based on Discrete Wavelet Transform (DWT) and Chaotic signals. DWT sparsens and codes the speech signal to the wavelet coefficients. The less impactful coefficients are eliminated to reduce the amount of data. After that, a new coding process which utilizes the chaotic signals is proposed to encode, in encrypted form, the residual coefficients. A High strength to the encryption process is realized by using four linked Hénon Chaotic Maps (HCM) in the proposed scheme. Multi HCM guarantees larger than 10^{240} of key space to the encryption process. The proposed system obtains up to -41.449 dB of spectral segmental signal-to-noise ratio, which measures and proves the strength of encryption. Also, at 10% compression ratio, signal-to-noise ratio of 11.549 dB and perceptual evaluation speech quality of 3.02945 demonstrate that the proposed system has high quality and intelligibility of the reconstructed speech.

Keywords Speech compression · Speech encryption · Wavelet transform · Chaotic map

1 Introduction

Speech is the most operative medium used in telephony, mobile communications and transmissions. Speech compression is one of the means that attempts to exploit all the available capabilities and resources of the communication systems. Compression is made by reducing the size or bit rate of the transmitted speech signal components [2, 4]. This process saves bandwidth of the communication channel. It also decreases the memory space which is needed for storage speech files. Speech compression is done according to the fact that a large number of redundant information is originated in the speech signals. By coding necessary speech

✉ Abbas Salman Hameed
abbasfuture@yahoo.com

¹ College of Engineering, University of Diyala, Diyala, Iraq

information and neglecting non-essential information, a compressed signal is generated. The amount of discarded information must be suited to the level which is desired to restore the original speech with high intelligibility. Today, speech compression is used in many different applications such as voice mail, video teleconferencing systems, satellite and cellular communications [10]. For the purpose of preparing the speech signal to compress, some transforms such cosine and wavelet are utilized [11, 14, 19, 31]. These transforms have the ability to deal with speech's time and frequency domains with high resolution.

Discrete wavelet transform decomposes the speech signal by decorrelating its samples into sets of coefficients. Many of these coefficients are almost zero [3, 27]. Thereby, the compression can be performed by coded effective coefficients only and give zero to the other coefficients when achieves a decompression process. Hence, wavelet transform based compression is still an interesting field by the researchers. Supatinee K. et al., [14] examined the application of Haar, Biorthogonal and Discrete Meyer wavelets to compress speech signals. The experiments indicated that the Biorthogonal wavelet provides a good compression ratio and quality of the reconstructed signal compared with two others. The enforcement of Coiflet wavelet based on speech compression was tested by Snehanka G. et al., [19]. This paper seeks to get a high auditory quality to the recovered speech. Fatma z. Chelali et al., [3] used DWT for speech compression and denoising. The results showed that the compression of an acoustic signal using DWT outperforms compared with use of discrete cosine transform. In [27], speech compression using a hybrid wavelet is proposed by Rekha V., and Sachin S. Chauhan. The held energy for the speech frames coefficients, which sets as a threshold, controls the required compression levels. Another compression algorithm for a distributed speech recognition system is proposed by Syu-Siang W. et al., [29]. This algorithm uses suppression by selecting wavelets to achieve compress and efficient data transmission. DWT filters speech signals into two frequency levels. The low frequency level is kept then transmitted while the high frequency level is discarded.

To get an increment in the compression ratio of the signal compared with existing schemes based on wavelet, an interesting compression approach called Compressive Sensing (CS) has been used recently. This approach introduces reduction in acquisition time and complexity. But this scheme has a challenge represented by that the processing power at the decoder limits the reconstruction quality of the signals [6, 13, 21, 25]. In [20], Viniitha R. et al., proposes a compression system based on CS to enhance and compress speech signals. Improving compression ratio is presented through a scheme proposed by Maher K. M. Al-Azawi et al., [1]. CS and chaotic system are utilized. All these compression algorithms intend to increase the compression level while retaining as much as possible of the quality and intelligibility of the recovered speech signal.

In this paper a high-quality system of compression and encryption of a speech signal based on DWT and Hènon chaotic maps is proposed. The excellent sparsen process of the signal which is produced by multi-level wavelet decomposition guarantees a high compression ratio after thresholding. Also, the new efficient coding process of the remaining wavelet coefficients using chaotic signals upsurge the auditory quality of the reconstructed speech and reduce the appended information required to decompress speech signal. The substantial characteristics for the eight chaotic signals assurance high security level to the compressed speech. A unified framework of compression and encryption processes is provided by using a combination of DWT and chaotic signals in the proposed approach.

The paper sets as follows; Section 2 shows DWT and Hènon chaotic map. Section 3 illustrates the proposed system of the speech compression and encryption. Finally, the simulation results and conclusions are presented in Sections 4 and 5 respectively.

2 Discrete wavelet transform and chaotic map

2.1 Discrete wavelet transform (DWT)

Wavelet transform applies different window scales to split the data into various ranges of frequency components. It captures time location and frequency information of the signals with high resolution [5, 28]. The coefficients (W) of the DWT which are computed for a signal $S[n]$ is defined by the equation [28].

$$W(j, k) = \sum_n S[n] 2^{-\frac{j}{2}} \psi(2^{-j}n - k), \quad (1)$$

where j and k are the scale and shift parameters, respectively.

$\psi(t)$ is called the mother wavelet. There are many wavelet families, each one is characterized by its mother wavelet shape.

Haar is the oldest, simplest, an orthogonal wavelet and it has linear phase characteristic. Haar wavelet owns one vanishing moment with two filter coefficients and it doesn't use the overlapping windowing technique [17, 28]. The mother Haar wavelet is given by the equation.

$$\psi(t) = \begin{cases} 1 & 0 < t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

Daubechies- p (db) and Coiflets- p (coif) are another orthogonal wavelet with longer compactly supported length than that of haar. These families use overlapping windows to decompose the data samples. The daubechies filters have $2p$ coefficients while coiflets filters have $6p$ coefficients. Therefore, these families deal with each $2p$ and $6p$ adjacent data element respectively. The results of windowing processes produce a smoother representation in the wavelet domain to the signal than in haar. One more wavelet family which has two different wavelet functions is Biorthogonal wavelet (bior). It is orthogonal to the shifted base function under different scale factors. But for the same scale factor, it is not orthogonal [14, 19]. Figure 1 shows haar, db4, coif1, and bior2.2 mother wavelet families respectively.

To decompose the signal, DWT passes the input data through successive low and high pass filters which have dissimilar cutoff frequencies. This process produces an orthogonal set of wavelets which have almost zero information components. In multilevel, DWT analyzes the data into approximation and detail coefficients by pushing those data through filters. Down-sampling by two is carried out to characterize the wavelet signal [7, 16, 30]. Therefore, the time resolution is halved while frequency resolution is doubled. Approximation coefficients are decomposed and subsampling again for each next level. Mathematically, the output coefficient vector can be written as

$$W(n) = \begin{cases} \left(\sum_{k=-\infty}^{\infty} S[k] L[2n-k] \right) \downarrow 2, \\ \left(\sum_{k=-\infty}^{\infty} S[k] H[2n-k] \right) \downarrow 2, \end{cases} \quad (3)$$

Where S , L , and H are the input signal, low pass and high pass filters respectively.

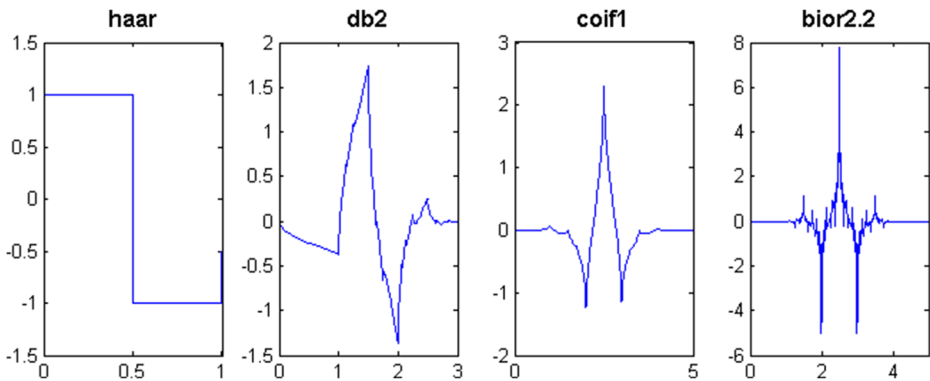


Fig. 1 Mother functions of several wavelet families

Several of one-dimension H and L filters are applied to obtain a two-dimensional (2D) wavelet (input data is a 2D matrix). First, each row of the input data is pushed through the two filters. Then after downsampling, each column of the resulting coefficient matrix is passed again through the filters. Three details subbands which represent highest resolution wavelet coefficients and one approximation subband which represents smooth coefficients of the original data. To conduct more analysis of the signal features with different scales, the last one subband is further decomposed through the next level [12, 15].

2.2 Hénon chaotic map

Chaotic maps produce deterministic sequences which have many unique properties such sensitivity to its parameters, noise-like behavior, and ergodicity. Thus, the chaotic signal confers additional confidentiality when the encryption scheme employs it [26]. In 1976, Michel Hénon introduced a chaotic map with two chaotic behavior signals. This map is defined as in the equation

$$\begin{aligned} x_n &= 1 - rx_{n-1}^2 + y_n, \\ y_n &= cx_{n-1}, \end{aligned} \quad (4)$$

where $(x_n, y_n) \in R$ are the generated chaotic sequences. r and c are the control parameters seed.

To guarantee chaotic performance of Hénon map signals, the control parameter may be having values $r \in (1.399, 1.4)$ and $c \in (0.299, 0.3)$ [18].

3 Proposed system of compression and encryption speech signal

Figure 2 shows the block diagram of the proposed compression and encryption system for the speech signals. This system utilizes DWT and Hénon chaotic signals to compress and encrypt speech signals. The speech signal is arranged into an almost squared matrix with respect to the length of speech signal. 2D spectrogram is constructed by applying the discrete cosine transform (DCT) for each column vector. DCT works as the first level of data sparsity. After that, multilevel 2D-DWT is applied on the spectrogram matrix to generate a wavelet coefficient matrix. A hard threshold value is chosen to attain the required compression level of the

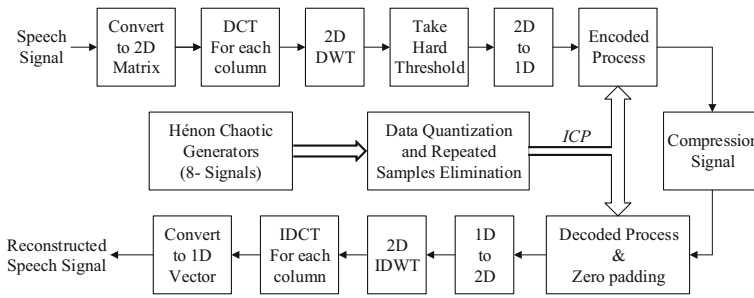


Fig. 2 The structure of proposed speech compression system

signal. All the details wavelet coefficients have values less than the threshold are reset to zero. The matrix is converted to 1D vector (W) to prepare it for the compression step. A new coding process is suggested for the compression scheme. This process is used to compress and encode the significant wavelet coefficients. Eight chaotic signals such $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ which are generated from four Hénon maps are employed for this mission. These signals are used to code each single, double, third, fourth, fifth, sixth, seventh, and eighth adjacent data samples respectively. The chaotic signals are joined together to assure high quality randomness of these signals. The joining process is accomplished by modifying ‘s equations such as

$$\left. \begin{aligned} y1_n &= c1 (x1_{n-1} + y2_{n-1}), \\ y2_n &= c2 (x2_{n-1} + y3_{n-1}), \\ y3_n &= c3 (x3_{n-1} + y4_{n-1}), \\ y4_n &= c4 (x4_{n-1} + y1_{n-1}), \end{aligned} \right\} \tag{5}$$

where $c1, c2, c3,$ and $c4$ are the control parameters for each map respectively.

Quantization process is applied for each chaotic signal to be compatible with bits per sample of speech signal. Then, to ensure there are no repeated values of chaotic samples, the elimination process cancels each sample which has the same values within an instantaneous chaotic plane (ICP). ICP is a stream of the eight quantized chaotic signals that have a thousand samples of each signal started with an instantaneous sample to the next thousand samples of those signals. ICP can be given as

$$ICP = \begin{bmatrix} \tilde{x}1_n & \tilde{x}1_{n+1} & \cdots & \tilde{x}1_{n+1000} \\ \tilde{y}1_n & \tilde{y}1_{n+1} & \cdots & \tilde{y}1_{n+1000} \\ \tilde{x}2_n & \tilde{x}2_{n+1} & \cdots & \tilde{x}2_{n+1000} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}4_n & \tilde{y}4_{n+1} & \cdots & \tilde{y}4_{n+1000} \end{bmatrix}, \tag{6}$$

where n is an instantaneous sample shifted with the process, \tilde{x} and \tilde{y} are quantized chaotic signals.

The compression and encoding of wavelet coefficients are performed using ICP planes. The number of adjacent information in the wavelet coefficients decides which ICP components are to be chosen to represent code of this information. To clarify that, if there are A adjacent information components (up to eight) that have non-zero value in the wavelet coefficients located at a P position with respect to the instantaneous thousand components, therefore $ICP_{A,P}$ is chosen to encode these information components. As an example, if the wavelet coefficients have a stream of data, such $\dots, 0, 0, W77, W78, 0, \dots$, the value in the 2nd

row (two non-zero adjacent samples) -77th column of ICP (\tilde{y}_{177}) is chosen to be coded. The compressed signal will be contained $\tilde{y}_{177}, W77, W78$. Figure 3 illustrates an example to the compression and encoding process.

After each thousand wavelet coefficients, a sample with zero value is inserted in the streams of the compressed signal. This sample identifies ICP length and is used to revive the decoding process if any error in the compressed signal samples occurs. To retrieve the speech signal from compressed data, the inverse steps of the compression process are performed at the receiver side. If the decoder receives $ICP_{A,P}$, the next A^{th} samples is put in the position started with P^{th} of the decoding vector. All next samples are set to zero until the next P^{th} position of the $ICP_{A,P}$ is detected. As an example, if the decoder receives data beginning with a sample which has the same \tilde{x}_{465} value, the next seven samples are put with position starting with 65 to 71 in the decoding vector. Then the eighth sample, next to the seven samples mentioned before, is compared with the ICP sample to know the corresponding position and data size. After whole this process is done, two-dimensional inverse Discrete wavelet (2D-IDWT) and inverse Cosine (ICT) transforms are applied respectively to the generated decoding vector.

4 Simulation results

Different speech files which are obtained from ‘NOIZEUS’, ‘CMU_Arctic’ and ‘TIMIT’ speech databases are used to experience the performance of the proposed speech compression system. All tested signals have samples of 16 bits. Eight level DWT- haar family (basically) is applied to get wavelet coefficients. As well as, four linked H enon maps which have different initial conditions and control parameters are utilized to generate eight chaotic signals. To prepare chaotic signals to the encoding process, all signals are quantized to 2^{16} levels respected to speech bit per sample (16 bits for speech files which are tested). Repeated samples in ICP , with dimensions 8×1000 , are eliminated to ensure correct reconstruction of speech signal at the receiver. Figure 4a shows y_2 samples, as an example, corresponding to the samples which have similar values in the other quantized chaotic signals. The red points indicate the similarity of the y_2 samples with that of other chaotic signals. Figure 4b shows the same relation after the eliminating process. So, it can be seen that the samples which have similar values within ICP are eliminated. The ICP is ready now to be used for the coding process.

Each one of the chaotic signals assigned to encode a specific group which have a certain number of adjacent samples in the wavelet coefficient samples. Up to eight adjacent samples can be encoded corresponding to eight chaotic signals. Figure 5 depicts the number of adjacent samples in the wavelet coefficients with respect to compression level for the $Sp21$ speech file in NOIZEUS database.

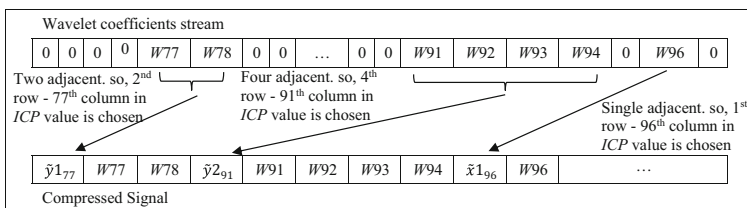


Fig. 3 An example of compression and encoding process

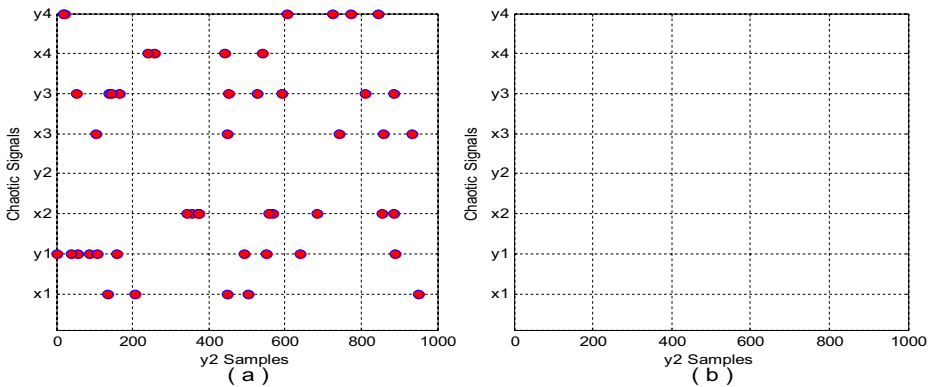


Fig. 4 ICP plane before and after repeated samples elimination process

Many performance statistical measures are used to evaluate the proposed speech compression system listed in the next subsection.

4.1 The performance measures

The objective measures are useful to measure the residual intelligibility and quality of compressed speech and retrieved signal respectively. The quality of the retrieved speech signal is mostly measured by Signal to Noise Ratio (SNR) [1]. High value of SNR corresponds with high quality of recovered speech. Higher quality of retrieved speech can be also measured by a higher value of another objective factor known as Peak Signal to Noise Ratio (PSNR) [10]. Perceptual Evaluation of Speech Quality (PESQ) [8, 24] is an accurate international standard factor for estimating speech quality. PESQ became a worldwide industry standard test for the applications which enhance speech quality used by voice processing and telephone networks. Moreover, Segmental Spectral Signal to Noise Ratio (SSSNR) [1] indicates the amount of residual intelligibility of encoded speech signal. The more negative value of SSSNR means

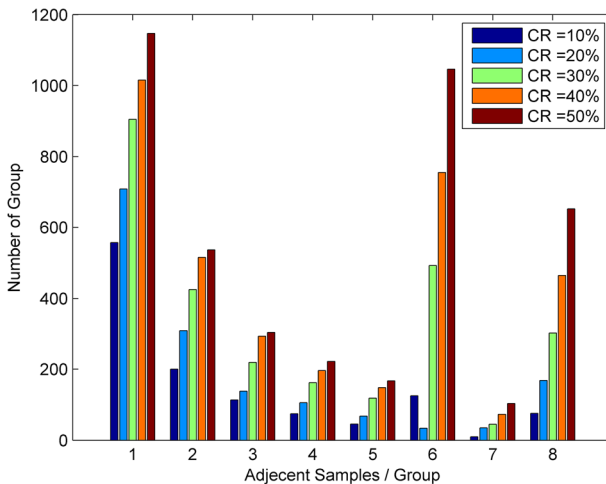


Fig. 5 Number of adjacent samples in the wavelet coefficients

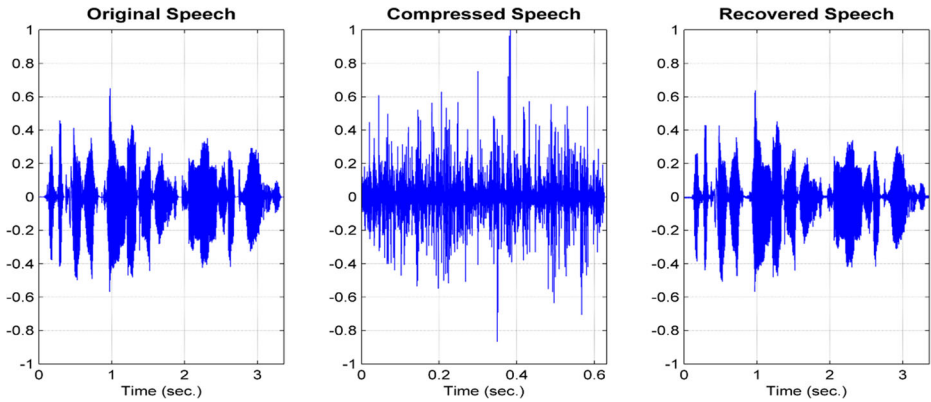


Fig. 6 Waveform of original, compressed, and recovered speech signals

more strength of the encryption process. Furthermore, Correlation coefficient (CF) [23] is a statistical measure used to test the signals similarity. CF has values between +1 to -1. The near zero value for CF means a large difference between the signals. When the CF value is almost one, the similarity is confirmed. Finally, Number of Non-Zero Coefficients (NNZC) before thresholding process and Number of Wavelet Coefficients (NWC) are suggested here. NNZC is applied to compute the percentage ratio of non-zero coefficients (coefficients which are processed to acquire a compressed signal) to the total coefficients before thresholding process. NWC is employed to compute the percentage ratio of increment in the number of decomposition coefficients for a wavelet family to that in haar wavelet family. NNZC and NWC are given as in the following equations:

$$NNZC = \frac{\text{number of non-zero before threshold}}{\text{total elements}} * 100\% . \tag{7}$$

$$NWC = \left(\frac{\text{number of coeff. of a family}}{\text{number of coeff. of haar wavelet}} - 1 \right) * 100\% . \tag{8}$$

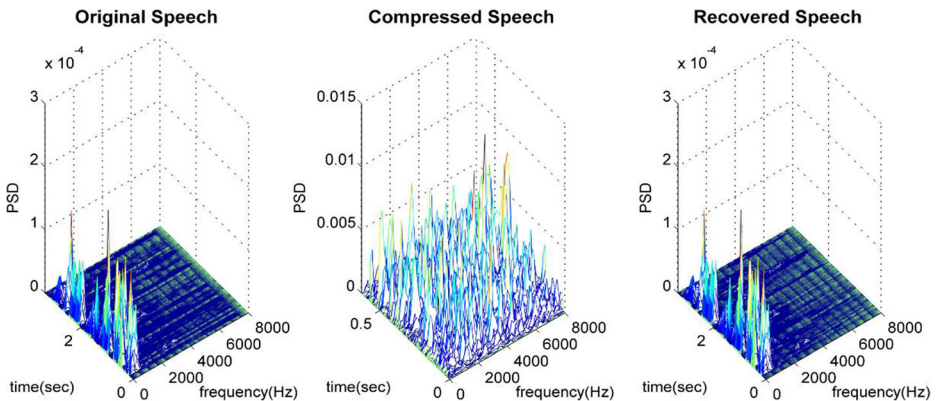


Fig. 7 Spectrogram of original, compressed, and recovered speech signals

Table 1 Simulation results of proposed compression speech system

Speech	CR%	SNR (dB)	PSNR (dB)	PESQ	CF	SSSNR (dB)
NOIZEUS Database	10	11.5496	58.21	3.02945	0.96437	- 39.6507
	20	19.2357	59.33	3.26646	0.99402	- 41.449
	30	24.2552	66.20	3.55866	0.99812	- 40.995
	40	28.9898	69.72	3.83787	0.99937	- 40.554
	50	33.8277	72.94	3.97434	0.99979	- 38.742
CMU_Arctic Database	10	14.3461	55.60	2.83454	0.98145	- 29.4533
	20	20.1368	59.46	3.29377	0.99514	- 28.9119
	30	24.9781	62.68	3.48302	0.99841	- 27.4141
	40	29.0439	65.39	3.67659	0.99937	- 27.9536
	50	33.3375	68.26	3.91564	0.99976	- 26.4618
TIMIT Database	10	14.1722	60.08	2.80504	0.98068	- 30.1716
	20	20.3514	64.46	3.16686	0.99537	- 29.9337
	30	25.6765	67.75	3.49660	0.99864	- 29.4170
	40	30.7095	71.11	3.80355	0.99957	- 27.9209
	50	35.6917	74.43	4.06825	0.99986	-26.2443

All these statistical measures are computed with respect to Compression Ratio (CR). CR is used to obtain the percentage ratio of the size of compressed signal to that of the original speech signal.

4.2 The results of proposed speech compression system

The performance of the proposed system illustrates in this subsection. Figures 6 and 7 show the waveform and spectrogram of the original, compressed and recovered speech signals respectively. Acoustically, these figures depict high intelligibility and quality of the retrieved signal at a high compression level (CR=18%). The compressed signal is like noise and it is dissimilar with respect to the original speech. This analysis is supported by the statistical results that are shown in Tables 1 and 2.

Table 1 indicates the SNR, PSNR, PESQ, and CF results for the retrieved speech signals and SSSNR for the encrypted and compressed signal all with various CR levels. By observation, it is found the high values of all simulated objective measures. High CR results are realized by the efficient encoding of wavelet coefficients. This process is accomplished with minimum information which is required to retrieve the speech signal. Also, the linear phase property of the haar wavelet attains a good reconstructing to the speech signal. That is clearly

Table 2 Comparison results of proposed compression system based on various wavelet families

Wavelet family	SNR (dB)	PSNR (dB)	PESQ	CF	NNZC (%)	NWC (%)
haar/db1/bior1.1	24.255	66.20	3.55866	0.99812	98.5	–
	20.309	63.93	3.37155	0.99533	100	5.77
db3	16.034	60.88	3.23349	0.98651	100	11.20
db4	11.610	58.13	3.05760	0.96487	100	18.21
coif1	16.869	61.64	3.32832	0.98966	100	11.21
coif2	5.887	54.32	2.68542	0.86151	100	31.68
bior1.3	16.809	61.60	3.33020	0.98952	100	11.27
bior1.5	6.289	54.58	2.75333	0.87510	100	23.55
bior2.2	15.657	60.83	3.25382	0.98632	100	11.23

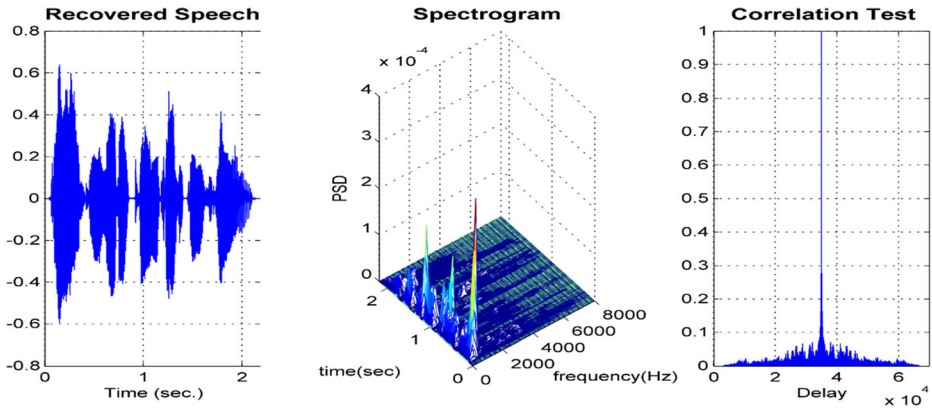


Fig. 8 Waveform, spectrogram and correlation test without any change in Hènon parameters

by these results which reflect high quality for the reconstructed speech. As well, Low SSSNR values (gets between -41.449 and -26.4618) confirm the strength of the encryption process and refer to high level immunity against any attacks.

To test the effects of applying another wavelet family, Table 2 represents the comparison results of using some types of db, coif, and bior wavelet instead of haar to compressing a speech from the NOIZEUS database at CR = 30%. From this Table and except haar wavelet, NNZC results indicate that all coefficients of the decomposition process which are produced by db, coif, or bior multilevel 2D-DWT haven't zero value before the thresholding process. This fact is a result of a highly smoothing representation of the signal by inherent overlapping windowing property for these families. But it leads to loss of more information through thresholding. Also, the NWC results indicate that the db, coif, and bior wavelets produce more coefficients (gets between 5% to 31%) compared to the coefficients which haar produces. Where those families have many samples in its FIR L and H filters compared with haar that has two samples only. For these reasons, the execution of haar wavelet in the proposed scheme appears superior in terms of SNR, PSNR, PESQ, and CF with respect to CR.

Figures 8 and 9 show the waveform, spectrogram, and correlation of the recovered speech signal with and without 10^{-15} change in $r1$ parameter value respectively, all with CR=40%.

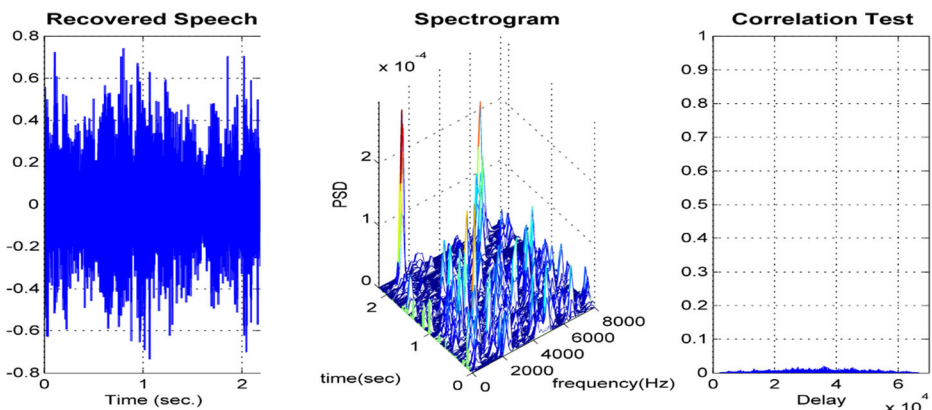


Fig. 9 Waveform, spectrogram and correlation test with $\pm 10^{-15}$ change in $r1$ parameter only

Table 3 SNR and CF of retrieved speech when tiny change in a control parameter

Speech	CR%	SNR (dB)	CF
NOIZEUS Database	10	4.6801-	0.0015
	20	-4.7572	0.0021
	30	-4.7103	0.0016-
	40	-5.0203	0.0023
	50	-4.6894	0.0013
CMU_Arctic database	10	-4.3897	0.0007
	20	-4.7335	0.0001
	30	-4.3743	0.0004
	40	-4.8366	0.0013
	50	-4.6907	0.0046
TIMIT database	10	-4.5656	0.0031
	20	-4.8910	0.0018
	30	-5.0210	0.0021
	40	-5.0300	0.0018
	50	-5.0610	0.0008

The spectrograms and correlation test results depict huge differences between original and recovered speech. The waveforms of the recovered speech assert that also.

Table 3 displays the SNR and CF with various CR values at a tiny change in $r1$ control parameter ($\pm 10^{-15}$). The results of SNR and CF clearly reflect that change by $\pm 10^{-15}$ to a control parameter makes it impossible to retrieve the original speech signal. The low CF values indicate there is no relation between the original and recovered signals.

In the proposed system, Wavelet family, wavelet level, all the control parameters, and initial values of the four Hénon maps are exploited as secret keys. Generally, increasing of keys in an encryption system leads to a high key space of that system. Sixteen parameters of four Hénon maps give $(10^{15})^{16} = 10^{240}$ key space.

The performance results of the proposed compression and encryption system have been compared with schemes that are presented in [1, 11, 19, 20, 22, 27], and [9]. Table 4 sets forth results summary of CR, SNR, PSNR, and PESQ for the proposed system and some of these objective measures for the compared schemes. It is clear that the proposed compression system outperforms the other schemes for the same CR values. The SNR and PESQ values indicate that the proposed system can attain a high level of quality and intelligibility for the

Table 4 Comparison results of proposed compression system with various compression Schemes

CR%	Reference-Year	SNR (dB)	PESQ	PSNR (dB)
8.24	[22] -2019	6.30	2.88	N/A
	Proposed Scheme	10.058	2.915	N/A
10	[19] -2014	9.2014	N/A	14.9047
	[11] -2016	6.5955	N/A	14.9747
	[27] -2018	N/A	N/A	37
	Proposed Scheme	11.5496	3.029	58.21
40	[20] -2015	6.49	N/A	N/A
	[1] -2018	11.836	2.734	N/A
	[27] -2018	N/A	N/A	48
	Proposed Scheme	28.989	3.837	65.39
50	[20] -2015	8.79	N/A	N/A
	[1] -2018	18.874	3.422	N/A
	Proposed Scheme	33.827	3.974	68.26

reconstructed signal. Also, the encryption strength of the proposed system is confirmed by lower SSSNR = -38.74 dB compared with -20.78 and -14.834 for [9] and [1] respectively.

The sparse process of the speech information by the multilevel 2D-DWT and efficient proposed encoding of the valuable coefficients give the proposed compression scheme dominance in terms of compression ratio and quality of retrieved speech in comparison with the other schemes.

5 Conclusions

In this paper, the proposed system compresses and encrypts the speech signal simultaneously. Discrete wavelet Transform sparsens speech information and then the proposed coding process which is based on eight signals of Hénon map encodes the weighty coefficients. The simulation results show outperforms the proposed system in terms of SNR, PESQ, PSNR, CF, and SSSNR with respect to CR ratio. At low CR value equals to 10%, the results get SNR= 11.5496 dB, PSNR= 58.21 dB, PESQ= 3.02945 , CF= 0.96437 . These results reflect high intelligibility and quality of reconstructed speech signals. As well the proposed system guarantees high encryption strength with large key space for compressed speech. Where it can note that the SSSNR values (get between -41.449 and -26.4618 dB) are very low. Consequently, it is harder to extract the original speech signal by any intruder.

Authors' contributions Not applicable.

Data availability Not applicable.

Compliance with ethical standards

Conflicts of interest Not applicable.

Code availability Not applicable.

References

1. Al-Azawi MKM, Gaze AM (2018) Combined speech compression and encryption using chaotic compressive sensing with large key size. *IET Signal Processing* 12:214–218. <https://doi.org/10.1049/iet-spr.2016.0708>
2. Cernak M, Asaei A, Hyafil A (2018) Cognitive speech coding. *IEEE Signal Process Mag* 35:97–109. <https://doi.org/10.1109/MSP.2017.2761895>
3. Chelali FZ, Cherabit N, Djeradi A, Falek L (2018) Wavelet transform for speech compression and denoising. *International Conference on Multimedia Computing and Systems -Proceedings 2018-May:1–7*. <https://doi.org/10.1109/ICMCS.2018.8525996>
4. Dusan S, Flanagan JL, Karve A, Balaraman M (2007) Speech compression by polynomial approximation. *IEEE Trans Audio Speech Lang Process* 15:387–395. <https://doi.org/10.1109/TASL.2006.881705>
5. Graps A (1995) An introduction to wavelets. *IEEE Comput Sci Eng* 2:50–61. <https://doi.org/10.1109/99.388960>
6. Gunawan TS, Khalifa OO, Shafie AA, Ambikairajah E (2011) Speech compression using compressive sensing on a multicore system. 2011 4th international conference on mechatronics: integrated engineering for industrial and societal development, ICOM'11 - conference proceedings 17–19. <https://doi.org/10.1109/ICOM.2011.5937130>

7. Hameed AS (2017) Image encryption based on fractional order lorenz system and wavelet transform. *Diyala journal of engineering sciences* 10:81–91. <https://doi.org/10.24237/djes.2017.10108>
8. ITU-T (2001) Perceptual evaluation of speech quality (PESQ). ITU-T Recommendation P862 862:749–752
9. Jawad AK, Abdullah HN, Hreshee SS (2018) Secure speech communication system based on scrambling and masking by chaotic maps. *International conference on advances in sustainable engineering and applications, ICASEA 2018 - proceedings* 7–12. <https://doi.org/10.1109/ICASEA.2018.8370947>
10. Joseph SM, Anto PB (2011) Speech compression using wavelet transform. *International Conference on Recent Trends in Information Technology, ICRTIT 2011* 754–758. <https://doi.org/10.1109/ICRTIT.2011.5972258>
11. Joseph SM, Babu AP (2016) Wavelet energy based voice activity detection and adaptive thresholding for efficient speech coding. *International Journal of Speech Technology* 19:537–550. <https://doi.org/10.1007/s10772-014-9240-x>
12. Karajeh H, Khatib T, Rajab L, Maqableh M (2019) A robust digital audio watermarking scheme based on DWT and Schur decomposition. *Multimed Tools Appl* 78:18395–18418. <https://doi.org/10.1007/s11042-019-7214-3>
13. Katzberg F, Member S, Mazur R, et al (2018) A compressed sensing framework for dynamic sound-field measurements. *IEEE/ACM Transactions on audio, speech, and Language processing* PP:1. <https://doi.org/10.1109/TASLP.2018.2851144>
14. Kornsing S, Srinonchat J (2012) Enhancement speech compression technique using modern wavelet transforms. *Proceedings - 2012 international symposium on computer, consumer and control, IS3C 2012* 393–396. <https://doi.org/10.1109/IS3C.2012.106>
15. Lee Y, Seo Y, Kim D (2019) Digital blind watermarking based on depth variation prediction map and DWT for DIBR free-viewpoint image. *Signal Process Image Commun* 70:104–113. <https://doi.org/10.1016/j.image.2018.09.004>
16. Makbol NM, Khoo BE, Rassem TH (2016) Block-based discrete wavelet transform-singular value decomposition image watermarking scheme using human visual system characteristics. *IET Image Process* 10:34–52. <https://doi.org/10.1049/iet-ipr.2014.0965>
17. Mehra M (2018) *Wavelets theory and its applications*. Springer Nature Singapore
18. Meranza-Castillón MO, Murillo-Escobar MA, López-Gutiérrez RM, Cruz-Hernández C (2019) Pseudorandom number generator based on enhanced Hénon map and its implementation. *AEU - International Journal of Electronics and Communications* 107:239–251. <https://doi.org/10.1016/j.aeu.2019.05.028>
19. Narkhedkar SG, Patel PK (2014) Recipe of speech compression using coiflet wavelet. *Proceedings of 2014 international conference on contemporary computing and informatics, IC3I 2014* 1135–1139. <https://doi.org/10.1109/IC3I.2014.7019767>
20. Ramdas V, Mishra D, Gorthi SS (2015) Speech coding and enhancement using quantized compressive sensing measurements. *2015 IEEE international conference on signal processing, Informatics, Communication and Energy Systems, SPICES 2015* 2–6. <https://doi.org/10.1109/SPICES.2015.7091436>
21. Rani M, Dhok SB, Deshmukh RB (2018) A systematic review of compressive sensing: concepts, implementations and applications. *IEEE Access* 6:4875–4894. <https://doi.org/10.1109/ACCESS.2018.2793851>
22. Sankar MSA, Sathidevi PS (2019) A scalable speech coding scheme using compressive sensing and orthogonal mapping based vector quantization. *Heliyon* 5:e01820. <https://doi.org/10.1016/j.heliyon.2019.e01820>
23. Sheela SJ, Suresh KV, Tandur D (2017) A novel audio cryptosystem using chaotic maps and DNA encoding. *Journal of Computer Networks and Communications* 2017:1–13. <https://doi.org/10.1155/2017/2721910>
24. Stankovi L (2018) Analysis of the reconstruction of sparse signals in the DCT domain applied to audio signals. *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING* 9290: 1–18. <https://doi.org/10.1109/TASLP.2018.2819819>
25. Strohmer T (2012) Measure what should be measured: Progress and challenges in compressive sensing. *IEEE Signal Processing Letters* 19:887–893. <https://doi.org/10.1109/LSP.2012.2224518>
26. Su Y, Tang C, Chen X, Li B, Xu W, Lei Z (2017) Cascaded Fresnel holographic image encryption scheme based on a constrained optimization algorithm and Henon map. *Opt Lasers Eng* 88:20–27. <https://doi.org/10.1016/j.optlaseng.2016.07.012>
27. Vig R, Chauhan SS (2018) Speech compression using multi-resolution hybrid wavelet using DCT and Walsh transforms. *Procedia Computer Science* 132:1404–1411. <https://doi.org/10.1016/j.procs.2018.05.070>
28. Waldekar S, Saha G (2020) Analysis and classification of acoustic scenes with wavelet transform-based mel-scaled features. *Multimed Tools Appl* 79:7911–7926. <https://doi.org/10.1007/s11042-019-08279-5>

29. Wang SS, Lin P, Tsao Y, Hung JW, Su B (2018) Suppression by selecting wavelets for feature compression in distributed speech recognition. *IEEE/ACM Transactions on Audio Speech and Language Processing* 26: 564–579. <https://doi.org/10.1109/TASLP.2017.2779787>
30. Yu ZQ, Bin QS, Bo HY, Zhang T (2018) A high-performance speech perceptual hashing authentication algorithm based on discrete wavelet transform and measurement matrix. *Multimed Tools Appl* 77:21653–21669. <https://doi.org/10.1007/s11042-018-5613-5>
31. Zhao D, Ma SQ (2010) Speech compression with best wavelet packet transform and SPIHT algorithm. *ICCMS 2010–2010 International Conference on Computer Modeling and Simulation* 1:360–363. <https://doi.org/10.1109/ICCMS.2010.68>

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