Transactions on Machine Intelligence 4(3) (2021) 109-117

ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

Concealment of Speech Signals within Speech by Combining Cryptography and Steganography

J. Shirazi[1,*]

[1] Assistant Professor, Department of Electrical Engineering, Islamic Azad University, Gonabad Branch, Razavi Khorasan, Iran

ARTICLE INFO

Article History:
Received 4 June 2021
Received in revised form 16 July 2021
Accepted 5 September 2021
Available online 5 September 2021

Keywords:
Cryptography, Steganography, Scrambling, Discrete Cosine Transform, Discrete Wavelet Transform, Chaotic Mapping

ABSTRACT

The concealment of speech signals, particularly within other speech signals, encompasses the integration of cryptographic techniques and steganographic methods. This literature review synthesizes current research findings and identifies existing gaps in knowledge, while also suggesting potential future research directions. In this paper, to enhance data security in speech transmission, a combination of steganography and cryptography is utilized. In the encryption phase, the Discrete Cosine Transform (DCT) is applied to frames of the speech signal. This phase involves scrambling time-domain samples, transform coefficients, and time-domain samples obtained from the inverse transform of the signal using a chaotic function that generates random numbers. In the steganography phase, the encrypted data replaces the low-value coefficients of the Discrete Wavelet Transform (DWT) of the host speech signal. To evaluate the proposed hybrid method, both qualitative and quantitative metrics were used. The results indicate a high level of various metric measures for the method.

1. INTRODUCTION

The concealment of speech signals within other speech signals is a critical area of research in secure communications, particularly in contexts where privacy and data integrity are paramount. This literature review synthesizes current research findings on the integration of cryptography and steganography for enhancing the security of concealed speech signals. The review is structured to highlight the advancements in cryptographic techniques, the application of steganographic methods, and the implications of these findings for future research directions.

1.1. Cryptographic Foundations

Recent advancements in cryptography, particularly in the context of quantum computing, have significant implications for the security of speech signals. For instance, the exploration of post-quantum cryptography by the National Institute of Standards and Technology (NIST) suggests that new cryptographic algorithms could provide

* Corresponding Author: j_shirazi@iau-gonabad.ac.ir
Assistant Professor, Department of Electrical Engineering, Islamic Azad University, Gonabad Branch, Razavi Khorasan, Iran

http://dx.doi.org/10.47176/TMI.2021.109

robust frameworks for encrypting speech signals. This is crucial as the potential for quantum attacks on traditional public-key cryptography poses a threat to secure communications [1,2].

The hybrid encryption techniques exemplified in studies on image data can be adapted for speech signal encryption. For example, integrating elliptic curve cryptography (ECC) with symmetric algorithms like Advanced Encryption Standard (AES) allows for secure and efficient encryption of speech signals prior to their embedding within other audio signals [3]. Such hybrid methods enhance the security of concealed information, which is essential for ensuring that hidden speech signals remain imperceptible to eavesdroppers [4].

1.2. Steganographic Techniques

Steganography, the art of concealing information within other data, has evolved to include sophisticated methods for hiding information in audio signals. One notable approach is the use of Least Significant Bit (LSB) coding, which allows for the embedding of secret information in audio files without significantly altering the perceptual quality of the audio [5]. The incorporation of cryptographic techniques alongside LSB steganography has proven effective in enhancing the security of concealed messages, making this dual approach a focal point of research in this domain [6].

Moreover, advancements in deep learning have led to improved methods for analyzing and processing audio signals. For instance, deep neural networks (DNNs) can be employed for packet loss concealment in digital speech transmission, which is relevant when ensuring the intelligibility of concealed speech signals [7]. The ability to reconstruct missing frames enhances the quality of the concealed speech, thereby improving the overall effectiveness of steganographic techniques.

1.3. Integration of Cryptography and Steganography

The combination of cryptography and steganography presents a promising avenue for secure communication. Research indicates that integrating these two techniques can form a robust communication framework that addresses the vulnerabilities of each method when used independently. For instance, the application of filter bank ciphers for encryption, combined with discrete wavelet transforms for steganography, demonstrates how speech signals can be effectively concealed while maintaining high security [8].

This dual approach not only secures the speech signals from interception but also enhances the perceptual quality of the concealed information. The metrics for evaluating the effectiveness of these techniques, such as Peak Signal to Noise Ratio (PSNR) and entropy, can be applied to assess the quality of the concealed speech signals [9].

1.4. Knowledge Gaps and Future Research Directions

Despite the advancements outlined, there remain significant knowledge gaps in the application of cryptographic and steganographic techniques specifically tailored for speech signal concealment. One area that warrants further exploration is the development of lightweight cryptographic algorithms that can be efficiently implemented in resource-constrained environments, such as Internet of Things (IoT) devices [10].

Other research has also been conducted in this field, a few of which are listed here. In [11], encrypted image hiding within an audio/speech signal is achieved. The method employed hides the encrypted image within the transform coefficients of the audio signal. The method's efficiency against various attacks has been evaluated. In [12], encrypted audio signal hiding within an image is examined, considering various audio signals. In [13], using DCT and wavelet transform for encryption combined with a chaotic system to generate keys for scrambling time-domain samples and the transform is proposed. In [14], DCT is used for transformation and a chaotic system for generating random numbers for encryption. Similarly, in [15], DCT and a chaotic system are used for encryption, and the method's security and resistance to different attacks are confirmed using various quantitative metrics.

This research proposes a method to enhance the security of speech signal transmission by combining cryptography and steganography. For speech encryption, DCT and scrambling of time and frequency domain samples, along with adding random numbers generated by a chaotic system to the signal's transform samples, are used. For steganography, the encrypted speech signal is embedded within the DWT coefficients of the host speech signal. The second section of the paper discusses several encryption methods. The third section details the proposed method for signal encryption, the chaotic function used, the construction of the encrypted signal, and its embedding within the

host speech signal. The fourth section evaluates the proposed method, and the final section presents conclusions and recommendations.

2. INTRODUCTION TO VARIOUS SPEECH SIGNAL ENCRYPTION METHODS

Encryption of signals is based on scrambling or permutation of the signal in both frequency and time domains. Methods used in the frequency domain include frequency inversion algorithms, frequency shifting, permutation of frequency bands, permutation of discrete Fourier transform (DFT) samples, and the use of spread spectrum techniques.

In frequency inversion, frequency samples are arranged from end to beginning, resulting in an encrypted signal in the time domain. In frequency shifting, which can accompany frequency inversion, the frequency spectrum is shifted. In frequency band permutation, the spectrum is divided into sub-bands that are then rearranged irregularly. In spread spectrum methods, each sub-band is shifted to a part of the spread spectrum with a different carrier frequency. Other encryption methods involve performing the encryption in the domain of mathematical transforms. In these methods, the signal is first transformed, and then the resulting coefficients are scrambled. Finally, the inverse transform is applied to obtain the encrypted signal in the time domain. Permutation of Fourier or discrete cosine transform (DCT) samples and scrambling these samples are examples of such frequency domain encryption methods.

Time domain methods used for encryption include time inversion, domain permutation, and amplitude masking. Time coefficient inversion involves reversing the time-domain coefficients from end to beginning within each signal window. Permutation can be achieved by rearranging the coefficients within a signal window. Another approach is to divide each time window into several sub-sections and permute these sub-sections or the internal elements of the sub-sections.

To enhance security, combined time and frequency domain methods are used, known as two-dimensional speech encryption algorithms. Examples include frequency inversion-time domain permutation and frequency band permutation-time domain permutation algorithms.

Encryption in the domain of transforms is one of the most commonly used encryption methods. These methods, considered some of the best and most secure for signal scrambling, are applied to the discrete transforms of the speech signal. Widely used transforms in this method include the discrete cosine transform (DCT) and wavelet transform (DWT).

3. DESCRIPTION OF THE EMPLOYED ENCRYPTION AND STEGANOGRAPHY METHOD

3.1. Encryption Method

In this encryption method, the speech signal is sampled at a rate of 8 kHz, and the obtained samples are framed into 32-millisecond windows. In the next step, for each frame, the samples are first permuted based on random numbers derived from a logistic chaotic system. The discrete cosine transform (DCT) is then applied to the permuted samples, and random numbers from the chaotic function are added to the resulting transform coefficients. The transform coefficients are then permuted based on random numbers from the chaotic system, and the inverse discrete transform is applied to the scrambled transform coefficients to obtain the signal in the time domain. Subsequently, the time-domain samples of the frame are scrambled based on the aforementioned random number code. The time-domain signal obtained from these steps is the encrypted speech signal resulting from the applied encryption method. Figure 1 illustrates the stages of this speech encryption method. The random number code output of the chaotic system is periodically altered in a specific order to ensure the scrambling of transform and time-domain samples does not follow a uniform pattern, significantly increasing encryption complexity. This also enhances the scrambling of the encrypted signal's spectrum, reducing the likelihood of its detection from the signal spectrum.

3.2. Logistic Chaotic Function

Chaotic functions are highly sensitive to initial values, and any change in these values or the function parameters leads to significant changes in the function's behavior and the generated values. The logistic map is a very simple chaotic function used as a generator for producing random numbers in this paper. This map is described by equation 1. The parameters of this function are x0 and R, where x0 is chosen between zero and one, and R is chosen between zero and four [16]. The output of this map in each iteration is a random number between zero and one.

$$x_{n+1} = \mathrm{R}\, x_n\,(1 - x_n\,) \tag{1}$$

In this paper, x0 is set to 0.66 and R to 3.999.

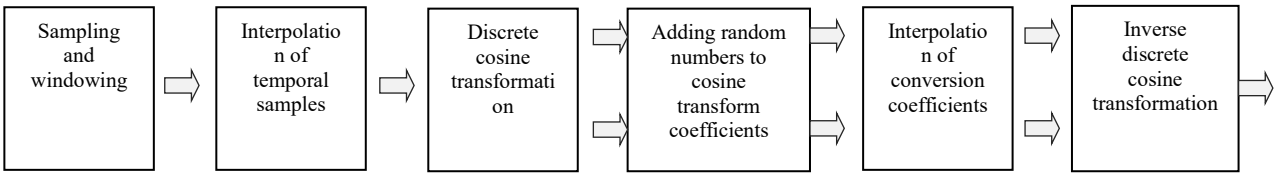

Fig.1. Steps of Creating an Encrypted Signal

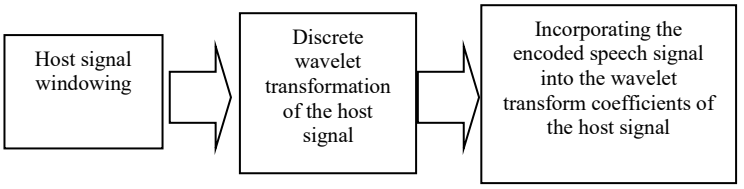

Fig.2. Steps of Steganography

3.3. Steganography Method

The steganography steps, as shown in Figure 2, are performed as follows:

1. The host speech signal is framed similarly to the initial speech signal at the transmitter.

2. For each host frame, a discrete wavelet transform (DWT) with a resolution level, which is the secret key between the sender and receiver, is applied to the discrete samples, and the wavelet coefficients are calculated.

3. The encrypted speech signal data string is embedded within the low-value wavelet coefficients of the host signal.

4. An inverse discrete wavelet transform is performed to obtain the final signal.

At the receiver, to reconstruct the original speech signal, the inverse of the steps applied at the transmitter is performed. To extract the encrypted message from the received signal, the receiver must have the initial random number code used at the transmitter and the code alteration key, which are the secret keys between the sender and receiver. During decryption, the receiver must know the number of wavelet decomposition levels, the specific level of substitution, and the substitution locations of the guest data string, which are also the secret keys between the sender and receiver.

4. EXPERIMENTS AND EVALUATION OF THE PROPOSED METHOD

To conduct experiments and evaluate the proposed method, we first tested the encryption phase. For this purpose, we used 10 diverse audio files. Each file was encrypted, decrypted without steganography, reconstructed, and stored. To evaluate this phase, we employed an auditory quality criterion by subjecting the original, encrypted, and reconstructed audio signals to a listening test by several individuals. The results consistently indicated that the

encrypted signal was incomprehensible, while the reconstructed signal was understandable and completely similar to the original signal. Figures 3 and 4 illustrate an example of the speech signal.

Figure 3 shows the time-domain plots of the original speech signal, the encrypted signal at the transmitter, and the reconstructed signal at the receiver. It can be observed that the time-domain plots of the original and reconstructed speech signals are identical, whereas the encrypted speech signal bears no resemblance to the original signal.

Figure 4 displays the spectra of the original speech signal, the encrypted signal at the transmitter, and the reconstructed signal at the receiver. The figure shows that the spectrum of the encrypted speech signal is scrambled, making it difficult to deduce the encryption method.

For steganography, we embedded the speech signal into the host signal with varying storage capacities and measured the transparency of the resulting signal by calculating the signal-to-noise ratio (SNR) using Equation 2.

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_n x^2(n)}{\sum_n \left[x(n)-y(n)\right]^2} \tag{2}$$

In this equation, x(n) represents the host speech string, and y(n) represents the final speech string containing both guest and host information. Table 1 shows the SNR results for different storage capacities. It is evident that as storage capacity decreases, the SNR increases.

Table 1. Signal-to-Noise Ratio Based on Storage Capacity

SNR	Capacity Percentage
9.9598	**41.4063**
10.7923	**35.1563**
11.1950	**29.6875**
12.3507	**23.4375**
15.6174	**11.7188**

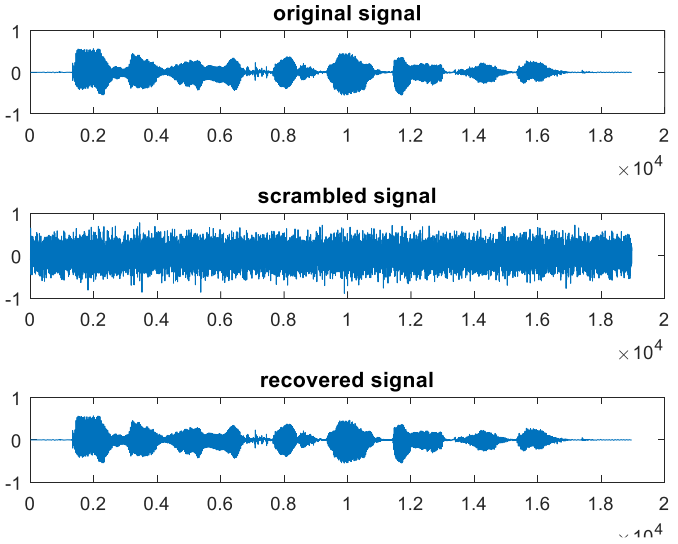

Fig.3. Time-domain plots of the original speech signal, the encrypted signal, and the reconstructed signal at the receiver without steganography

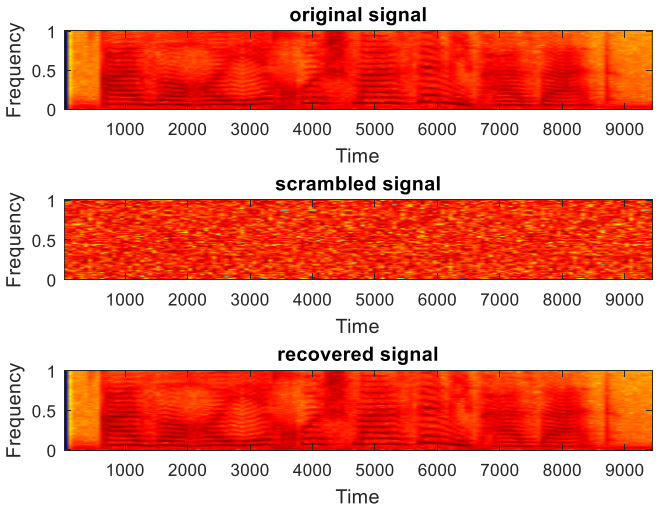

Fig.4. Spectra of the original speech signal, the encrypted signal, and the reconstructed signal without steganography

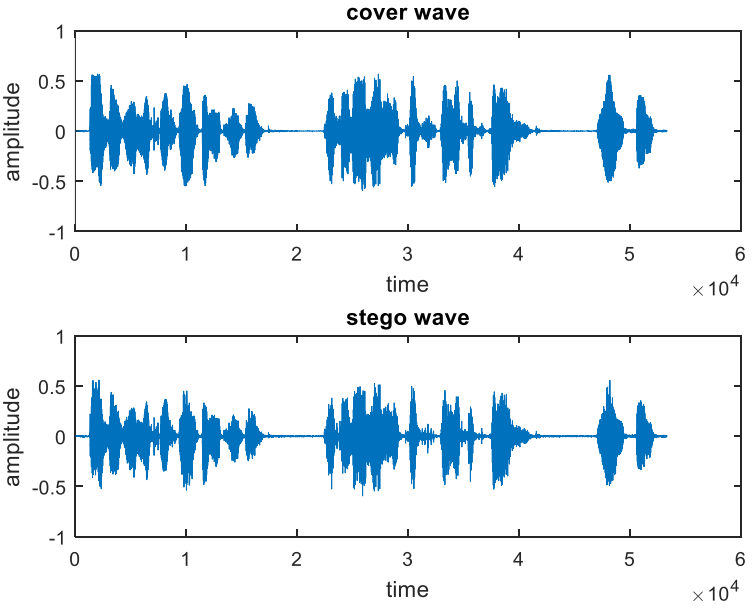

Fig.5. Time-domain plots of the host speech signal and the steganographically embedded speech signal at the transmitter

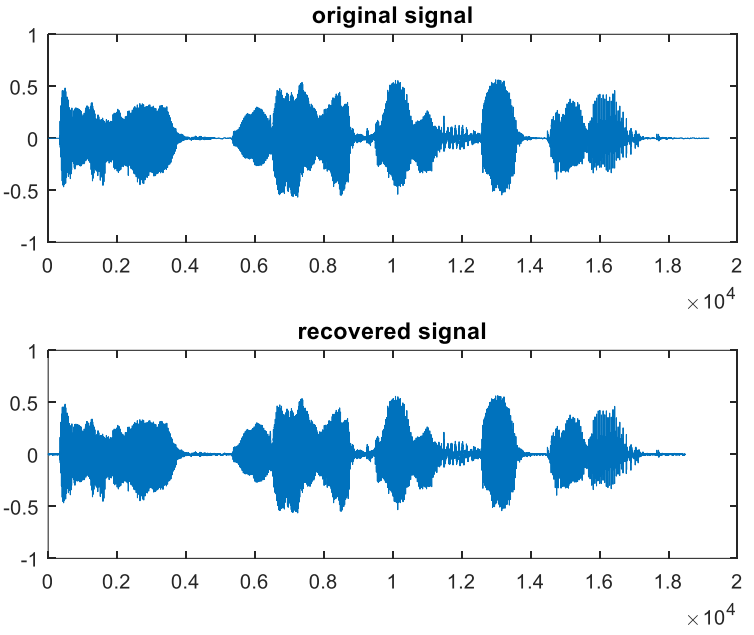

Fig.6. Time-domain plots of the original speech signal at the transmitter and the reconstructed speech signal at the receiver

Figure 5 shows the time-domain plots of the host speech signal and the steganographically embedded speech signal at the transmitter. It can be observed that the time-domain plots of the host and the embedded speech signals are identical. Auditory tests conducted with listeners of various ages indicated that the listeners could not distinguish the host speech signal from the steganographically embedded speech signal.

Figure 6 shows the time-domain plots of the original speech signal at the transmitter and the reconstructed speech signal at the receiver, demonstrating that these signals are identical. Auditory tests at this stage revealed that listeners could not distinguish the original speech from the reconstructed speech.

5. CONCLUSION

In this paper, to enhance the security of speech signal transmission, we proposed and implemented a hybrid system of cryptography and steganography. In the encryption phase, we utilized the discrete cosine transform (DCT) and added random numbers to the transform coefficients. Based on a sequence of random numbers derived from a logistic chaotic system, we scrambled the initial time-domain samples, the transform coefficients, and the time-domain samples from the inverse transform to construct the encrypted signal. The sequence of random numbers was periodically altered according to a specific key to increase the signal's security, complicating the discovery of the encryption method and making the reversal of steps difficult. In the steganography phase, we embedded the encrypted speech signal into the low-value coefficients of the host speech signal's discrete wavelet transform (DWT) and constructed the steganographic signal using the inverse wavelet transform. To evaluate the encryption and steganography methods, we employed both qualitative and quantitative criteria. The results indicated high levels of the applied criteria for the proposed hybrid method. The encryption complexity and the use of multiple secret keys between the sender and receiver constitute the innovation of this research. For future work, examining the effects of noise and various unauthorized receiver attacks on the proposed method is recommended.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Yin, Juan., Li, Yuhuai., Liao, Shengkai., Yang, Meng., Cao, Yuan., Zhang, Liang., Ren, Ji-Gang., Cai, Wenqi., Liu, Weiyue., Li, Shuang-Lin., Shu, R.., Huang, Yongmei., Deng, Lei., Li, Li., Zhang, Qiang., Liu, Nai-Le., Chen, Yu-Ao., Lu, Chaoyang., Wang, Xiang-Bin., Xu, Feihu., Wang, Jian-Yu., Peng, Cheng-Zhi., Ekert, A.., & Pan, Jian-Wei. (2020). Entanglement-based secure quantum cryptography over 1,120 kilometres. Nature , 582 , 501 - 505 . http://doi.org/10.1038/s41586-020-2401-y

[2] Yesina, M.V.., Ostrianska, Ye.V.., & Gorbenko, I.. (2022). Status report on the third round of the NIST post-quantum cryptography standardization process. Radiotekhnika . http://doi.org/10.30837/rt.2022.3.210.05

[3] Kumari, S.., Karuppiah, Marimuthu., Das, A.., Li, Xiong., Wu, Fan., & Kumar, Neeraj. (2017). A secure authentication scheme based on elliptic curve cryptography for IoT and cloud servers. The Journal of Supercomputing , 74 , 6428 - 6453 . http://doi.org/10.1007/s11227-017-2048-0

[4] Memon, I.., Hussain, Ibrar., Akhtar, Rizwan., & Chen, Gencai. (2015). Enhanced Privacy and Authentication: An Efficient and Secure Anonymous Communication for Location Based Service Using Asymmetric Cryptography Scheme. Wireless Personal Communications , 84 , 1487 - 1508 . http://doi.org/10.1007/s11277-015-2699-1

[5] Haeb-Umbach, R.., Watanabe, Shinji., Nakatani, T.., Bacchiani, M.., Hoffmeister, Björn., Seltzer, M.., Zen, H.., & Souden, M.. (2019). Speech Processing for Digital Home Assistants: Combining signal processing with deep-learning techniques. IEEE Signal Processing Magazine , 36 , 111-124 . http://doi.org/10.1109/MSP.2019.2918706

[6] Chowdhary, C. L.., Patel, Pushpam Virenbhai., Kathrotia, Krupal Jaysukhbhai., Attique, M.., P, Kumaresan., & Ijaz, M.. (2020). Analytical Study of Hybrid Techniques for Image Encryption and Decryption. Sensors (Basel, Switzerland) , 20 . http://doi.org/10.3390/s20185162

[7] Lee, Bong-Ki., & Chang, Joon-Hyuk. (2016). Packet Loss Concealment Based on Deep Neural Networks for Digital Speech Transmission. IEEE/ACM Transactions on Audio, Speech, and Language Processing , 24 , 378-387 . http://doi.org/10.1109/TASLP.2015.2509780

[8] P.Sathiyamurthi, Dr.., & Ramakrishnan, S.. (2017). Speech encryption using chaotic shift keying for secured speech communication. EURASIP Journal on Audio, Speech, and Music Processing , 2017 . http://doi.org/10.1186/s13636-017-0118-0

[9] Zhou, Xinyi., Gong, Wei., Fu, W.., & Jin, L.. (2016). An improved method for LSB based color image steganography combined with cryptography. 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS) , 1-4 . http://doi.org/10.1109/ICIS.2016.7550955

[10] Sadhukhan, Dipanwita., Ray, Sangram., Biswas, Gautam., Khan, M.., & Dasgupta, Mou. (2020). A lightweight

remote user authentication scheme for IoT communication using elliptic curve cryptography. The Journal of Supercomputing , 77 , 1114 - 1151 . http://doi.org/10.1007/s11227-020-03318-7

[11] Saadi, S., Merrad, A., & Benzian, A. (2019). Novel secured scheme for blind audio/speech norm-space watermarking by Arnold algorithm. Signal Processing, 154, 74-96. https://doi.org/10.1016/j.sigpro.2018.08.011

[12] Al-Najjar, A. J., Alvi, A., Idrees, S. U., & Al-Manea, A. (2007). Hiding encrypted speech using steganography. In Proceedings of the 7th WSEAS International Conference on Multimedia, Internet & Video Technologies (pp. 275-281). Beijing, China.

[13] Peyvandi, H., & Park, S. J. (2011). Security in data communication and privacy in conversations for underwater wireless networks using scrambled speech scheme. In IEEE OCEANS 2011 (pp. 19-22). Waikoloa, HI. https://doi.org/10.23919/OCEANS.2011.6106900

[14] Zghair, H. K., Mehdi, S. A., & Sadkhan, S. B. (2021). Speech scrambler based on discrete cosine transform and novel seven-dimension hyper chaotic system. Journal of Physics: Conference Series, 1804(1), 012048. https://doi.org/10.1088/1742-6596/1804/1/012048

[15] Habib, Z., Khan, J. S., Ahmad, J., Muazzam, A. K., & Khan, A. F. (2017). Secure speech communication algorithm via DCT and TD-ERCS chaotic map. In 4th International Conference on Electrical and Electronics Engineering (ICEEE). https://doi.org/10.1109/ICEEE2.2017.7935827

[16] Enayatifar, R., Abdullah, A. H., & Isnin, I. F. (2014). Chaos-based image encryption using a hybrid genetic algorithm and a DNA sequence. Optics and Lasers in Engineering, 56, 83-93. https://doi.org/10.1016/j.optlaseng.2013.12.003