

Article

# A Multilayered Audio Signal Encryption Approach for Secure Voice Communication

Hanaa A. Abdallah \*  and Souham Meshoul \* 

Department of Information Technology, College of Computer and Information Sciences,  
Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

\* Correspondence: haabdullah@pnu.edu.sa (H.A.A.); sbmeshoul@pnu.edu.sa (S.M.)

**Abstract:** In this paper, multilayer cryptosystems for encrypting audio communications are proposed. These cryptosystems combine audio signals with other active concealing signals, such as speech signals, by continuously fusing the audio signal with a speech signal without silent periods. The goal of these cryptosystems is to prevent unauthorized parties from listening to encrypted audio communications. Preprocessing is performed on both the speech signal and the audio signal before they are combined, as this is necessary to get the signals ready for fusion. Instead of encoding and decoding methods, the cryptosystems rely on the values of audio samples, which allows for saving time while increasing their resistance to hackers and environments with a noisy background. The main feature of the proposed approach is to consider three levels of encryption namely fusion, substitution, and permutation where various combinations are considered. The resulting cryptosystems are compared to the one-dimensional logistic map-based encryption techniques and other state-of-the-art methods. The performance of the suggested cryptosystems is evaluated by the use of the histogram, structural similarity index, signal-to-noise ratio (SNR), log-likelihood ratio, spectrum distortion, and correlation coefficient in simulated testing. A comparative analysis in relation to the encryption of logistic maps is given. This research demonstrates that increasing the level of encryption results in increased security. It is obvious that the proposed salting-based encryption method and the multilayer DCT/DST cryptosystem offer better levels of security as they attain the lowest SNR values,  $-25$  dB and  $-2.5$  dB, respectively. In terms of the used evaluation metrics, the proposed multilayer cryptosystem achieved the best results in discrete cosine transform and discrete sine transform, demonstrating a very promising performance.



**Citation:** Abdallah, H.A.; Meshoul, S.

A Multilayered Audio Signal Encryption Approach for Secure Voice Communication. *Electronics*

2023, 12, 2. <https://doi.org/10.3390/electronics12010002>

Academic Editors: Namgi Kim, Hyunsoo Yoon and Qiang Lai

Received: 27 October 2022

Revised: 15 December 2022

Accepted: 16 December 2022

Published: 20 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** audio; fusion; encryption; chaotic systems; random projection; salting; chaotic baker map

## 1. Introduction

When transmitting data across communication channels, security and privacy represent major challenges. Speech signals can be conveyed via numerous means, such as telephone networks and private or public radio networks. Over time, speech has become the most common method of human interaction among all modes of communication. Speech communication is simple, natural, flexible, and efficient, which explains its widespread use and practicality. There are two possible ways to encode a speech signal: analog and digital. In an analog representation, the speech signal is viewed as a waveform that specifies the signal's frequency and amplitude. In digital form, the speech signal is the numerical equivalent of the analog version, where the signal is encoded as a sequence of zeros and ones. Certain situations need the transmission of confidential information, including diplomatic and military communications during times of war and peace.

Because voice is so redundant when compared to written material, ensuring security is extremely difficult. Speech security can be achieved using either analog scrambling or digital ciphering; two distinct techniques. Speech scrambling has always piqued the interest of researchers due to its low-bandwidth consumption, ease of implementation, and effective handling of asynchronous transmission.

Speech codes are used in digital telecommunication networks to reduce the required transmission bandwidth. Other vocoders such as LPC-10 (Linear Prediction Coding), CELP (Code Excited Linear Prediction), MELP (Mixed Excitation Linear Prediction), and others have been developed. The most secure communication systems are built on LPC techniques. The main reason for this is that LPC voice coding ensures low bit rates and excellent voice clarity. Cryptographic algorithms frequently have stringent implementation criteria, such as minimal resource use, reduced memory, and logic gate count, and efficient power use. Designing systems that meet all of these needs in this context is a difficult task that requires extensive research.

More specifically, an implementation must be fast enough to avoid a significant slowdown in the system caused by the operation of cryptographic methods. Because software implementations have shown to be unable to deliver the required level of performance at an affordable price, hardware acceleration is adopted to accomplish this [1,2].

Only a small portion of available resources are dedicated to cryptography, making it difficult to develop high-security methods.

Real-time processing is an essential requirement for secure voice communications. In this case, the task of framing the incoming data becomes vital. It is crucial to strike a balance between the block size and all other elements since depending on the sampling rate, either short or large buffers may result in buffer overflows and delays. Because of their complexity, cryptographic algorithms need to be implemented on flexible platforms to meet real-time voice encryption requirements.

To improve security and address issues with existing techniques, we suggest investigating the use of multilayer encryption (fusion, substitution, and permutation) to achieve greater privacy by increasing confusion, which complicates the relationship between original and encrypted signals. The main contributions of this work can be summarized as follows:

- The encryption of audio signals using random projection and salting in the time domain and frequency domain.
- The encryption of audio signals using three layers of encryption that include fusion, substitution, and permutation in the frequency domain.

The following outlines how this paper is organized. Section 2 introduces the related work on audio signal encryption. Section 3 presents the proposed audio cryptosystems. The experimental study, simulation results, and discussion are included in Section 4. Finally, Section 5 summarizes the work and provides the concluding remarks.

## 2. Related Work

As mentioned previously, several attempts for audio data encryption have been proposed. For instance, numerous studies have examined audio encryption algorithms with the ultimate goal to achieve a good balance of speed and security [3]. Sharma et al. [4] developed a method for encrypting audio recordings with the RSA algorithm. Several algorithms, such as those outlined in [5], encrypted text files and images using a variety of shuffling techniques. The RSA algorithm was used to encrypt speech files with each word extracted and translated to text, as described in [6]. Using a chaotic map, several encryption keys are generated and used during each iteration of the encryption process, as in the chaotic audio encryption approach described in [7]. Five steps were employed to implement the audio encryption technique; these phases use various diffusion and table substitution algorithms to encrypt audio data without data loss [8]. To reduce the time required for audio encryption, the authors of [9] recommended employing the discrete Fourier transform to encrypt selected parts of the audio stream.

In [8], the dual-channel audio data is encrypted with a one-time key using a chaotic system with both the confusion and diffusion technique. To guard against brute-force attacks, the approach uses a big key space. Block by block, the cosine number transformation has already been used to generate a secret key from non-compressed 16-bit audio data [10]. The sequence is based on a narrative that combines Henon and economic maps. When

computing encrypted audio data, the confusion and diffusion technique is repeatedly used to encrypt plain audio data [11]. Audio data has also been confused and disseminated using DNA coding and chaotic systems. The chaotic system's initial value is calculated using the hash value of the audio [12], the audio signal is transformed into data using a lifting wavelet methodology in a new encryption method, and it is then encrypted using a chaotic dataset and a hyperbolic function. Block-by-block encryption for audio files uses the principles of a block cipher and chaotic maps. In the permutation stage, a chaotic tent map is utilized. After that, a key block and the obtained block were XORed. The multiplication inverse-based method of substitution is used to replace the resultant block [13]. With self-adaptive scrambling, chaotic maps, DNA coding, and cipher feedback mechanisms, a novel audio transmission technique is discussed. A pseudo-random number is produced by combining five separate chaotic maps and eight control settings [14]. The chaotic circle map and modified rotation equations are utilized to produce the pseudo-random number in a proposed encryption technique for audio data [15]. The permutation of audio samples using a discrete modified Henon map, followed by a substitution operation, is used to create a proposed audio encryption method. The modified Lorenz–hyperchaotic system yields the keystream. To assess the quality of the encryption method, many quality criteria have been established [9]. The use of numerous chaotic maps and cryptographic protocols is described as a novel approach to voice signal encryption. The input signal is separated into four parts using a cubic map as part of the scrambling procedure. The blowfish algorithm is used in combination with the private key to protect all of the chaotic map parameters. The blowfish key of the system and the hashing algorithm are implemented between the sender and receiver endpoints. The message digest is used to authenticate and verify the chaotic map parameters during secure communication. To demonstrate the effectiveness of the technique, several statistical tests are conducted [16]. A chaos-based cryptosystem was used to develop a novel multiuser speech encryption technique. Chua chaotic systems are used in transmitters and receivers to generate chaotic encryption and decryption keys. The XOR operation is combined with a chaotic matrix operation for randomization to encrypt the speech stream. The security analysis demonstrates the vulnerability of secret keys and the need for a large key space to withstand a brute-force attack. Strong diffusion and confusion processes have increased the battery life of the transmitter [17].

A novel audio encryption system has been investigated using a substitution–permutation algorithm and DNA encoding [18]. As part of the key generation process, the chaotic logistic map generates a new key block for each plain block. Several security assaults are carried out to evaluate the system. A cycle attack, selected plaintext attacks, and chosen cipher text are all successfully demonstrated [18]. Effective cryptography is required for secure communication to exhibit confidentiality, privacy, efficiency, and accuracy. The security considerations and methods incorporated into the design and implementation of the most popular symmetric encryption algorithms were examined. Researchers estimated and compared the performance of various encryption algorithm parameters, including encryption and decryption times, throughput, key size, the avalanche effect, memory requirements, correlation assessment, and entropy, to determine which encryption algorithm is best for each application. With the use of a substitution–permutation algorithm and DNA encoding, a novel audio encryption system has been investigated. With the chaotic logistic map, a new key block is created for each plain block as part of the key generation process. To assess the system, several security assaults are carried out. A cycle attack, the selected plaintext attacks, and the selected cipher text are all successfully shown [18]. Effective cryptography is necessary for secure communication to have the properties of confidentiality, privacy, efficiency, and correctness. The security considerations and procedures addressed in the design and implementation of the most widely used symmetric encryption algorithms were examined. To determine which encryption algorithm is best for each application, researchers estimated and compared the performance of various encryption algorithm parameters, including encryption and decryption times, throughput, key size, the avalanche effect, memory requirements, correlation assessment, and entropy.

The study described in [19] proposes a new chaotic one-dimensional map. The second hyperbolic tangent term in this connection is delayed to prevent dynamical degradation. The map's consistent chaotic behavior for practically all parameter values is demonstrated by computing its bifurcation diagrams and Lyapunov exponent diagrams. The suggested map is then used to construct a high-key space pseudo-random bit generator. Then, using this generator as a foundation, a proposed password generator application is built. The goal of this program is to develop an algorithm that takes a user-supplied, easy-to-remember key as the input and generates a strong password suitable for website security or file security.

The authors of [20] propose a brand-new, straightforward chaotic one-dimensional map. The chaotic properties have been declared using Lyapunov exponent analysis and bifurcation analysis. They also propose a new picture encryption scheme based on this novel chaotic map. The shuffling algorithm and the substitution method both use this map. According to numerous statistical tests and security analyses, this approach has outstanding security performance and can compete with certain other recently presented picture encryption algorithms.

Authors of [21] investigate the problem of chaos-based image encryption. The first step is to create and investigate a generalization of the one-dimensional chaotic map proposed by Talhaoui et al. in *The Visual Computer*. The generalized map, like the original image, depicts areas of persistent chaos. Using the new map, a statistically safe pseudo-random bit generator is created and used in the encryption procedure. To introduce an image encryption technique based on rearranging the bit levels of an image, the bits are first organized into a three-dimensional matrix, and then a three-level shuffling is applied to each row, column, and bit level of the 3D matrix. After the bits have been shuffled, they are subjected to an exclusive OR operation.

In light of the provided literature review, we were motivated to develop a multilayer encryption system to increase privacy.

To the best of our knowledge, no prior research has investigated the impact of using layers of encryption. The main contributions of this study are described in the following section.

### 3. Proposed Approaches to Multilayer Audio Data Encryption

In the following proposed cryptosystems, audio signals and personal data are encrypted before being transmitted over wireless networks. The encrypted information can either be used as-is for authentication in cancelable biometric applications at the receiving end, or it can be recovered using a secret key and a decryption algorithm to be used for additional audio analysis. In this research, we consider the discrete wavelet transform (DWT), the discrete cosine transform (DCT), the discrete sine transform (DST), random projection, salting, and the chaotic baker map. As depicted in Figure 1, these methods are arranged into three-layer encryption systems where numerous combinations are investigated. The cryptosystems alleviate the issue of audio low-activity intervals by combining the audio signal with a speech signal that lacks silence periods. The audio signal and the speech signal undergo preprocessing to meet the fusion requirements. During the preprocessing phase, the audio signal is converted from a one-dimensional to a two-dimensional signal. The 2D DWT is then applied to the audio signal. The same approach is used for the speech signal. In the proposed cryptosystems, the first layer is a fusion layer that employs either random projection or salting. The second layer is a substitution layer based primarily on the DCT or DST techniques. The third layer is a permutation layer that uses the chaotic baker map. Before delving into these strategies, we outline the overall structure first. The idea behind proposing a three-layer system is to strengthen audio signal security and preserve privacy and confidentiality. As seen in Figure 1, audio and speech input signals are transformed into 2D signals prior to DWT processing. The two signals are subsequently fused using either random projection or salting. The result of applying an inverse DWT (IDWT) on the fused signal is sent to the next layer for the second level of encryption using a substitution operation with DCT or DST. The result of the substitution is then sent to the permutation

layer using a chaotic baker map for the third level of encryption. In order to obtain the encrypted signal, an inverse DCT (IDCT) or inverse DST (IDST) is performed at the end. Algorithm 1 outlines this process. In Figures 2 and 3, the various combinations of fusion with random signal and substitution layer approaches are shown. Figure 2 depicts the use of DWT with random projection or salting, whereas Figure 3 depicts the use of DCT with random projection or salting. Further details concerning the proposed strategies follow.

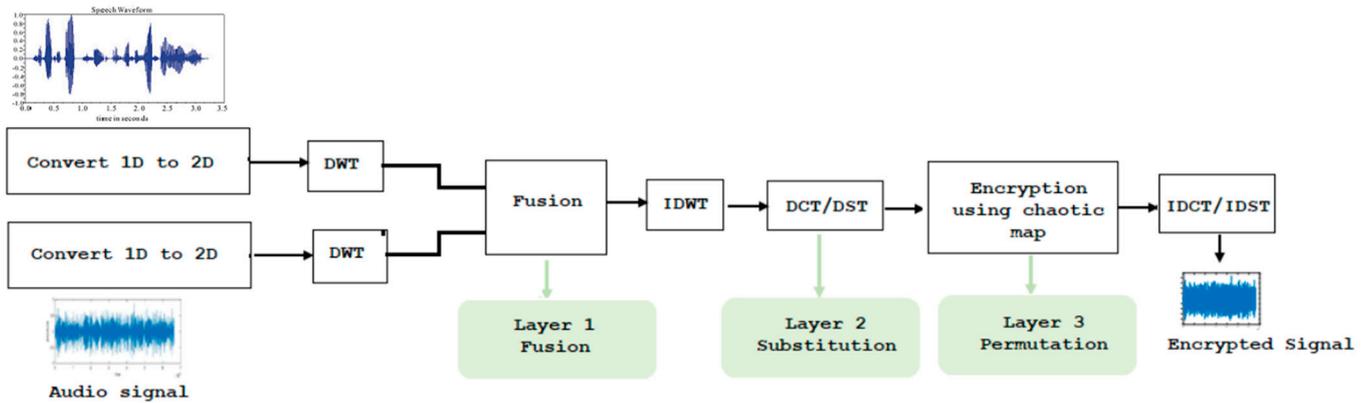


Figure 1. Cryptosystem based on three layers of encryption.

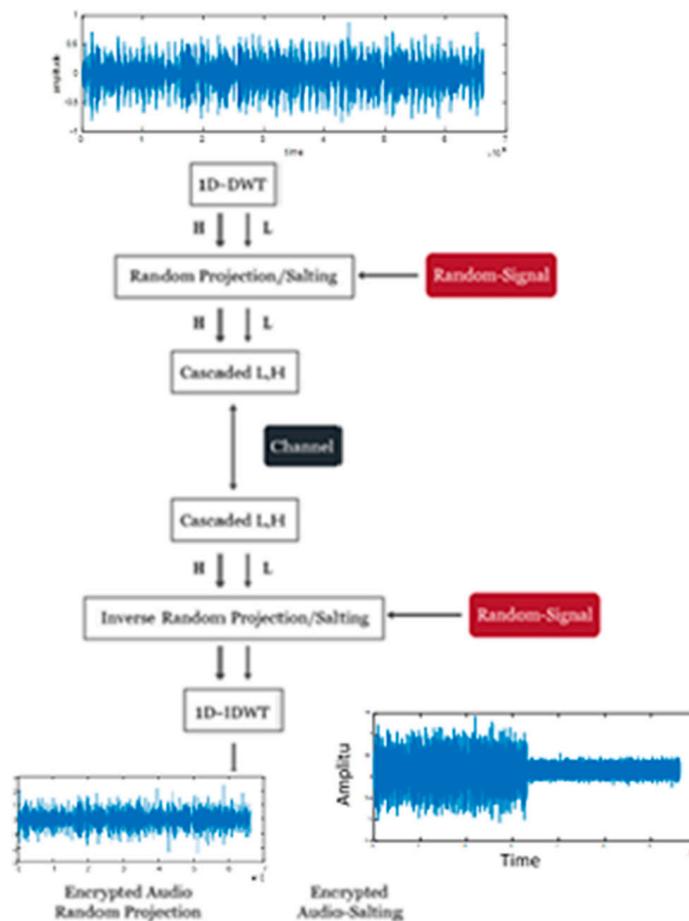
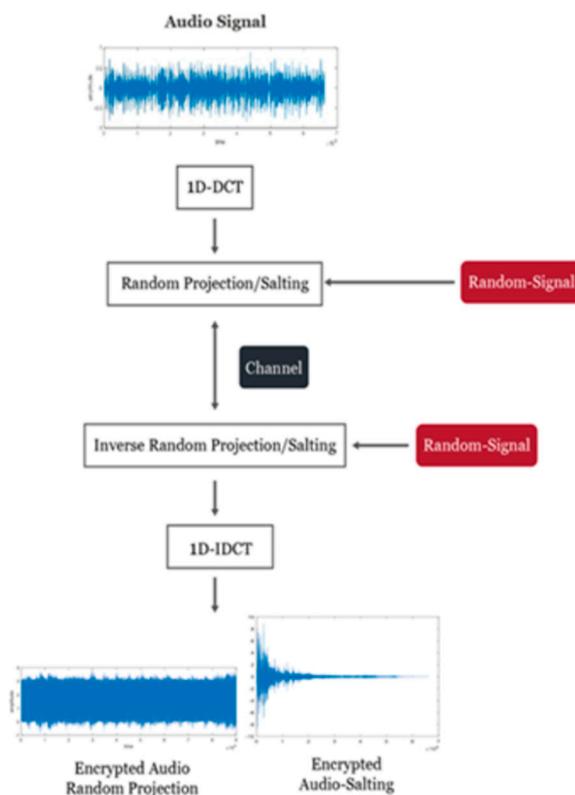


Figure 2. A 1D DWT fusion system for encryption with random projection and salting.

**Algorithm 1:** Encryption**Input:** Original audio signal and speech signal**Output:** Encrypted audio signal**Begin:**

- 1-Get audio signal.
- 2-Get speech signal.
- 3-Preprocess the two signals by converting them to 2D signals.
- 4-Apply 2D DWT to the 2D signals to decompose each signal into four subbands.
- 5-Apply fusion using random projection by multiplying the coefficients of the subbands of the two signals or applying salting by adding the coefficients.
- 6-Apply IDWT to the results of random projection or salting.
- 7-The results of step 6 undergo a DCT or DST to replace the values with the frequency coefficients.
- 8-The coefficients from step 7 are permuted using the chaotic baker map with key  $S\_key = [n1, n2, n3] = [2, 4, 2]$ .
- 9-Apply the IDCT or IDST to the output of step 8 results.
- 10-Convert the 2D signal obtained in step 9 to a 1D signal to obtain the encrypted audio.

**End****Figure 3.** Diagram of the proposed 1D-DCT encryption of audio signals.**3.1. Random Projection Based on 1D-DWT/1D-DCT**

The proposed algorithm for audio signal encryption based on random projection in the DWT domain is illustrated in Figure 2. The block diagram suggests combining the 1D DWT of the audio signal with a random signal.

The DWT first divides the signal into an approximation signal obtained by using a low-pass filter ( $L$ ) and details obtained by using a high-pass filter ( $H$ ). The details represent the high-frequency component, while the approximation is the low-frequency component. The audio signal is split into two sub-signals, each half its original length, by the DWT. One sub-signal is a running average or trend ( $L$ ), and the other is a running difference (or fluctuation) ( $H$ ). To generate a random signal, the audio signal is divided in half.

The random projection transformation is then used to encrypt the wavelet coefficients. Random projection transformation is frequently used for cancelable biometric templates [22]. “Random projection” refers to the act of projecting the original signal or feature vector onto a random space. It can be implemented by multiplying with a random matrix. For example, a random matrix  $F$  can be used to multiply the vector of an audio signal,  $L$ , to produce a random vector,  $L^*$ . Also, a random matrix  $F$  can be used to multiply an audio signal’s vector,  $H$ , to produce a random vector,  $H^*$ , as can be seen in Equation (1) where  $k$  and  $d$  refer to the dimension of the matrix.

$$L^*_{k \times 1} = F_{k \times d} \cdot L_{d \times 1}, H^*_{k \times 1} = F_{k \times d} \cdot H_{d \times 1} \quad (1)$$

The wavelet coefficients of an audio signal are transformed into a random signal using this method. The encrypted audio signal is created after the random projection transformation. As a result, the relative separations between any two samples in the encrypted audio signal space are preserved in the output random space. Then, at the receiver, the audio signal is subjected to the inverse random projection transform in preparation for further processing. Finally, the inverse DWT method is used to reconstruct the original audio signal so that it can be compared to those in the database if used for authentication.

Another combination is considered in our study. As illustrated in Figure 3, the discrete cosine transform (DCT) can be used in place of the DWT. In this case, the random projection algorithm is used to encrypt the DCT coefficients.

### 3.2. Salting Based on 1D-DWT/1D-DCT

Figures 2 and 3 show the other combinations considered in our study, where the fusion rule is achieved this time through salting. Depending on the type of random sequence, the salting process involves adding a random sequence to the audio stream or its extracted features (Gaussian or uniform). The noise power affects the performance of the salting-based technology. This method entails blending a pattern with the original audio pattern’s DWT or DCT elements. The mixed patterns may contain pure random noise or artificial patterns. The relative strength of the noise patterns influences the cryptosystem’s effectiveness. As a result, picking a pattern with a lot of actions makes sense. Using a simple subtraction technique, the original audio signal is extracted at the receiver for further processing.

### 3.3. Permutation Using Chaotic Baker Map

The third method in the suggested cryptosystems for permuting the coefficients from layer 2 to obscure the output of the second layer is the chaotic baker map. The horizontal and vertical stretching and folding of the coefficients provide the foundation of the baker map. The positions of each coefficient in the output of the second layer are changed by repeating this process. More specifically, the chaotic baker map is a two-dimensional map that shifts each square matrix element [23]. These maps are used to nonlinearly randomize the elements of the square matrix in a framework that is sensitive to the initial conditions. To achieve confusion and diffusion, cryptosystems must be extremely sensitive to the starting conditions. The two key properties of the chaotic baker map are as follows:

1. Its output is highly random, unpredictable, and correlated little with the input.
2. It is highly dependent upon the input parameters, starting values, and early conditions.

Let  $B(n_1, \dots, n_k)$  denote the discretized map, where the vector  $[n_1, \dots, n_k]$  represents the secret key,  $S_{key}$ . Defining  $N$  as the number of data items in one row, the secret key is chosen such that each integer  $n_i$  divides  $N$ , and  $n_1 + \dots + n_k = N$ . Let  $N_i = n_1 + \dots + n_i$ . The data item at the indices  $(q, z)$ , is moved to the indices: as shown in Equation (2).

$$B_{(n_1, \dots, n_i)}(q, z) = \left( \frac{N}{n_i}(q - N_i) + z \bmod \left( \frac{N}{n_i} \right), \frac{n_i}{N} \left( Z - Z \bmod \left( \frac{N}{n_i} \right) \right) + N_i \right) \quad (2)$$

where  $N_i \leq q < N_i + n_i$ , and  $0 \leq z < N$ .

The chaotic permutation is performed in the following steps:

1. An  $N \times N$  square matrix is divided into  $k$  rectangles of width  $n_i$  and the number of elements  $N$ .
2. The elements in each rectangle are rearranged into a row in the permuted rectangle. Rectangles are taken from left to right beginning with upper rectangles and then lower ones.
3. Inside each rectangle, the scan begins from the bottom left corner toward the upper elements.

Figure 4 below shows a popular example that explains how permutation is performed for a chaotic map of an  $(8 \times 8)$  square image (i.e.,  $N = 8$ ). The secret key,  $S_{key} = [n_1, n_2, n_3] = [2, 4, 2]$ .

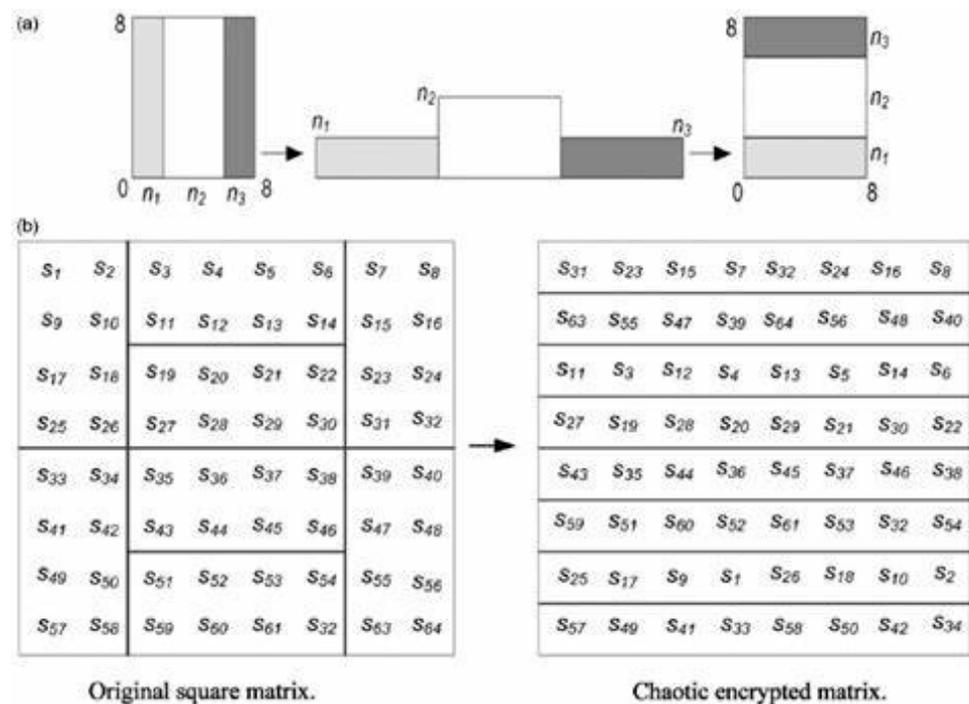


Figure 4. (a) The  $8 \times 8$  matrix divided into blocks, and (b) represents the matrix after applying the 2D baker map [24].

### 3.4. Decryption Process of the Proposed Three Layers Technique

The inverse of the encryption process is used in the decryption algorithm. As shown in Figure 5, the encrypted signal is transformed using DCT or DST and the resulting coefficients are passed through a pipeline that begins with chaotic decryption followed by an inverse discrete cosine transform or inverse discrete sine transform, then DWT. After this stage, inverse fusion is performed, whereas in salting, we subtract the speech coefficients from the audio coefficients (in case of a random projection we apply division), then an inverse DWT is applied, and finally, a conversion operation is performed to obtain the decrypted signal.

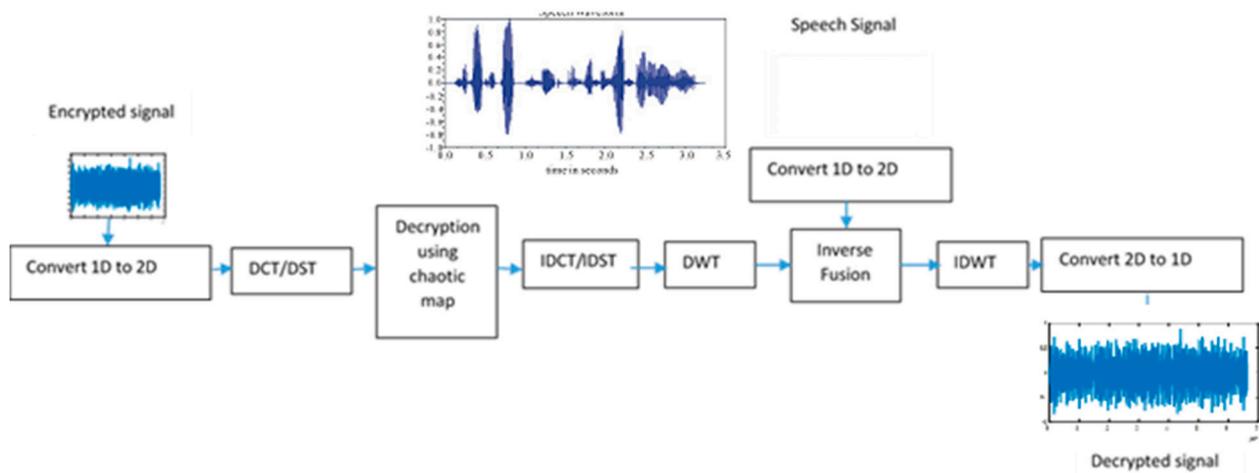


Figure 5. Decryption process of the proposed three layers technique.

#### 4. Simulation Results and Discussions

In order to evaluate the proposed cryptosystems, the performance of each technique was evaluated in a simulated environment using MATLAB. We also considered logistic map encryption, which is used in many comparable publications, for comparison purposes. The logistic map is one of the traditional chaotic maps that can be used to generate keys for an encryption scheme [24]. The 1D logistic map is defined as follows:

$$x_{m+1} = hx_m(1 - x_m) \tag{3}$$

where  $x_0$  is the initial state,  $m$  is the number of iterations and  $h$  is a system control parameter with values ranging from 0 to 4, and  $x_{m+1}$  has a value range of 0 to 1 for all values of  $m$ . The value of  $h$  determines how the logistic map behaves. As the repetitions become entirely chaotic,  $h$  is chosen for encryption purposes to be 3.57. There are different behaviors of  $x$  depending on the  $h$  value. The 1D logistic map system is used to construct the new S-Box used to permute samples.

Due to the 1D logistic map’s limitations, a modified 2D logistic map has been developed for encryption. It is defined as follows [24]:

$$x_{m+1} = h(3y_m + 1)x_m(1 - x_m) \tag{4}$$

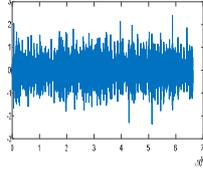
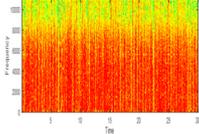
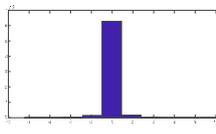
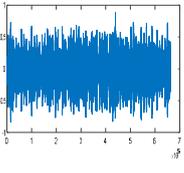
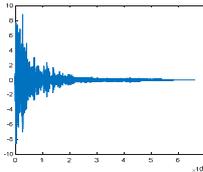
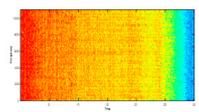
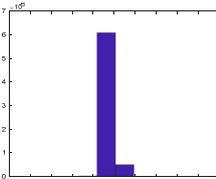
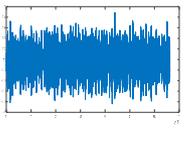
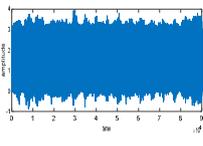
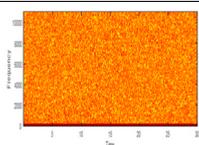
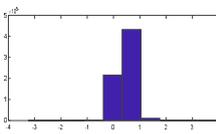
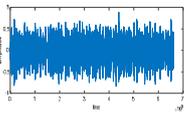
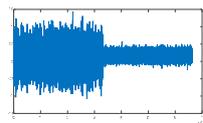
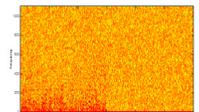
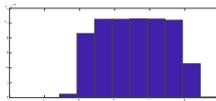
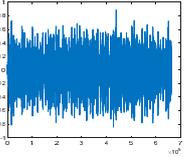
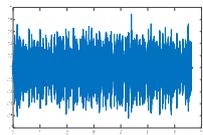
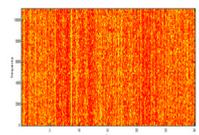
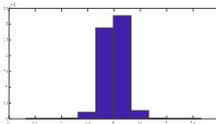
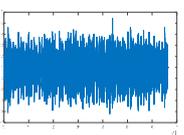
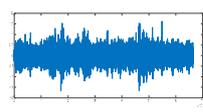
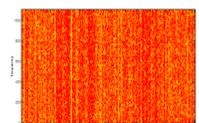
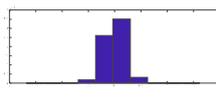
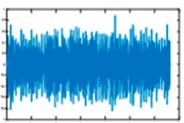
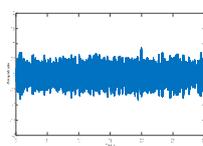
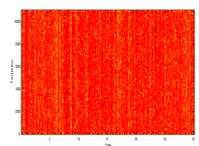
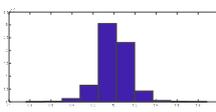
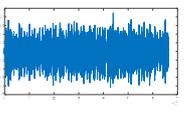
$$y_{m+1} = h(3x_{m+1} + 1)y_m(1 - y_m) \tag{5}$$

##### 4.1. Statistical Metrics

Each signal in the database has an average runtime of 3 min and a 360 Hz sampling frequency. Various measurements and viewpoints have been taken into account during the performance analysis process.

To evaluate the quality of the encryption and decryption processes, histograms of the encrypted and decrypted signals were calculated. Table 1 illustrates this comparison. It is important to note from Table 1 that both the proposed random projection based on 1D DWT and the cryptosystem based on three layers provide uniform histograms for the encrypted signals. These methods are hence resistant to statistical attacks. Unfortunately, the random projection based on DCT, salting, and logistic map encryption has histogram distributions that are not uniform enough, making them less resistant to histogram-based statistical assaults.

**Table 1.** Comparison between the results of different encryption schemes.

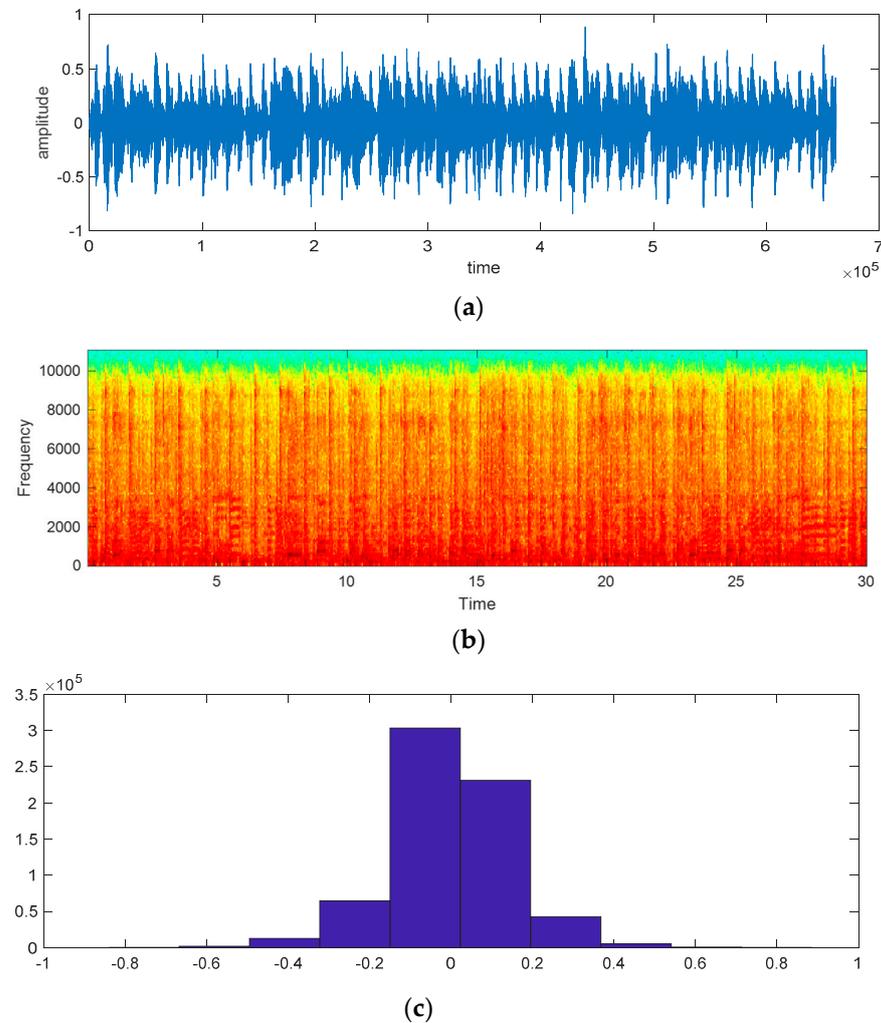
Encryption Method	Encrypted Signal	Spectrogram of Encrypted Signal	Histogram of Encrypted Signal	Decrypted Signal
Random projection based on 1D-DWT				
Random projection based on 1D-DCT				
Salting based on 1D-DCT				
Salting based on 1D-DWT				
Multilayer encryption in the DST domain				
Multilayer encryption in the DCT domain				
Encryption using a Logistic map				

SNR, SNRseg, LLR, and spectral distortion are examples of quality measures that can be used to assess the effectiveness of audio signal encryption (SD). Based on the original and encrypted audio signals, all of these parameters can be approximated. The quality of the encryption increases with decreasing SNR and SNRseg. Additionally, the encryption quality improves with greater LLR and SD values. We represent SNR in Equation (6), and we suggest referring to [25] for further details.

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N (x(n) - y(n))^2} \tag{6}$$

where  $x(n)$  is the plain audio signal,  $y(n)$  is the encrypted signal, and  $n$  is the time index.

Figure 6 shows the original audio, histogram, and spectrogram of the audio signal, whereas Figure 7 depicts a comparison of the proposed encryption techniques in terms of SNR. The suggested salting-based encryption technique and the three-phase DCT/DST cryptosystem provide the highest level of security compared to the other approaches since they achieve the lowest SNR values ( $-25$  dB and  $-2.5$  dB, respectively).



**Figure 6.** Original audio signal representations: (a) audio signal in the time domain, (b) audio spectrogram in the frequency domain, and (c) histogram for the audio signal.

Over a brief frame of the predefined signal, the average SNR values can be determined. SNRseg is a well-known metric that uses the following Equation (7) to compute the average SNR values for a given signal over a particular period of time.

$$\text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \sum_{i=Nm}^{Nm+N-1} \left( \frac{x^2(i)}{x(i) - y(i)} \right)^2 \right) \quad (7)$$

where  $M$  is the total number of frames in the audio signal, and  $N$  is the frame duration which can be set between 15 and 20 ms.

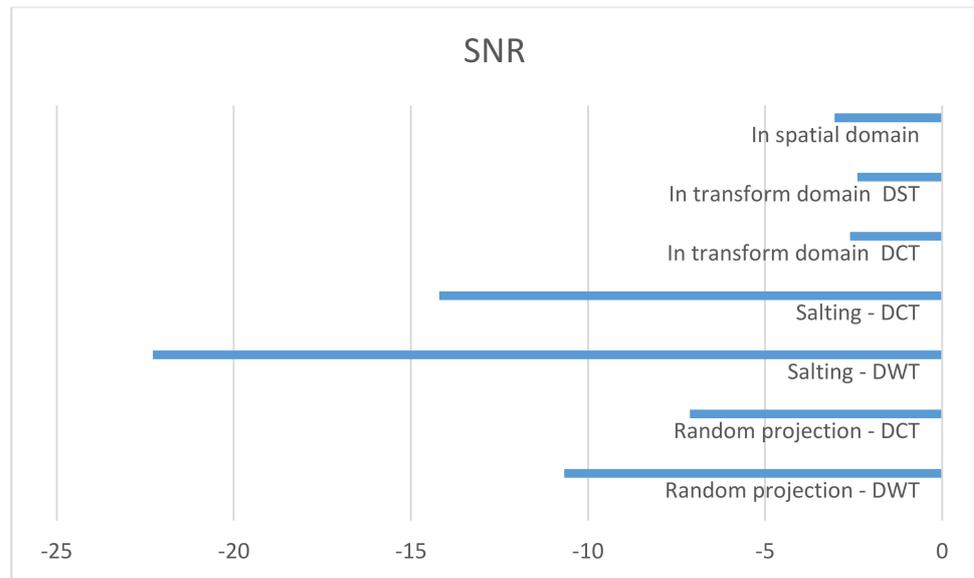


Figure 7. Comparison of SNR (dB) between the proposed cryptosystems.

The log-likelihood ratio (LLR), an objective metric described in [26], is based on the distance between two vectors of linear prediction coefficients (LPC) computed on plain and encrypted signals. It can be calculated using the following mathematical equation:

$$LLR = \left| \log \left[ \frac{\vec{a}_x \vec{r}_x \vec{a}_x^T}{\vec{a}_y \vec{r}_y \vec{a}_y^T} \right] \right| \tag{8}$$

LPC coefficients for the original audio signal are  $(a_x)$ , and those for the encrypted audio signal are  $(a_y)$ . The autocorrelation matrix for the original audio signal and the encrypted audio signal are determined by the values of  $(r_x)$  and  $(r_y)$ , respectively.

The LLR values of the suggested alternatives in comparison to the logistic map encryption method are shown in Figure 8. The suggested three-phase cryptosystem offers the maximum level of encryption since, when compared to the other systems, it obtains the highest LLR value.

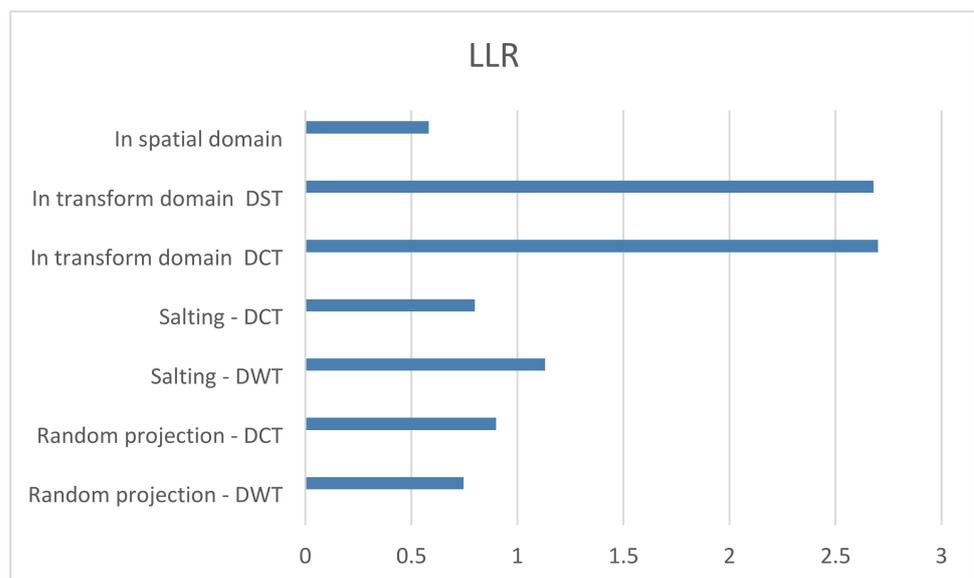


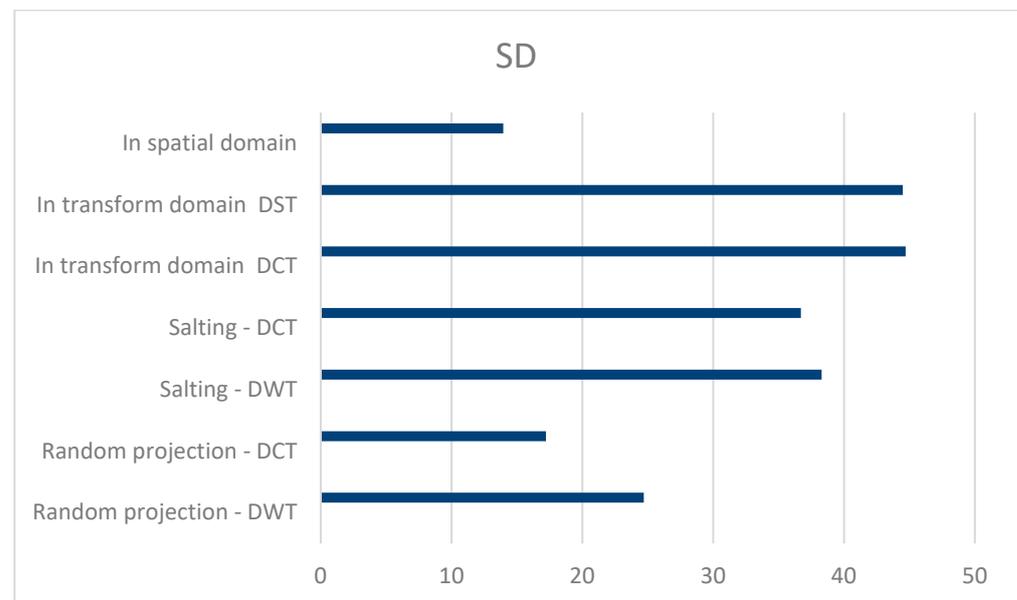
Figure 8. Comparison of LLR values for the proposed cryptosystem.

The spectral distortion (SD), which has maximum values of 44 to 45, indicates how much the encrypted signal spectrum differs from the original signal spectrum. In the frequency domain, it is calculated as:

$$SD = \frac{1}{M} \sum_{m=0}^{M-1} \sum_{i=N_m}^{Nm+N-1} |V_s(i) - V_y(i)| \quad (9)$$

where  $V_s(i)$  and  $V_y(i)$  represent the spectrum of the original and encrypted audio signals, respectively; greater simple signals result in a smaller SD value. Figure 9 compares several encryption methods in terms of SD values. It is apparent that the SD value of the 1D-DWT random projection, which achieves the most spectral distortion during the encryption process, yields the best results. Accordingly, the correlation between the original and encrypted audio signals is evaluated in order to determine how well the cryptosystem works:

$$r_{xy} = \frac{c_v(x,y)}{\sqrt{D(x).D(y)}} \quad (10)$$



**Figure 9.** Comparison of SD values for the proposed cryptosystem.

The correlation between the decrypted signal  $y$  and the original signal  $x$  is denoted by  $r_{xy}$  while  $c_v(x,y)$  refers to the covariance between them. The variances of these signals are represented by  $D(x)$  and  $D(y)$ , respectively. Further details on this measure can be found in [26]. The lower the value of the correlation coefficient  $r$ , the better the performance of the encryption.

Table 2 compares the normal and encrypted audio signals in terms of SNR,  $r$ , SSIM, SD, SNRseg, and LLR. Similar comparisons are conducted between plain and decrypted signals, as shown in Table 3, in order to determine how effectively the encryption systems perform after being decrypted.

**Table 2.** Comparison of various encryption schemes using the evaluation metrics.

Encryption							
Encryption Method	Algorithm	SNR	r	LLR	SD	SNRseg	SSIM
Fusion	Random projection-DWT	−10.35	0.01	0.746	24.7	−10.67	0.001
	Random projection-DCT	−6.34	0.01	0.90	17.21	−7.12	0.9
	Salting-DWT	−22.2804	0.06	1.13	38.29	−22.29	0.9
	Salting-DCT	−1.6	0.01	0.8	36.7	−14.2	0.9
Multilayer cryptosystem	In transform domain-DCT	−2.69	0.002	2.7	44.7	−2.6	0.2
	In transform domain-DST	−2.4	0.018	2.68	44.49	−2.39	0.2
Logistic map	In spatial domain	−3.0131	$−6.3977 \times 10^{-4}$	0.5818	13.9475	−3.0379	0.7

**Table 3.** Comparison of different decrypted signals.

Decryption							
Encryption Method	Algorithm	SNR	r	LLR	SD	SNRseg	SSIM
Fusion	Random projection-DWT	222.8	1	0	0	222.7	1
	Random projection-DCT	302.7	1	$8.3 \times 10^{-15}$	$1.5772 \times 10^{-14}$	302.7	1
	Salting-DWT	303.8	1	$1.2 \times 10^{-14}$	$1.4 \times 10^{-14}$	303.7	1
	Salting-DCT	300.4	1	$1.3897 \times 10^{-14}$	$2.0368 \times 10^{-14}$	300.4	1
Cryptosystem based on multilayers	In transform domain-DCT	304	1	0	0	304	1
	In transform domain-DST	302	1	0	0	302	1
Logistic map	In spatial domain	27.9444	0.9992	0.0610	0.8008	27.9120	0.999

The three suggested options offer high levels of security, as shown in Table 2, where the correlation values are near zero. However, the logistic chaotic encryption yields a correlation of just 0.48. The correlation values between the original and decrypted audio signals equal 1, as given in Table 3, further demonstrating the high quality of the decrypted audio signals. These tables reveal a significant discrepancy between the original and encrypted signals, revealing remarkably low SNR values. The findings demonstrate close to zero values for r, low values for SNR, and SSIM, big values for LLR, and high values for SD. These results confirm that the suggested audio encryption methods are extremely effective.

#### 4.2. Infiltration of Encrypted Signals

Noise attack is a considerable risk when the encrypted signals are transmitted through a communication channel. This section analyses how noise attacks impact the suggested approaches.

Here, the encrypted signal is subjected to salt and pepper noise with a variance equal to 0.019. The impact of noise on the decrypted signal is seen in Table 4. The noise disrupted the decryption process and the recovered signals suffered from data loss as a consequence. Table 5 displays the outcome of the cropping experiment using 40% of the encrypted signal. Contrary to salt and pepper noise, cropping was considerably less impactful. According to Tables 4 and 5, the random projection based on 1D DWT and the three-phase encryption cryptosystem record the best results, demonstrating that these two systems are highly resistant to noise attacks.

**Table 4.** Comparison of different decrypted signals after applying salt and pepper noise.

Decryption after Adding Salt and Pepper Noise 0.019							
Encryption Method	Algorithm	SNR	r	LLR	SD	SNRseg	SSIM
Fusion	Random projection-DWT	−50.69	0.015	0.87	28	−80.97	0.98
	Salting-DCT	−17.92	0.03	0.286	34.6	−44.6	0.99
	Random projection-DCT	−25.21	0.021	0.91	42	−25.2	0.93
	Salting-DWT	−17.39	0.043	1.2	34	−17.4	0.98
Multilayers cryptosystem	In transform domain-DCT	−1.35	0.42	0.35	13.2	0.49	0.99
	In transform domain-DST	−1.36	0.42	0.4	13.24	−1.86	0.98
Logistic map	In spatial domain	−17	0.96	0.33	39.8	26.7	0.99

**Table 5.** Comparison of different decrypted signals after applying cropping.

Decryption after Cropping by 40%							
Encryption Method	Algorithm	SNR	r	LLR	SD	SNRseg	SSIM
Fusion	Random projection-DWT	26.21	0.58	0.30	15.23	26.19	1
	Random projection-DCT	63.14	0.99	0.004	0.38	63.15	1
	Salting-DWT	12.39	0.15	0.66	28.54	12.39	0.99
	Salting-DCT	12.38	0.149	0.236	32.97	4.41	0.98
Multilayers cryptosystem	In transform domain-DCT	7.54	0.68	0.093	14.37	7.54	0.99
	In transform domain-DST	6.70	0.68	0.092	14.56	6.7	1
Logistic map	In spatial domain	7.656	0.69	0.119	15.02	7.65	0.99

#### 4.3. Performance Comparison of Audio Encryption Algorithms

In this section, we provide a comparison of audio encryption techniques from the literature in terms of correlation (r) and SNR. Table 6 displays the results of this comparative study.

**Table 6.** Comparison based on statistical analysis.

Method	Correlation Coefficient (r)	SNR
Ref. [25]	0.0011	−10.6357
Ref. [27]	0.0471	−1.47
Ref. [28]	0.0233	−33.7464
Ref. [29]	0.0029	−23.89
Ref. [30]	0.0491	−44.8
Ref. [31]	0.0321	−54.89
AES	0.00971	−1.4461
Triple DES	0.1704	−0.250
Fusion Random projection -DWT	0.01	−10.35
Fusion Random projection-DCT	0.01	−6.34
Fusion Salting-DWT	0.06	−22.2804
Fusion Salting-DCT	0.01	−1.6
Proposed multilayers in transform domain-DCT	0.002	−2.69
Proposed multilayers in transform domain-DST	0.018	−2.4
Logistic map in the spatial domain	$−6.3977 \times 10^{-4}$	−3.0131

As shown in Table 6, the proposed encryption methods achieved competitive results compared to the state-of-the-art methods showing low scores for both correlation coefficient  $r$  and SNR. The best score in terms of correlation score is obtained with the logistic map. However, its SNR score is not the best. The results clearly show that the proposed methods achieved a better balance in terms of  $r$  and SNR. Because the correlation scores are so low, we can conclude that a statistical attack by an intruder will not reveal any valuable data. On the other hand, the algorithm's SNR results are low, implying that the proposed algorithms have a lot of noise, making them more attack-resistant.

#### 4.4. Signal-to-Noise Ratio between Plain and Encrypted Audio Files

Since the employed measures show low SNRs, it is clear from Table 7 that all test findings show that the audio encryption technique is secure. According to Table 7, all of the observed SNR values are negative, indicating that the encrypted files are extremely noisy and that the encryption approach has entirely destroyed the clear signal in the plain audio files.

**Table 7.** Evaluation using four audio samples.

	SNR					
	Fusion Random Projection-DWT	Fusion Random Projection-DCT	Fusion Salting-DWT	Fusion Salting-DCT	Multilayer Cryptosystems in Transform Domain-DCT	Multilayer Cryptosystems in Transform Domain-DST
Audio 1	−10.35	−6.34	−22.28	−1.6	−2.69	−2.4
Audio 2	−12.43	−8.65	−23.52	−2.67	−5.98	−4.5
Audio 3	−9.81	−7.222	−21.31	−1.98	−3.43	−3.86
Audio 4	−10.98	−6.97	−22.56	−1.93	−2.72	−2.22

## 5. Conclusions

In this study, the generation of audio signals was framed as a security risk and three cryptosystems were developed and evaluated accordingly. Each cryptosystem approaches the problem differently. The first method involves the projection of random numbers onto the DWT coefficients of audio signals. The second method employs salting algorithms, whereas the third way enhances security by applying multiple levels of processing for encryption and functioning in three stages: fusion, substitution, and chaotic permutations. This method combines the original audio with a speech signal to compensate for lengthy audio signals with low-activity intervals. It then employs the 2D chaotic map to produce permutations. The proposed approaches were assessed in a simulated environment. The results demonstrated that the three-phase cryptosystem provides the best performance and attack resilience. In future research, we intend to use deep learning to extract the most prominent features from audio signals and to expand its use in further applications such as authentication.

**Author Contributions:** Conceptualization, H.A.A. and S.M.; methodology, H.A.A. and S.M.; software, H.A.A. and S.M.; validation, H.A.A. and S.M.; formal analysis, H.A.A. and S.M.; investigation, H.A.A. and S.M.; resources, S.M. and H.A.A.; data curation, H.A.A. and S.M.; writing—original draft preparation, H.A.A. and S.M.; writing—review and editing, H.A.A. and S.M.; visualization, H.A.A. and S.M.; supervision, H.A.A. and S.M.; project administration, H.A.A. and S.M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project is supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R196), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Acknowledgments:** The authors would like to acknowledge the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R196), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pedre, S.; Krajník, T.; Todorovich, E.; Borensztein, P. Accelerating embedded image processing for real time: A case study. *J. Real-Time Image Process.* **2016**, *11*, 349–374. [[CrossRef](#)]
2. Joao, J.A.; Mutlu, O.; Patt, Y.N. Flexible reference-counting-based hardware acceleration for garbage collection. *ACM SIGARCH Comput. Arch. News* **2009**, *37*, 418–428. [[CrossRef](#)]
3. Albahrani, E.A.; Alshekly, T.K.; Lafta, S.H. A Review on Audio Encryption Algorithms Using Chaos Maps-Based Techniques. *J. Cyber Secur. Mobil.* **2021**, *11*, 53–82. [[CrossRef](#)]
4. Sheetal, S.; Kumar, L.; Sharma, H. Encryption of an Audio File on Lower Frequency Band for Secure Communication. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2013**, *3*, 79–84.
5. Yahya, A.; Abdalla, A. An AES-Based Encryption Algorithm with Shuffling. In Proceedings of the 2009 International Conference on Security & Management, SAM 2009, Las Vegas, NV, USA, 13–16 July 2009; pp. 113–116.
6. Yousif, S.F. Encryption and Decryption of Audio Signal Based on Rsa Algorithm. *Int. J. Eng. Technol. Manag. Res.* **2020**, *5*, 57–64. [[CrossRef](#)]
7. Chaudhary, N.; Shahi, T.B.; Neupane, A. Secure Image Encryption Using Chaotic, Hybrid Chaotic and Block Cipher Approach. *J. Imaging* **2022**, *8*, 167. [[CrossRef](#)] [[PubMed](#)]
8. Liu, H.; Kadir, A.; Li, Y. Audio encryption scheme by confusion and diffusion based on multi-scroll chaotic system and one-time keys. *Optik* **2016**, *127*, 7431–7438. [[CrossRef](#)]
9. Farsana, F.; Devi, V.; Gopakumar, K. An audio encryption scheme based on Fast Walsh Hadamard Transform and mixed chaotic keystreams. *Appl. Comput. Inform.* **2019**. *Online ahead of print.* [[CrossRef](#)]
10. Oyiza, A.H.; Maarof, M.A. An Improved Discrete Cosine Transformation Block Based Scheme for Copy-move Image Forgery Detection. *Int. J. Innov. Comput.* **2019**, *9*, 2. [[CrossRef](#)]
11. Adhikari, S.; Karforma, S. A novel audio encryption method using Henon–Tent chaotic pseudo random number sequence. *Int. J. Inf. Technol.* **2021**, *13*, 1463–1471. [[CrossRef](#)]
12. Wang, X.; Su, Y. An Audio Encryption Algorithm Based on DNA Coding and Chaotic System. *IEEE Access* **2019**, *8*, 9260–9270. [[CrossRef](#)]
13. Albahrani, E.A. A new audio encryption algorithm based on chaotic block cipher. In Proceedings of the 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT), Baghdad, Iraq, 7–9 March 2017; pp. 22–27. [[CrossRef](#)]
14. Abdelfatah, R.I. Audio Encryption Scheme Using Self-Adaptive Bit Scrambling and Two Multi Chaotic-Based Dynamic DNA Computations. *IEEE Access* **2020**, *8*, 69894–69907. [[CrossRef](#)]
15. Kordov, K.; Bonchev, L. Using Circle Map for Audio Encryption Algorithm. *Math. Softw. Eng.* **2017**, *3*, 183–189.
16. Yasser, I.; Mohamed, M.; Samra, A.; Khalifa, F. A Chaotic-Based Encryption/Decryption Framework for Secure Multimedia Communications. *Entropy* **2020**, *22*, 1253. [[CrossRef](#)] [[PubMed](#)]
17. Hashemi, S.; Pourmina, M.A.; Mobayen, S.; Alagheband, M.R. Multiuser wireless speech encryption using synchronized chaotic systems. *Int. J. Speech Technol.* **2021**, *24*, 651–663. [[CrossRef](#)]
18. Belazi, A.; El-Latif, A.A.A.; Belghith, S. A novel image encryption scheme based on substitution-permutation network and chaos. *Signal Process.* **2016**, *128*, 155–170. [[CrossRef](#)]
19. Kafetzis, I.; Moysis, L.; Tutueva, A.; Butusov, D.; Nistazakis, H.; Volos, C. A 1D coupled hyperbolic tangent chaotic map with delay and its application to password generation. *Multimed. Tools Appl.* **2022**, 1–20. [[CrossRef](#)]
20. Liu, L.; Miao, S. A new simple one-dimensional chaotic map and its application for image encryption. *Multimed. Tools Appl.* **2018**, *77*, 21445–21462. [[CrossRef](#)]
21. Moysis, L.; Kafetzis, I.; Tutueva, A.; Butusov, D.; Volos, C. Chaos-Based Image Encryption Based on Bit Level Cubic Shuffling. In *Cybersecurity; Studies in Big Data*; Abd El-Latif, A.A., Volos, C., Eds.; Springer: Cham, Switzerland, 2022; Volume 102. [[CrossRef](#)]
22. Soliman, R.F.; Amin, M.; El-Samie, F.E.A. A Modified Cancelable Biometrics Scheme Using Random Projection. *Ann. Data Sci.* **2018**, *6*, 223–236. [[CrossRef](#)]
23. El-Bendary, M.A.M.; El-Azm, A.E.A.; El-Fishawy, N.A.; Al-Hosarey, F.S.M.; Eltokhy, M.A.; El-Samie, F.E.A.; Kazemian, H. JPEG image transmission over mobile network with an efficient channel coding and interleaving. *Int. J. Electron.* **2012**, *99*, 1497–1518. [[CrossRef](#)]
24. Wu, Y.; Yang, G.; Jin, H.; Noonan, J.P. Image encryption using the two-dimensional logistic chaotic map. *J. Electron. Imaging* **2012**, *21*, 013014. [[CrossRef](#)]
25. Belmeguenai, A.; Ahmida, Z.; Ouchtati, S.; Djemii, R. A novel approach based on stream cipher for selective speech encryption. *Int. J. Speech Technol.* **2017**, *20*, 685–698. [[CrossRef](#)]
26. Zhou, Y.; Long, B.; Chen, C. A new 1D chaotic system for image encryption. *Signal Process.* **2014**, *97*, 172–182. [[CrossRef](#)]

27. Augustine, N.; George, S.N.; Pattathil, D. An audio encryption technique through compressive sensing and Arnold transform. *Int. J. Trust Manag. Comput. Commun.* **2015**, *3*, 74. [[CrossRef](#)]
28. Sathiyamurthi, P.; Ramakrishnan, S. Speech encryption using chaotic shift keying for secured speech communication. *EURASIP J. Audio Speech Music. Process.* **2017**, *2017*, 20. [[CrossRef](#)]
29. Farsana, F.; Gopakumar, K. A Novel Approach for Speech Encryption: Zaslavsky Map as Pseudo Random Number Generator. *Procedia Comput. Sci.* **2016**, *93*, 816–823. [[CrossRef](#)]
30. Belazi, A.; Khan, M.; El-Latif, A.A.; Belghith, S. Efficient cryptosystem approaches: S-boxes and permutation–substitution-based encryption. *Nonlinear Dyn.* **2017**, *87*, 337–361. [[CrossRef](#)]
31. Abd El-Latif, A.A.; Li, L.; Wang, N.; Han, Q.; Niu, X. A new approach to chaotic image encryption based quantum chaotic system, exploiting color spaces. *Signal Process.* **2013**, *93*, 2986–3000. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.