electronics

MDPI

Article

Simultaneous Audio Encryption and Compression Using Compressive Sensing Techniques

Rodolfo Moreno-Alvarado [1], Eduardo Rivera-Jaramillo [2], Mariko Nakano [2] and Hector Perez-Meana [2,*]

[1] División de Ingeniería y Ciencias Exactas, Universidad Anahuac Mayab, Merida Yucatan 97302, Mexico; moar09@hotmail.com

[2] Mechanical and Electrical Engineering School, Culhuacan Campus, Instituto Politécnico Nacional, Mexico City 04440, Mexico; lalomxp@hotmail.com (E.R.-J.); mnakano@ipn.mx (M.N.)

* Correspondence: hmperezm@ipn.mx; Tel.: +52-5551-02-9405

Received: 6 April 2020; Accepted: 15 May 2020; Published: 22 May 2020

Abstract: The development of coding schemes with the capacity to simultaneously encrypt and compress audio signals is a subject of active research because of the increasing necessity for transmitting sensitive audio information over insecure communication channels. Thus, several schemes have been developed; firstly, some of them compress the digital information and subsequently encrypt the resulting information. These schemas efficiently compress and encrypt the information. However, they may compromise the information as it can be accessed before encryption. To overcome this problem, a compressing sensing-based system to simultaneously compress and encrypt audio signals is proposed in which the audio signal is segmented in frames of 1024 samples and transformed into a sparse frame using the discrete cosine transform (DCT). Each frame is then multiplied by a different sensing matrix generated using the chaotic mixing scheme. This fact allows that the proposed scheme satisfies the extended Wyner secrecy (EWS) criterion. The evaluation results obtained using several genres of audio signals show that the proposed system allows to simultaneously compress and encrypt audio signals, satisfying the EWS criterion.

Keywords: extended Wyner secrecy (EWS); compressive sensing (CS); M-sequence; chaotic mixing; Pearson correlation coefficient; UACI; NSCR

1. Introduction

The large amount of digital information transmitted over unsecure channels has led to the necessity of developing efficient schemes for increasing the amount of information transmitted over the existing unsecure communication channels, as well as improving the security of the transmitted information. Thus, to meet these two requirements, many efforts have been undertaken that intend to develop encoding schemes able to simultaneously compress and encrypt audio signals, before their transmission over unsecure communication channels [1,2]. These topics have attracted the attention of a significant number of researchers, consequently leading to the development of several efficient schemes, which firstly compress and subsequently encrypt the compressed information. These schemes intuitively simplify the encryption task because the redundant information has been eliminated during the compression operation. However, because the compressed information is stored before encryption, its security may be compromised because it can be accessed before performing the encryption task. To overcome this problem, several schemes have been proposed in which the information is firstly encrypted and then the resulting information is compressed [1]. The main disadvantage of such schemes is the fact that a lossless compression scheme must be used to avoid the encrypted information being destroyed. A suitable approach to reduce these limitations is the

development of algorithms allowing the simultaneous encryption and compression audio signals, such as those based on compressive sensing [3,4], which is a suitable scheme for encryption of digital information [5].

Because of the growing number of practical applications, compressive sensing has attracted the attention of a large number of researchers working in fields such as audio, image, and video processing [6,7]. As a result, several algorithms able to simultaneously encrypt and compress digital information, based on compressive sensing techniques, have been proposed during the last years [4,8,9], because these schemes have the capacity to meet these requirements simultaneously, using simple matrix operations. In encryption systems based on compressive sensing, the encoding signal is estimated transforming the input frame into a sparse one, using a discrete cosine transform (DCT). The transformed frame is then multiplied by a sensing matrix whose row number is much smaller than its columns; such that when a compressive sensing (CS) approach is used, the audio signal can be simultaneously compressed and encrypted. Thus, using CS, an encrypted signal is also obtained, because to properly decode the encoded signal, the sensing matrix used for decoding must be the same as that used in the encoding stage [5]. Thus, the sensing matrix can be considered as a private key of the CS-based encryption-compression system [4,8,9].

The CS-based joint compression and encryption system has several advantages. The decoding is carried out using only standard matrix operation, and thus it generally has lower computational complexity, compared with other previously proposed systems [4,8,9]. Because the signals to be encoded must be firstly segmented and transformed into sparse signals before applying the CS, each audio segment can be independently encoded and sent in any order to the receiver side. This is an important advantage of the CS-based system and other block-based encryption schemes, used to jointly compress and encrypt any kind of audio signals. However, several drawbacks must be solved to develop trustworthy CS-based audio compression and encryption systems. Firstly, to properly recover the original signal, the sensing matrixes used in the transmission and reception stages must be the same [5]; it is necessary to have a mechanism to allow the generation of the same sensing matrix in both the encryption and des-encryption stages. Second, because both the encryption and des-encryption stages use only linear operations, the security of the CS-based encryption system must be ensured [3,10–12].

Taking in account the requirements described above, this paper proposes a compression-encryption system based on CS, in which the audio signal is firstly segmented into L non overlapping frames of 1024 samples. Each frame is then independently compressed and encrypted using a different sensing matrix for each frame, which is generated using three secret keys provided by the user. These secret keys are transmitted to the receiver side before sending the digital information, encrypted with a public key cryptography algorithm. The rest of this paper is organized as follows. Section 2 presents the development of CD-based proposed system, together with a review of the compressive sensing and an analysis of the security of proposed system. Section 3 provides a detailed evaluation of proposed system and, finally, Section 4 contains the conclusions of this research.

2. Proposed System for Simultaneously Compression and Encryption of Audio Signals

This section provides a description of proposed encoding and decoding system, shown in Figure 1, that allows the simultaneous encryption and compression of audio signals. In the proposed structure, the incoming audio signal X(k) is segmented in frames of "n" samples, which are encrypted with a compression rate of n/m, where m is the number of samples in the compressed frame, using a CS approach. To this end, firstly, the user inserts the values n and m into the encoder stage together with three secret keys k_1, k_2, and k_3 provided by the user, which are then used to estimate the sensing matrix in the transmission stage. Because these secret keys are also required for estimation of the sensing matrix in the decoder stage, they are transmitted to the receiver side encrypted using the Rivest, Shamir, and Adleman (RSA) public key algorithm. This allows that the proposed system operates in a multiuser form and even with different frame sizes and compression rates for each possible user.

In the receiver side, using the secret keys k_1, k_2, and k_3, the sensing matrix required for decoding is generated.

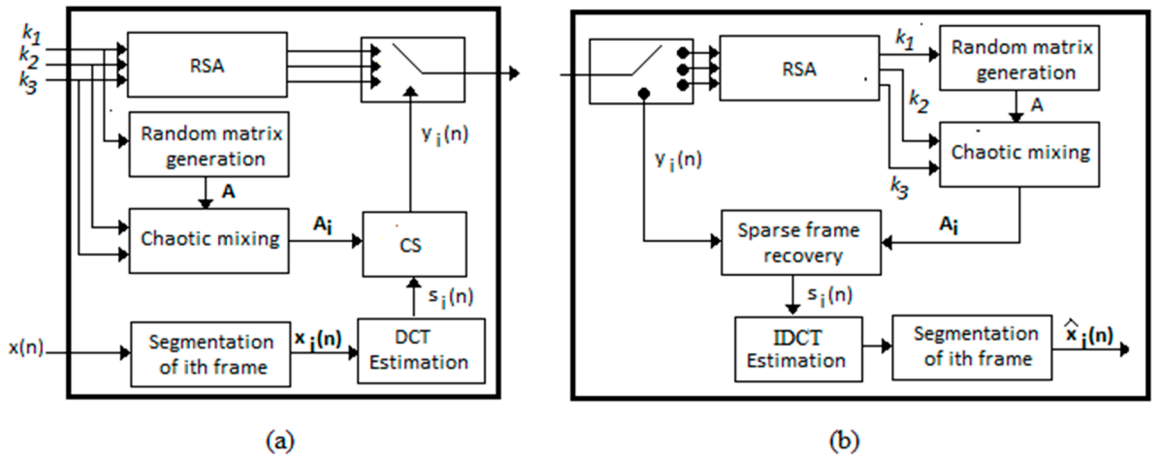


Figure 1. Proposed encryption/des-encryption system. (**a**) Encryption and compression stage, (**b**) des-encryption and decompression stage. DCT, Discrete Cosine Transform; IDCT, Inverse DCT; CS, Compressive Sensing; RSA, Rivest, Shamir, and Adleman.

After matrix $\boldsymbol{A}$ is generated, the sensing matrix $\boldsymbol{A}_0$, of size n × m, used for compressing and encrypting the first block of audio signal, is estimated using the chaotic mixing approach described in Section 2.1. Next, the first block of the input signal given by $\boldsymbol{X}_0(\mathrm{k}) = \boldsymbol{X}(\mathrm{k})$, k = 1, 2, . . . , n is extracted, which is transformed using the (DCT) to estimate a sparse representation of such block, $\boldsymbol{S}_\mathrm{i}$. Then, $\boldsymbol{S}_1$ is multiplied by the sensing matrix $\boldsymbol{A}_0$ to generate the compressed and encrypted version of the first block of input signal $\mathbf{y}_1$, which is transmitted to the received side. In general, the input signal $\mathbf{X}(\mathrm{k})$ is segmented in non-overlapped blocks given by $\boldsymbol{X}_\mathrm{i}(\mathrm{k}) = \boldsymbol{X}_\mathrm{i}(\mathrm{k} + \mathrm{in})$, k = 1, 2, . . . , n; i = 1, 2, . . . , which are then transformed using the DCT to generate a sparse frame, $\mathbf{S}_\mathrm{i}$. Next, using the chaotic mixing method [13,14], the n × m sensing matrix of the *i*-th frame, $\boldsymbol{A}_\mathrm{i}$, is generated from the random matrix $\boldsymbol{A}$. This allows to generate a different sensing matrix for each frame, without significantly increasing the computational complexity, satisfying at the same time the extended Wyner secrecy (EWS) criterion [12]. Next, the sparse vector $\mathbf{S_i}$, estimated using the DCT, is multiplied the sensing matrix $\boldsymbol{A}_\mathrm{i}$ to obtain the encrypted frame $\boldsymbol{y}_1$ with a compression rate of n/m, which is sent to the reception side.

In the receiver stage, provided in Figure 1b, the received information is decoded using the RSA des-encryption module, which allows to recover the values of m and n as well as the users secret keys k_1, k_2, and k_3. These parameters are then used to generate the matrix $\boldsymbol{A}$ and then, using the chaotic mixing method, the sensing matrix for the *i*-th block, $\boldsymbol{A}_\mathrm{i}$, is generated in the same form as it is estimated in the encoding stage. Next, the sensing matrix $\mathbf{A}_\mathrm{i}$ and the input frame $\mathbf{y}_\mathrm{i}(n)$ are fed into the CS recovery stage to obtain $\hat{S}_l$. Then, the inverse DCT (IDCT) of $\hat{S}_l$ is computed to obtain $\hat{X}_l$, which is then concatenated with the previously decoded frames to estimate the decoded signal. The encoding and decoding process described above is performed with each frame of input signal $\boldsymbol{X}$. The following sections provide a description of each stage of the proposed system.

2.1. Sensing Matrix Generation

The sensing matrix, $\boldsymbol{A}$, required in a CS-based audio compression system, becomes the secret key used for the proposed encryption system. Thus, to obtain sufficiently accurate signal decoding, the sensing matrix $\boldsymbol{A}$ must satisfy the restrictive isometry property (RIP) given by [6,15,16]

$$(1 - \delta_k) \left[\!\left[\boldsymbol{AS}\right]\!\right]_2^2 \leq \left[\!\left[\boldsymbol{AS}\right]\!\right]_2^2 \leq (1 + \delta_\mathrm{k}) \left[\!\left[\boldsymbol{AS}\right]\!\right]_2^2, \quad (1)$$

where $0 \leq \delta_k \leq 1$. Thus, because

$$\| \boldsymbol{AS} \|_2^2 = (\boldsymbol{AS})^{\mathrm{T}} \mathbf{AS} = \boldsymbol{S}^T \boldsymbol{A}^T \boldsymbol{AS} \tag{2}$$

and assuming that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \sigma^2\mathbf{I}$, from Equation (1), it follows that

$$(1-\delta_k)\,[\![\boldsymbol{S}]\!]_2^2 \leq \boldsymbol{\sigma}^2\,[\![\boldsymbol{S}]\!]_2^2 \leq (1+\delta_{\mathrm{k}})\,[\![\boldsymbol{S}]\!]_2^2. \tag{3}$$

Thus, $\boldsymbol{A}$ satisfies the RIP if $\sigma^2 = 1$. Then, if the sensing matrix, $\boldsymbol{A}$, satisfies (3), the signal **S** can be accurately recovered [6]. Then, in the encoding stage, the sensing matrix **A** is constructed using a pseudo random number generator whose initial value is the key, k_1, provided by the user, while in the decoding stage, using the same user key, k_1, the sensing matrix, **A,** required to decode **S**, is generated. To this end, firstly, $(L/2)^2$ pairs of uniformly distributed random numbers (U_j, V_j), j = 0, 1, . . . , $L/2 - 1$, are generated [16]. Next, using the Marsaglia polar method [16], the $(L/2)^2$ pairs of uniformly distributed random numbers (U_j, V_j) are converted into L^2 Gaussian distributed random numbers used to estimate the matrix **A**. To this end, L uniformly distributed random numbers are computed (U_j, V_j), j = 0, 1, . . . , $L/2 - 1$.

$$S^2 = V_j^2 + U_j^2, \tag{4}$$

then

$$\boldsymbol{A}(p, 2j) = U_j \sqrt{\frac{-2\ln(S)}{S}}, \;\; j = 0,\, 1,\, \ldots, \frac{L}{2} - 1; \;\; p = 0,\, 1,\, \ldots,\, L-1 \tag{5}$$

$$\boldsymbol{A}(p, 2j+1) = V_j \sqrt{\frac{-2\ln(S)}{S}}, \;\; j = 0,\, 1,\, \ldots, \frac{L}{2} - 1; \;\; p = 0,\, 1,\, \ldots,\, L-1 \tag{6}$$

The matrix $\boldsymbol{A}$, described by (5) and (6), will be used to generate the sensing matrixes required to compress and encrypt the input signal, simultaneously satisfying the RIP property [15,17] and the EWS criterion [10].

To satisfy the EWS criterion [10], a different sensing matrix must be used in each frame, which must also satisfy the RIP. To generate a different sensing matrix for each frame, the chaotic mixing scheme [13,14] is applied to random matrix **A**, as described in Figure 2 and Equations (7)–(15).

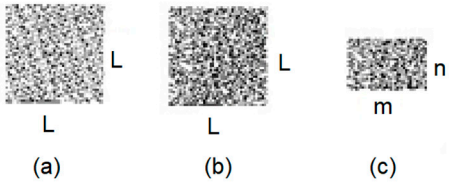


Figure 2. Generation of matrix $\mathbf{A}(x, y)$, using chaotic mixing. (**a**) Matrix **A** during frame $j-1$, (**b**) matrix **A** during frame j, and (**c**) matrix $\mathbf{A}_i(x, y)$ during frame j.

To modify only the position of the matrix elements and not their values themselves, the chaotic mixing method is used, because it performs a mapping from $M_L \rightarrow M_L$. To achieve this goal, the location of the element, (x, y) of the matrix, $\boldsymbol{A}(x, y)$, is modified using a matrix $\boldsymbol{B}_{\mathrm{L}}$ given by [13,14]

$$\boldsymbol{B}_{\mathrm{L}} = \begin{pmatrix} 1 & 1 \\ k_2 & k_2+1 \end{pmatrix}, \tag{7}$$

where $\boldsymbol{B}_{\mathrm{L}}$ satisfies that $\det(\boldsymbol{B}) = 1$ and $trace(\boldsymbol{B}) = \lambda_1 + (1/\lambda_1)$, where λ_1 is the largest eigenvalue of $\boldsymbol{B}_{\mathrm{L}}(k)$. Consider the largest and smallest eigenvalues of $\boldsymbol{B}_{\mathrm{L}}(k)$, which are given by [13,14]

$$\lambda_1 = 1/2\left[k_2 + 2 + \sqrt{4k_2 + k_2^2}\right] \tag{8}$$

and

$$\lambda_2 = 1/2\left[k_2 + 2 - \sqrt{4k_2 + k_2^2}\right]. \tag{9}$$

Next, from (5) and (6), it follows that

$$\lambda_1\lambda_2 = \frac{1}{4}\left(\left[k_2 + 2 + \sqrt{4k_2 + k_2^2}\right]\left[k_2 + 2 - \sqrt{4k_2 + k_2^2}\right]\right), \tag{10}$$

$$trace(\boldsymbol{B}) = k_2 + 2 = \lambda_1 + \lambda_2 = \lambda_1 + (1/\lambda_1), \tag{11}$$

and

$$\det(\boldsymbol{B}) = (k_2 + 1) - k_2 = 1. \tag{12}$$

Thus, because $\boldsymbol{B}_{\mathrm{L}}$ is not singular and thus $\boldsymbol{B}_{\mathrm{L}}^{-1}$ exists, from (8)–(12), it follows that positions of the elements of $\boldsymbol{A}(x, y)$ are estimated using $\boldsymbol{B}_{\mathbf{L}}$ as follows

$$\begin{pmatrix} x_{r+1} \\ y_{r+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k_2 & k_2 + 1 \end{pmatrix} \begin{pmatrix} x_r \\ y_r \end{pmatrix} mod(L),\ r = 1, 2, \ldots, k_3, \tag{13}$$

$$x_{r+1} = (x_r + y_r)\ mod(L);\ r = 1, 2, \ldots, k_3, \tag{14}$$

$$y_{r+1} = (k_2 x_r + (k_2 + 1) y_r)\ mod(L);\ r = 1, 2, \ldots, k_3, \tag{15}$$

where k_2 and k_3 are the secret keys provided by the user, where user's key k_3 determines the required iterations. Thus, Equations (14) and (15) are iterated from $r = 1, 2, \ldots, k_3$, where $(x_r, y_r)\ \epsilon\ [0, L-1]$ denotes the position of (x, y) during the r-th iteration. Thus, using the chaotic mixing method, a new matrix is estimated in each frame. Thus, using the chaotic mixing approach, firstly, the random matrix $\boldsymbol{A}(x, y)$, at the i-th frame, is estimated after iterating k_3 times, using (14) and (15), the random matrix estimated in the $(i-1)$th frame. Next, using $\boldsymbol{A}(x, y)$, the $n \times m$ sensing matrix during i-th frame, $\boldsymbol{A}_i(x, y)$, is given by $\boldsymbol{A_i}(x, y) = \boldsymbol{A}(x, y)$, $x = 0, 1, \ldots, n-1$; $y = 0, 1, 2 \ldots, m-1$.

2.2. Public Key Encryption of Secrete User Key

In the encoding stage of the proposed system, firstly, the compression parameters m and n together with the user secret keys k_1, k_2, and k_3 are encrypted with the public key RSA algorithm [14,18–21], whose security depends on the difficulty of factoring large integer numbers into their prime components. In order for the transmitter $\boldsymbol{A}$ to be able to send the above information to the receiver $\boldsymbol{B}$, using the RSA algorithm, the receiver $\boldsymbol{B}$ must send to the transmitter $\boldsymbol{A}$ the product of two prime secrete numbers, $\boldsymbol{N_B}$, where $N_B = p_B q_B$, and p_B *and* q_B are two secrete prime numbers of B, together with a no-secrete public encryption exponent E_B. Thus, for sending the public key, the transmitter, A, must firstly receive from the receiver, B, the product of two prime secrete numbers, N_B, together with its public encryption exponent E_B. Additionally, the receiver generates its secret decryption exponent D_B. Thus, using the parameters received from $\boldsymbol{B}$, the secret key of the proposed scheme is that transmitted by the encoder stage as [18,21]

$$Y_i = (K_i)^{E_B} mod(N_B); i = 1,\ 2,\ 3. \tag{16}$$

Next, using D_B, the receiver decrypts the secret keys, sent by the transmitter and used for generating the sensing matrix, $\mathbf{A}_\mathrm{i}$, as follows [18,21]:

$$K_i = (Y_i)^{D_B} mod(N_B); i = 1,\ 2,\ 3, \tag{17}$$

where D_B satisfies the relation

$$E_B D_B = 1\ mod\ [(p_B - 1)(q_B - 1)], \tag{18}$$

where p_B *and* q_B are two prime secret numbers generated in the reception side. Finally, the receiver sent to the transmitter a conformation massage, given by [18,21]

$$C = M^{D_B} mod(N_B), \tag{19}$$

which is decrypted by the transmitter as follows [18,21]

$$M = C^{E_B} mod(N_B). \tag{20}$$

Because E_B is public, the message M can be recovered by any member of the network, however, only the receiver B may have sent the confirmation of the encryption of the message M.

2.3. Encrypted and Compressed Signal

In the transmission stage, firstly, the input audio signal, **X**(m), is segmented in a set of nonoverlapped frames, such that its i-th frame is given by $\mathbf{X}_i(k) = \mathbf{X}(k + in)$, where $0 \leq k \leq n$. $\mathbf{X}_i$(k) is then transformed to the DCT domain, which provides a k sparse representation with only k << n terms different to zero, that is, $S_i = \Psi \boldsymbol{X}_i$. Finally, the encrypted and compressed signal is computed by multiplying the sparse vector $\boldsymbol{S}_i$ by the i-th, $n \times m$ size sensing matrix $\boldsymbol{A}_i$. Thus, the i-th frame of the transmitted signal is given by [6,17,22].

$$\boldsymbol{y} = \boldsymbol{A}_i \boldsymbol{S}_i = \boldsymbol{A}_i \Psi \boldsymbol{X}_i, \tag{21}$$

where Ψ denotes the DCT basis functions. Thus, according to the compressive sensing theory, $\mathbf{S_i}$ can be reconstructed if the input signal is represented with at least m samples, where $m \geq O(k \log n)$ [6,17,22].

2.4. Decrypted and Decompressed Signal

The transmitted signal is decrypted and decompressed by minimizing the norm l_1 because, if the received frame $\mathbf{S_i}$ is sparse enough, the probability that the recovered signal is almost equal to the original one is very high [6,17,22], because the norm l_1 improves the signal reconstruction, that is, for a given sensing matrix $\mathbf{A}_i \in R^{n \times m}$ and a received vector $\boldsymbol{y}_i \in \mathrm{R^m}$, the i-th transmitted sparse vector, $\hat{\mathbf{S}}_i$, can be estimated minimizing [6,17,22]

$$\min_{y \in \mathrm{R^m}} \| \boldsymbol{y} - \boldsymbol{A}\hat{\mathbf{S}} \|_1 \text{ given that } \boldsymbol{y} = \boldsymbol{A}\mathbf{S}, \tag{22}$$

using orthogonal matching pursit (OMP) [6]. Finally, the transmitted vector **X** is estimated computing the inverse DCT of $\hat{\mathbf{S}}_i$, that is, $\boldsymbol{X}_i = \Psi^{-1}\mathbf{S}_i$. Because k, $m \ll n$ and $\boldsymbol{S_i}$ is sparse, it can be recovered with about $k \times n \times m$ operations [6], and then, the CS-based scheme is a highly competitive compression-encryption system. However, as in any other encryption system, its security is of great importance. Thus, attending to this fact, the next subsection presents a security analysis of the CS-based encryption system.

2.5. Security Analysis of CS-Based System

The security of the proposed system strongly depends on the fact that the encoding and decoding sensing matrixes be enough different from each other. To carry out this analysis, consider the binary hypothesis testing theory developed by Ramezani-Mayiami et al. [5], using the Norman–Pearson test. Using this theory, it can be shown that, when the same sensing matrix, $\mathbf{A}_i$, is used in both the encoding and decoding stages, the probability of correctly detecting the transmitted signal, $P_s(\alpha)$, is given by [5]

$$P_S(\alpha) = Q\left[Q^{-1}(\alpha) - \frac{\| \boldsymbol{A}_i \boldsymbol{S}_i \|_2}{\sigma}\right]. \tag{23}$$

Meanwhile, when two different matrixes $\mathbf{A}_i$ and $\mathbf{B}_i$ are used in the encoding and decoding stages, the probability of correctly detection, $P_d(\alpha)$, is given by

$$P_d(\alpha) = Q\left[Q^{-1}(\alpha) - \frac{\| \boldsymbol{A}_i\boldsymbol{S}_i \|_2 \langle \boldsymbol{A}_i\boldsymbol{S}_i, \boldsymbol{B}_i\boldsymbol{S}_i \rangle}{\sigma \| \boldsymbol{A}_i\boldsymbol{S}_i \|_2 \| \mathbf{B}_i\boldsymbol{S}_i \|_2}\right], \tag{24}$$

$$P_d(\alpha) = Q\left[Q^{-1}(\alpha) - \frac{\| \boldsymbol{A}_i\boldsymbol{S}_i \|_2 \| \boldsymbol{A}_i\boldsymbol{S}_i \|_2 \| \mathbf{B}_i\boldsymbol{S}_i \|_2 \cos(\theta)}{\sigma \| \mathbf{A}_i\boldsymbol{S}_i \|_2 \| \mathbf{B}_i\boldsymbol{S}_i \|_2}\right], \tag{25}$$

$$P_d(\alpha) = Q\left[Q^{-1}(\alpha) - \frac{\| \mathbf{A}_i\boldsymbol{S}_i \|_2 \cos(\theta)}{\sigma}\right], \tag{26}$$

where θ is the angle between sub-spaces $\| \mathbf{A}_i\boldsymbol{S}_i \|_2$ and $\| \mathbf{B}_i\boldsymbol{S}_i \|_2$. Because $0 \le \cos(\theta) \le 1$ from (23) and (26), it follows that, when $\| \mathbf{A}_i\boldsymbol{S}_i \|_2$ and $\| \mathbf{B}_i\boldsymbol{S}_i \|_2$ are orthogonal sub-spaces, that is, $\cos(\theta) = 0$, the decoded signal is completely useless because, in this situation, $P_d(\alpha) = 0.5$. From the information theoretic perspective, this situation is satisfied when the perfect secrecy is satisfied [5].

Equations (23)–(26) show that, to correctly decode the incoming signal, the encoding matrix must be the same as the decoding sensing matrix. Thus, a possible attack is to try to estimate the sensing matrix using several received frames by means of some blind signal separation methods, such as the independent component analysis (ICA). Thus, it is necessary to determine the conditions that allow to increase the security against ICA or other blind separation analysis. To analyze the security of the CS-based crypto system, it will be assumed that the secrecy of sensing matrix **A** is guaranteed. To this end, consider the plain text to be k sparse, where $k < n$, such that there is at most k elements different from zero in a frame $\mathbf{S}_i$ of length n, such that the CS-based encoded signal is given by [6,10]

$$\mathbf{y} = \boldsymbol{A}_i\boldsymbol{S}_i, \tag{27}$$

where $\boldsymbol{y} \in R^m$ is the encoded vector, and $\boldsymbol{A} \in R^{m\times n}$ and $\boldsymbol{S} \in R^n$ is the input vector. Next, defining $\boldsymbol{A} = \left[\boldsymbol{A}^a, \boldsymbol{A}^b\right]$ and $\boldsymbol{S} = \left[\boldsymbol{S}^a, \boldsymbol{S}^b\right]$, $\mathbf{A}^{\mathrm{a}} \in R^{m\times(n-k)}$, $\mathbf{A}^{\mathrm{b}} \in R^{m\times k}$, $\boldsymbol{S}^a \in R^{(n-k)\times 1}$, $\boldsymbol{S}^b \in R^{k\times 1}$, without loss of generality, assume [10]

$$\boldsymbol{S}^a = [S_1, S_2, S_3, \ldots, S_{n-k}]^T, \tag{28}$$

$$\boldsymbol{S}^b = [S_{n-k+1}, S_{n-k+2}, S_{n-k+3}, \ldots, S_n]^T, \tag{29}$$

$$\mathbf{A}^a = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \ldots, \boldsymbol{a}_{n-k}]^T, \tag{30}$$

$$\boldsymbol{A}^b = [\boldsymbol{a}_{n-k+1}, \boldsymbol{a}_{n-k+2}, \boldsymbol{a}_{n-k+3}, \ldots, \boldsymbol{a}_n]^T \tag{31}$$

and $\boldsymbol{a}_1 \in R^{m\times 1}$. Substituting (28)–(31) into (27),

$$\mathbf{y} = \boldsymbol{A}^a\boldsymbol{S}^a + \boldsymbol{A}^b\boldsymbol{S}^b \tag{32}$$

and using the Moore–Penrose inverse matrix, $\boldsymbol{S}^b$ is given by [10]

$$\boldsymbol{S}^b = \left((\boldsymbol{A}^b)^T\boldsymbol{A}^b\right)^{-1}(\boldsymbol{A}^b)^T\left(\mathbf{X} - \boldsymbol{A}^a\boldsymbol{S}^a\right). \tag{33}$$

Next, consider the conditional entropy function of $\boldsymbol{S}^a$ given **S**, which satisfies [10,19]

$$H(\mathbf{S}^a/\boldsymbol{S}) = \mathbf{0}, \tag{34}$$

where the entropy of **S** and conditional entropies satisfies

$$H(\boldsymbol{S}) \le H(\boldsymbol{S}^a) + H\left(\boldsymbol{S}^b\right), \tag{35}$$

$$H(\boldsymbol{S}/\boldsymbol{S}^a) = H(\boldsymbol{S}, \boldsymbol{S}^a) - H(\boldsymbol{S}^a), \tag{36}$$

$$H(S^a/S) = H(S, S^a) - H(S) \tag{37}$$

and then

$$H(S, S^{\mathrm{a}}) = H(S^a/S) + H(S). \tag{38}$$

Substituting (38) and (35) into (36), from (34), it follows that

$$H(S/S^a) = H(S^a/S) + H(S) - H(S^a), \tag{39}$$

$$H(S/S^a) = H(S^a) + H\left(S^b\right) - H(S^a), \tag{40}$$

$$H(S/S^a) = H\left(S^b\right). \tag{41}$$

Assuming that the entropy of S^b is smaller than or equal to the sum of the entropy of its elements, that is,

$$H\left(S^b\right) \leq H(S_{n-k+1}) + H(S_{n-k+2}) + \cdots + H\left(S_{n-k+j}\right) + \cdots + H(S_n). \tag{42}$$

Using the fact that all elements of S^b given by (33) have the same distribution, it follows that

$$H\left(S^b\right) \leq \sum_{j=1}^{K} \log(M) = k \log(M), \tag{43}$$

where $M = 2^B$ and B is the number of bits used for representing an information sample, that is, an audio sample in an audio signal or a pixel in an image.

Next, consider the conditional mutual information of $\boldsymbol{y}$ and S, given S^a, which is given as [6]

$$I(\boldsymbol{y}; \mathbf{S}/S^a) = H(S/S^a) - H(\mathbf{S}/\boldsymbol{X}, S^a). \tag{44}$$

Substituting (41) into (44), it follows that

$$I(\boldsymbol{y}; S/S^a) = H\left(S^b\right) - H(S/\boldsymbol{X}, S^a), \tag{45}$$

$$I(\boldsymbol{y}; S/S^a) = klog(M) - H(S/\boldsymbol{X}, S^a), \tag{46}$$

$$I(\boldsymbol{y}; S/S^a) \leq klog(M). \tag{47}$$

Next, consider the mutual information between the input vector $\boldsymbol{y}$ and the sensing matrix $\boldsymbol{A}$ given S^a, $I(\boldsymbol{y}; \boldsymbol{A}/S^a)$, which, using the chain rule and the fact that $\boldsymbol{A} = \left[A^a, A^b\right]$, can be expressed as follows [10]:

$$I(\boldsymbol{y}; \boldsymbol{A}/S^a) = I(\boldsymbol{y}; A^a, A^b/S^a), \tag{48}$$

$$I(\boldsymbol{y}; \boldsymbol{A}/S^a) = I(\boldsymbol{y}; A^b/S^a) + I(\boldsymbol{y}; A^a/A^b, S^a), \tag{49}$$

$$I(\boldsymbol{y}; \boldsymbol{A}/S^a) = I(\boldsymbol{y}; A^b/S^a) + H(A^a/A^b, S^a) - H(A^a/\boldsymbol{y}, A^b, S^a), \tag{50}$$

Because $\boldsymbol{A}^{\mathrm{a}}$ is independent of $\boldsymbol{A}^{\mathrm{b}}$ and S is independent of $\boldsymbol{A}$, besides that the elements of $\boldsymbol{A}$ and S are statistically independent, it follows that

$$H(\boldsymbol{A}^{\mathrm{a}}/\boldsymbol{A}^{\mathrm{b}}, \mathbf{S}) = H(\boldsymbol{A}^{\mathrm{a}}/S) - I(\boldsymbol{A}^{\mathrm{a}}, \boldsymbol{A}^{\mathrm{b}}/S), \tag{51}$$

$$H(\boldsymbol{A}^{\mathrm{a}}/\boldsymbol{A}^{\mathrm{b}}, S) = H(\boldsymbol{A}^{\mathrm{a}}/S) - H(\boldsymbol{A}^{\mathrm{b}}/S) - H(\boldsymbol{A}^{\mathrm{a}}, \boldsymbol{A}^{\mathrm{b}}/S). \tag{52}$$

As $\mathbf{A}^{\mathbf{a}}$ and $\mathbf{A}^b$ are statistically independent of S, from (52), it follows that [10]

$$H(\boldsymbol{A}^{\mathrm{a}}/\boldsymbol{A}^{\mathrm{b}}, S) = -H\left(\boldsymbol{A}^{\mathrm{b}}\right) + H(\boldsymbol{A}), \tag{53}$$

$$H(A^{\mathrm{a}}/A^{\mathrm{b}},S) = -H\left(A^{\mathrm{b}}\right) + H(A^{\mathrm{a}}) + H\left(A^{\mathrm{b}}\right), \tag{54}$$

$$H(A^{\mathrm{a}}/A^{\mathrm{b}},S) = \mathrm{H}(\mathbf{A}^{\mathrm{a}}). \tag{55}$$

Next, if $S^a = 0$, from (32), it follows that

$$\boldsymbol{y} = A^{\mathrm{b}}S^{\mathrm{b}}. \tag{56}$$

Then,

$$H(A^a/\boldsymbol{y},A^{\mathrm{b}},\mathbf{S}^a) = \mathrm{H}(A^a/A^{\mathrm{b}}S^{\mathrm{b}},A^{\mathrm{b}},S^a), \tag{57}$$

$$H(A^a/\boldsymbol{y},A^b,S^a) \geq H(A^a/A^b,S^b,A^b,S^a), \tag{58}$$

$$H(A^a/\boldsymbol{y},A^b,S^a) = H(A^a/A^b,\mathbf{S}). \tag{59}$$

Then, from (59), it follows that

$$H\left(A^a/\mathbf{y},\mathbf{A}^{\mathrm{b}},\mathbf{S}^a\right) = H\left(\mathbf{A}^a\right). \tag{60}$$

Next, consider Equation (58),

$$I(\boldsymbol{y};\boldsymbol{A}/\mathbf{S}^a) = \mathbf{I}(\boldsymbol{y};A^{\mathrm{b}}/S^a) + H(A^{\mathrm{a}}/A^{\mathrm{b}},\mathbf{S}^a) - H(A^{\mathrm{a}}/\boldsymbol{X},A^{\mathrm{b}},S^a) \tag{61}$$

and substituting (60) and (55) into (50), it follows that

$$I(\boldsymbol{y};\boldsymbol{A}/S^a) = I(\boldsymbol{y};\mathbf{A}^b/S^a) + H(\mathbf{A}^a) - H(\mathbf{A}^a), \tag{62}$$

$$I(\boldsymbol{y};\mathbf{A}/S^a) = I(\boldsymbol{y};\mathbf{A}^b/S^a). \tag{63}$$

Next, consider the mutual information of $\boldsymbol{y}$ and $\mathbf{A}$, given S^a, which is given by [10]

$$I(\mathbf{y};\mathbf{A}/S^a) = H(A^{\mathrm{b}}/S^a) - H\left(A^b/\mathbf{y},S^a\right). \tag{64}$$

Next, assuming that A^b has k × m entries, which are mutually independent and also independent of S^a, it follows that

$$H\left(A^{\mathrm{b}}/S^a\right) = H\left(A^{\mathrm{b}}\right), \tag{65}$$

$$H\left(A^{\mathrm{b}}/S^{\mathrm{a}}\right) = \sum_{i=1}^{m}\sum_{j=n-k+1}^{n} H\left(a_{ij}\right) \leq kmlog(C). \tag{66}$$

Because

$$I(\boldsymbol{y};\boldsymbol{A}/S^a) = H(A^{\mathrm{b}}/S^a) - H\left(A^{\mathrm{b}}/y,S^a\right) \tag{67}$$

and

$$H\left(A^{\mathrm{b}}/\boldsymbol{y},S^a\right) \geq 0 \tag{68}$$

it follows that

$$I(\boldsymbol{y};\boldsymbol{A}/S^a) \leq kmlog(C). \tag{69}$$

Thus, from (69) and (68), it follows that [10]

$$I(\boldsymbol{y};S/S^a) + I(\boldsymbol{y};\boldsymbol{A}/S^a) \leq klog(M) + kmlog(C). \tag{70}$$

Finally, considering that $S^a = 0$, from (70), it follows that

$$I(\boldsymbol{y};\boldsymbol{S}) + I(\boldsymbol{y};\boldsymbol{A}) \leq klog(M) + kmlog(C). \tag{71}$$

Then, from (68), it follows that

$$\lim_{n\to\infty} \frac{I(\boldsymbol{y};\boldsymbol{S}) + I(\mathbf{y};\boldsymbol{A})}{n} \leq \lim_{n\to\infty} \frac{klog(M) + kmlog(C)}{n} = 0. \tag{72}$$

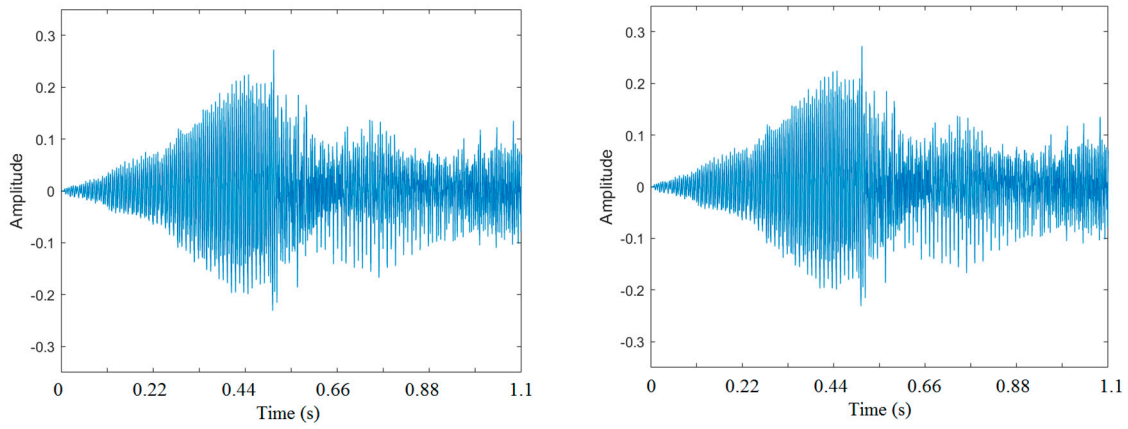
As the mutual information is always positive, that is, $I(\mathbf{y};\mathbf{S}) + I(\mathbf{y};\mathbf{A})$ is always non-negative, it approaches zero as n increases. Then the CS-based joint encryption-compression system satisfies the EWS criterion [10], when the key is used only once.

3. Experimental Results

To evaluate the compression and encryption capability of proposed algorithm, it is necessary to simultaneously compress and encrypt different genres of audio signals, such as Mexican, Caribbean, classic, pop, and rock music, as well as speech signals with different compression rates. To this end, these signals are encoded and decoded using either the same or different sensing matrixes. To evaluate the security performance of proposed system, several tests are performed that are described in the following subsections.

3.1. Waveform Plotting

One of the more common evaluations of the system performance is the waveform plotting, which allows a visual comparison about the similarity between the original audio and the decrypted/decompressed signals. Figure 3a–e show the plot of decrypted/decompressed violin audio signal segment of 1.1 s corresponding to a Bach concert with a sampling rate of 44 kHz and 16 bits/sample without compression, that is, with a bit rate of 704 kb/s, plot in Figure 3a. Figure 3b shows the decrypted signal using the same sensing matrix for both the encryption and decryption process without compression. Figure 3c shows the decrypted signal when the sensing matrix used for encryption is different from that used for decryption; in this case, the original signal was encoded without any compression. Figure 3d plots the decrypted/decompressed signal when the sensing matrixes used for both encryption/compression and decryption/decompression are the same. In this situation, the original signal was encoded with 176 kb/s. Finally, Figure 3e shows the decoded signal when the sensing matrix used for decoding is different to that used during the encoding process.



(**a**) (**b**)

Figure 3. *Cont.*

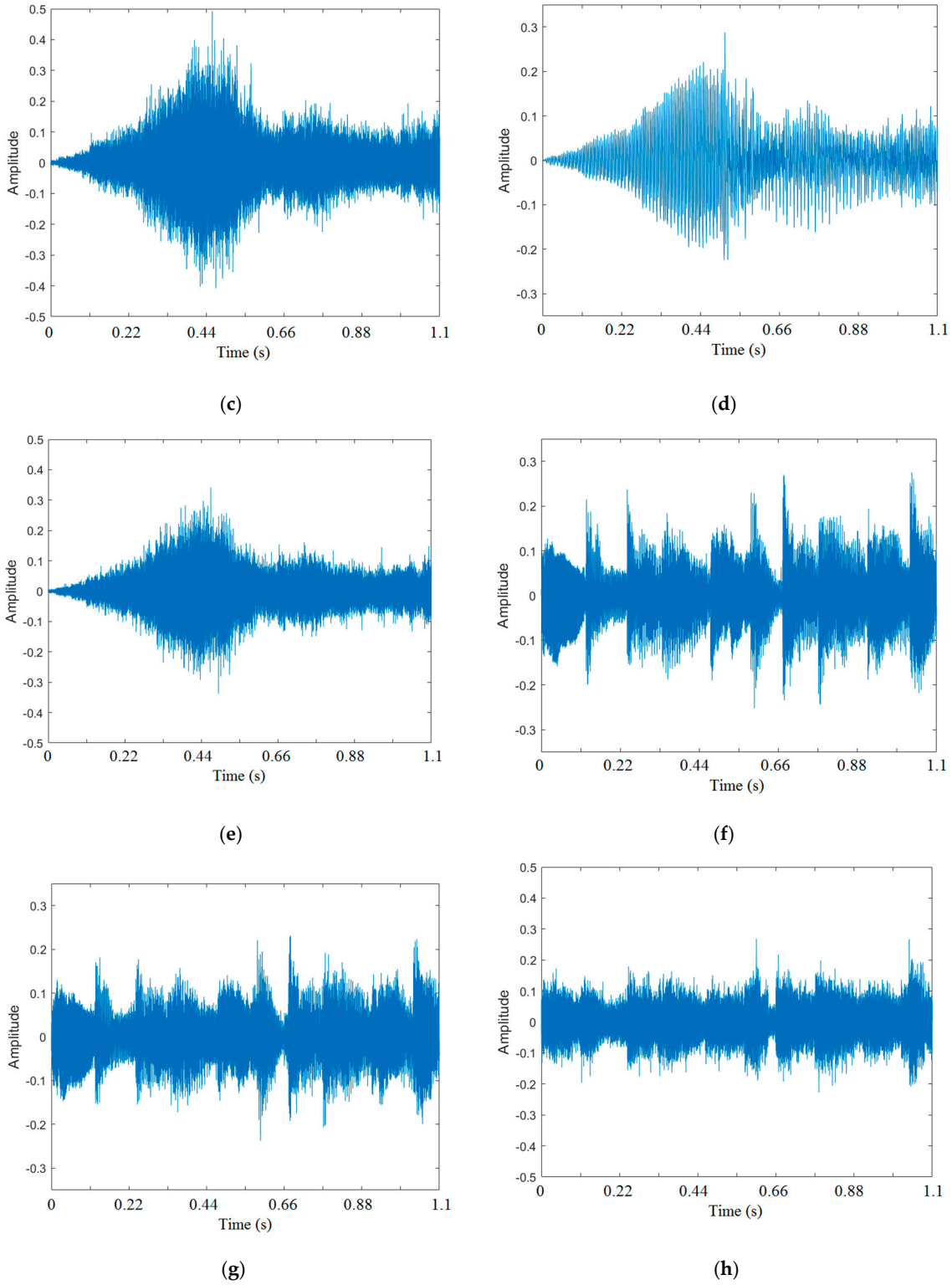


Figure 3. (**a**). Original violin signal with a bit rate of 704 kb/s. (**b**). Decoded violin signal using the same sensing matrix during the encoding process with a bit rate of 704 kb/s. (**c**). Decoded violin signal using different sensing matrixes when the encoding and decoding processes are different, with a bit rate of 704 kb/s. (**d**). Decoded violin signal using the same sensing matrix during the encoding process with a bit rate of 176 kb/s. (**e**). Decoded violin signal using different sensing matrix during the encoding and decoding processes, with a bit rate of 176 kb/s. (**f**). Original popular music segment with a bit rate of 704 kb/s. (**g**). Decoded popular signal using the same sensing matrix during the encoding and decoding process with a bit rate of 352 kb/s. (**h**). Decoded popular signal using the different sensing matrix during the encoding and decoding process with a bit rate of 352 kb/s.

Figure 3f–h show the plot of a decrypted/decompressed signal segment of popular music sampled at 44 kHz. Figure 3f plots the original signal. Figure 3g plots the decoded signal when the encoding and decoding sensing matrixes are the same, and the original signal was compressed to 352 kb/s. Finally, Figure 3h plots the decoded signal obtained when different sensing matrixes are used during the encoding and decoding processes. In this case, the transmission rate was equal to 352 kb/s. Figure 3a–h show that when the same sensing matrix is used for encoding and decoding, the decoded signal closely resembles the original one, independently of the audio signal genre and compression rate used. On the other hand, when the sensing matrix used for decoding is different from that used for encoding, the decoded signal is quite different to the original one, even though, for some genre signals, the envelope has some similarity.

3.2. Spectrogram

Another important evaluation method consists of the comparison of the spectral characteristics of the original, encrypted, and decrypted signals, using different compression rates. Figure 4a–f show the spectrogram of violin music obtained from a Bach concert. These signals are encrypted using compressive sensing with different compression rates. Figure 4a shows the spectrogram of the original Bach concert signal. Figure 4b shows the spectrogram of the encrypted signal without compression. Figure 4c shows the decrypted and decompressed signal when the encoded signal is decoded using the same sensing matrix used during the encoding process. The original signal is encoded with a bit rate of 176 kb/s. Figure 4d shows the spectrogram of the decoded signal when the decoded signal is obtained using a sensing matrix different from that used during the encoding process. Here, the original signal was encoded with a bit rate of 176 kb/s. Figure 4e shows the decoded signal obtained when the original signal is encoded with a bit rate of 88 kb/s and decoded using the same matrix used during the encoding process. Finally, Figure 4f shows the spectrogram of the signal decoded using a sensing matrix different from that used during the encryption and compression processes. These figures show that the spectrum obtained when the sensing matrix used for encoding and decoding is different, and is almost flat, and they strongly infer the signal, shown in Figure 4a, from the knowledge of the signal in Figure 4b. On the other hand, these figures also show that the spectrogram of the signals obtained when the sensing matrix used for encoding and decoding is the same clearly resembles to that of the original one, while they are quite different from those obtained when the sensing matrix used for encoding and decoding is different. Thus, when the decoded signal is obtained using the same sensing matrixes in both the encoding stage and decoding stages, it clearly resembles the original one, while the spectrum of the decoded signal obtained using different sensing matrix in both encoded and decoded stages is clearly different.

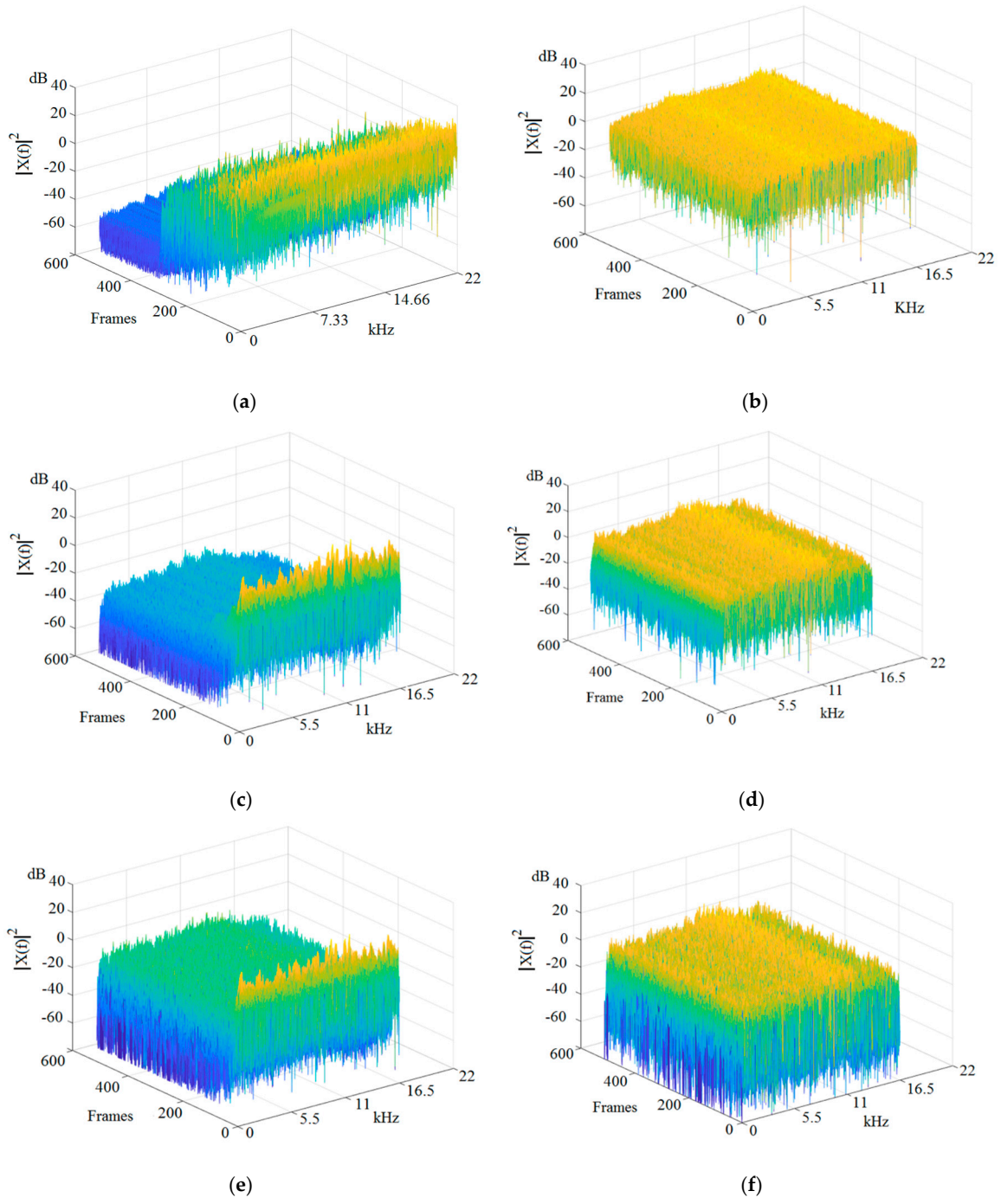


Figure 4. (**a**). Spectrogram of a music signal obtained from a Bach concert with a bit rate of 704 kb/s. (**b**). Spectrogram of a encrypted music Bach violin signal with a bit rate of 704 kb/s. (**c**). Spectrogram of a decoded signal obtained using the same sensing matrix for encoding and decoding process with a bit rate of 176 kb/s. (**d**). Spectrogram of a decoded signal obtained using different sensing matrixes during the encoding and decoding process with a bit rate of 176 kb/s. (**e**). Spectrogram of a decoded signal obtained using the same sensing matrix for encoding and decoding process with a bit rate of 176 kb/s. (**f**). Spectrogram of a decoded signal obtained using different sensing matrixes during the encoding and decoding process with a bit rate of 176 kb/s.

3.3. Pearson Correlation Analysis

Another important parameter used for evaluating the similarity between the original signal and the decoded one is the Pearson correlation coefficient, which is given as follows:

$$\frac{M\sum_{n=0}^{M} x_0(n)x_d(n) - \overline{x}_o(n)\overline{x}_d(n)}{\sqrt{M\overline{x_o{}^2}(n) - (\overline{x}_o(n))^2}\sqrt{M\overline{x_d{}^2}(n) - (\overline{x}_d(n))^2}}, \tag{73}$$

where

$$\overline{x}_{(o,d)}(n) = \sum_{n=0}^{M} x_{(o,d)}(n), \tag{74}$$

$$\overline{x_{(o,d)}{}^2}(n) = \sum_{n=0}^{M}\left(x_{(o,d)}(n)\right)^2 \tag{75}$$

and $x_{(o,d)}(n)$ denotes either the original signal, $x_o(n)$, or the decoded one, $x_d(n)$. Figure 5a–h show the comparison of the Pearson correlation coefficient obtained when the received signal is decoded using the same sensing matrix used for encoding, $R_{xx_S}(k)$, together with the Person correlation coefficient obtained when the received signal is decoded using a sensing matrix different to that used during the encoding process $R_{xx_d}(k)$. Figure 5a shows the Pearson correlation coefficients when the original signal is popular music with a bit rate of 704 kb/s. Figure 5b shows the Pearson correlation coefficients when the original signal is encoded using 352 kb/s. Figure 5c,d shows the Pearson correlation coefficients when popular music is encoded with 176 kb/s and 88 kb/s, respectively. Figure 5e,f show the Pearson correlation coefficients when the original signal is a segment of a Bach concert signal with bit rates of 704 kb/s and 352 kb/s, respectively. Moreover, Figure 5g,h show the Pearson correlation coefficients for each frame when the original Bach concert signal is encoded with bit rates of 176 kb/s and 88 kb/s, respectively. Figure 5i,j show the dispersion diagram of original and decoded popular music audio signals with a bit rate of 352 kb/s. Figure 5i shows that, when the same matrix is used for encoding and decoding, the dispersion diagram is close to a straight line with a slope. This means that, from the decoded audio signal, it is possible to obtain the input one. Figure 5j shows when the sensing matrix used for encoding and decoding is different, the dispersion diagrams are quite spread, such that the decoded signal cannot be inferred from the original one.

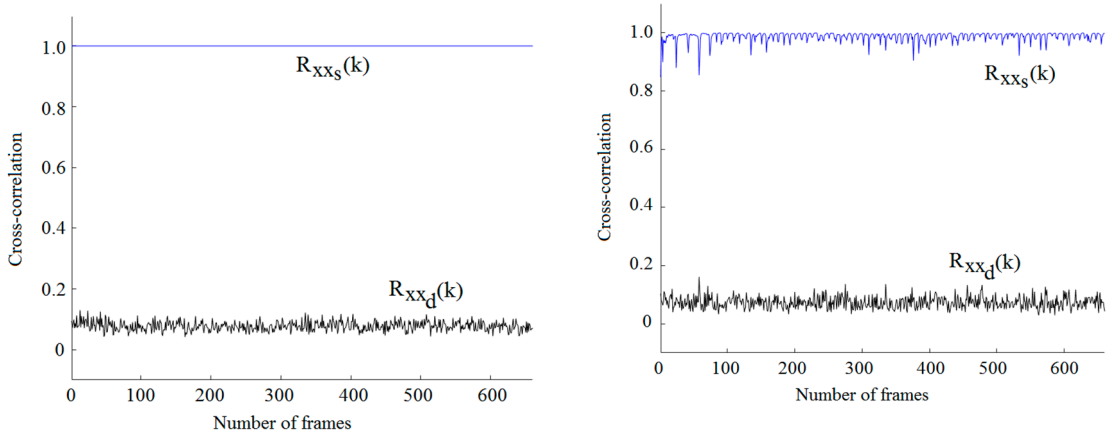


Figure 5. *Cont.*

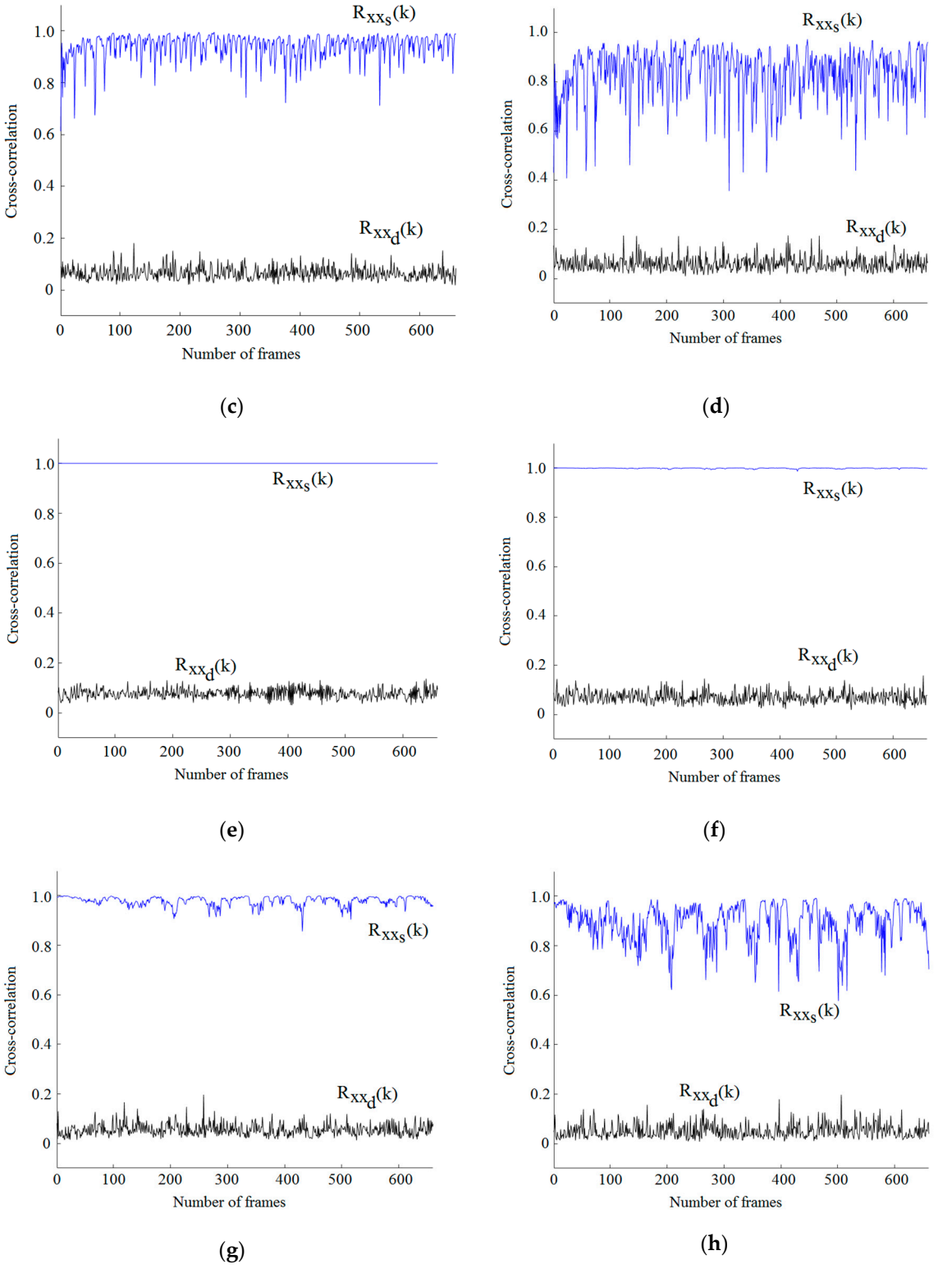
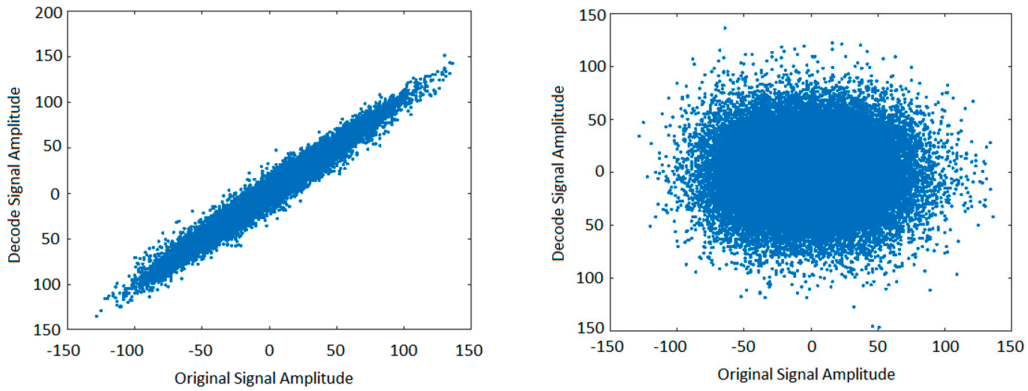


Figure 5. *Cont.*



(**i**) (**j**)

Figure 5. (**a**). Correlation coefficient of popular music using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 704 kb/s. (**b**). Correlation coefficient when popular music is decoded using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 352 kb/s. (**c**). Correlation coefficient of popular music using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 176 kb/s. (**d**). Correlation coefficient when popular music is decoded using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 88 kb/s. (**e**). Correlation coefficient when classic music is decoded using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 704 kb/s. (**f**). Correlation coefficient when classic music is decoded using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 352 kb/s. (**g**). Correlation coefficient when classic music is decoded using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 176 kb/s. (**h**). Correlation coefficient when classic music is decoded using the same, $R_{xx_S}(k)$, as well as a different sensing matrix, $R_{xx_d}(k)$, with 88 kb/s. (**i**). Dispersion diagram of the encoded and decoded signal when the same sensing matrix is used for encoding and decoding popular audio music. (**j**). Dispersion diagram of the encoded and decoded signal when different sensing matrixes are used for encoding and decoding a music signal.

The evaluation results show that the Pearson correlation coefficient, of each frame, between the original and decoded signal when the sensing matrixes used in the encoded and decoded stages are different, is close to zero, around 10^{-2}. From the dispersion diagram shown in Figure 5j, it follows that, if the sensing matrix used for encoding and decoding is different, the dependence between the original and signals decoded is too weak, such that the original signal cannot be estimated from the decoded one. Thus, it would be tough for an intruder to hack the audio signal during the transmission. On the other hand, when the sensing matrixes used for encoding and decoding are the same, the correlation coefficients for each frame are close to one, even when the bit rate used is relatively low. This fact can be observed from the dispersion diagram of Figure 5i, which plots the dispersion diagram between the original and decoded signal. Here, we can see that, when the same matrix is used, the decoded signal closely approaches the original one, grouping around a straight line with a slope. This means that the original signal can be accurately inferred from the decoded one. Thus, the proposed system allows secure and high-quality audio signal transmission.

3.4. Normalized Mean Square Error Analysis

Other important parameter used to evaluate the quality of the proposed system is the normalized mean square error between the original and decoded signals, when the sensing matrixes used for decoding are the same or different to those used for encoding. The normalized mean square error *(MSE)* is given by

$$MSE = \frac{\sum_{k=1}^{1024}(x_o(1024(k-1)+n) - x_d(1024(k-1)+n))^2}{\sum_{k=1}^{1024} x_o^2(1024(k-1)+n)}, \tag{76}$$

where $x_o(n)$ and $x_d(n)$ are the original and recovered signals, respectively. For evaluating the performance of the proposed system, several audio signals were used, such as popular Mexican and Caribbean music, POP music, classic music, and rock music signals sampled at 44 KHz. Each signal is encoded using 16 bits/sample, that is, a bit rate of 704 kb/s. For encoding, as described in Section 2.1, each signal is divided in frames of 1024 samples/frame before computing the DCT, whose resulting vector is multiplied by the sensing matrix. Figure 6a–h show the MSE obtained when the input signals are decoded using either the same or different sensing matrixes used for encoding, that is, the correct or incorrect private secret key. These figures show that the MSE obtained when the decoded sensing matrix is different to that used during the encoding process, that is, an incorrect private decoding key, is larger, whereas when a correct sensing matrix is used, the MSE is close to zero. This fact can be also observed from Figure 5i,f, which show that, when the same matrix is used for encoding and decoding, the decoded signal closely approaches to the original one, which results in an approximation error close to zero. If we consider that the MSE given by (76) can be considered as the inverse of the signal-to-noise ratio (SNR), that is, $MSE^{-1} = SNR$, the evaluation results show that, when the same matrix is used for encoding and decoding, a decoded signal with high SNR is obtained, that is, a high quality signal can be obtained. Meanwhile, when the sensing matrix used for encoding and decoding is different among them, a rather noisy decoded signal with SNR smaller than zero is obtained, which results in an unintelligible decoded signal. Thus, it can be expected that the proposed system allows the secure transmission of the high-quality signal. When the compression rate increases, as can be expected, the quality of the decoded signal becomes lower.

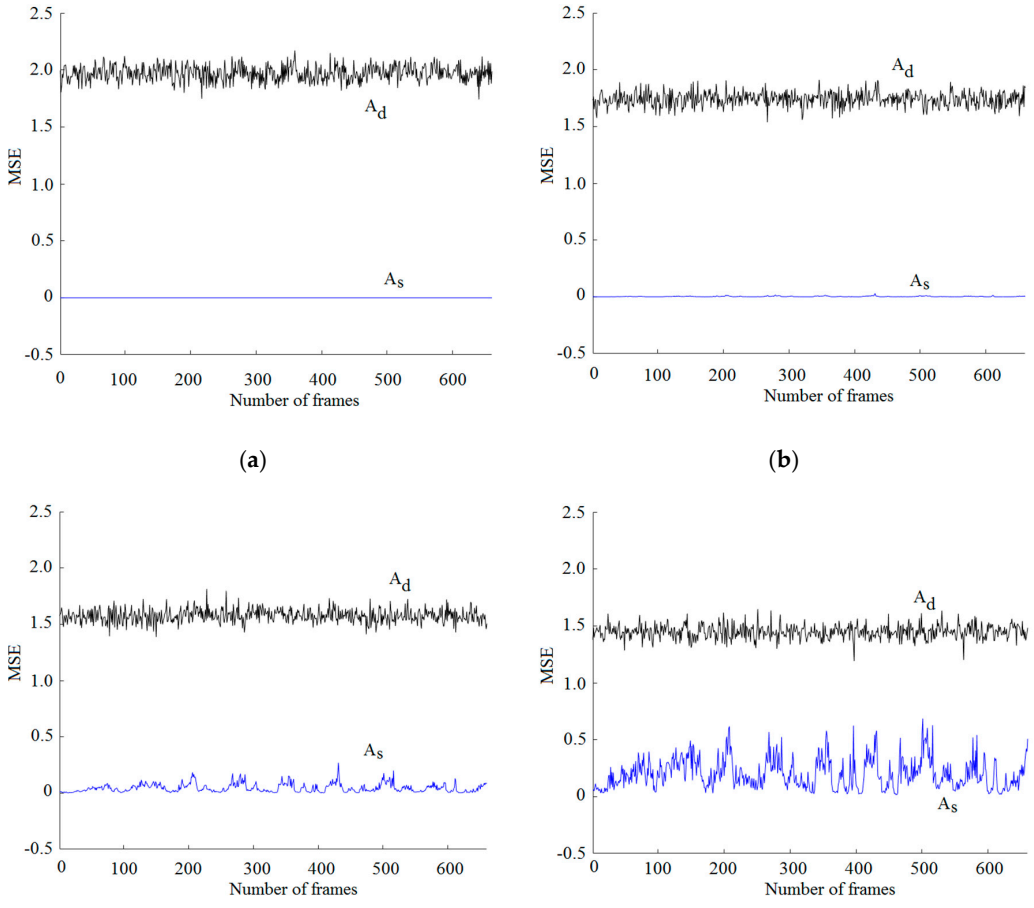


Figure 6. *Cont.*

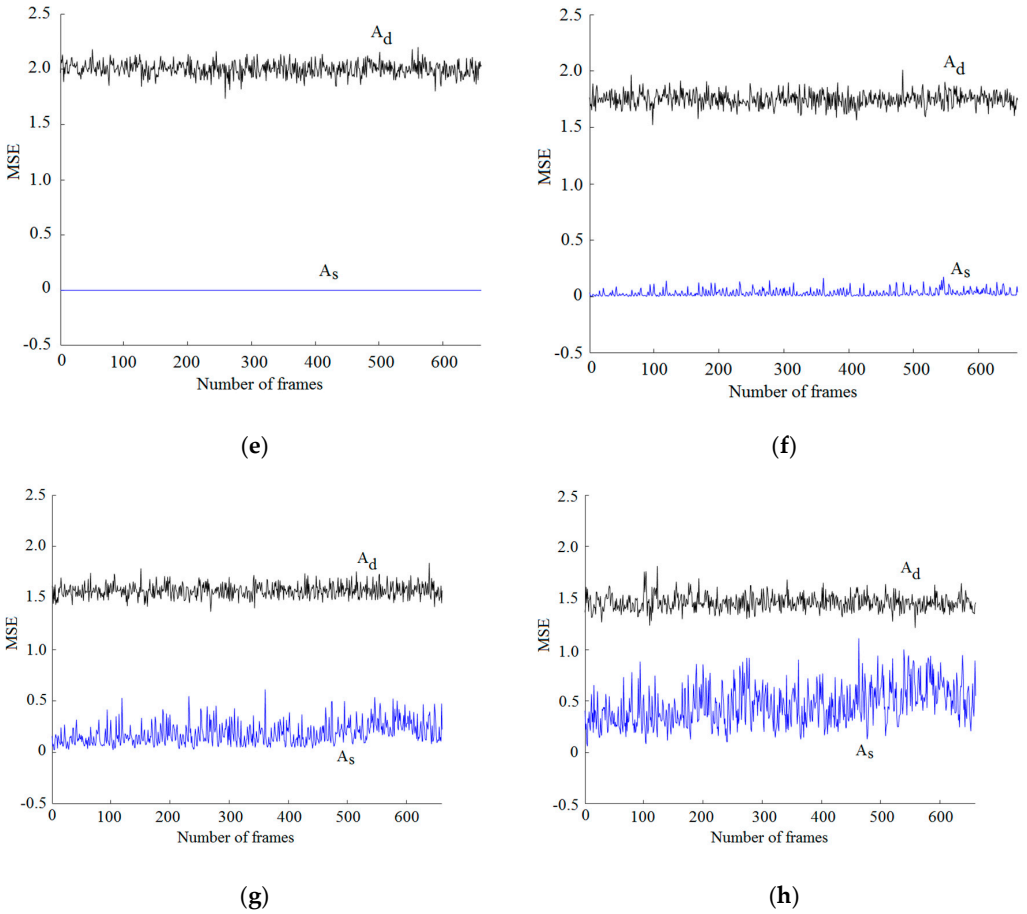


Figure 6. (**a**) Mean square error *(MSE)* estimated when classic music, with a bit rate of 704 kb/s, is decoded using the same, A_s, and different, A_d, sensing matrix. (**b**) *MSE* obtained when classic music, with a bit rate of 352 kb/s, is decoded using the same, A_s, and different, A_d, sensing matrix. (**c**) *MSE* obtained when classic music with a bit rate of 176 kb/s is decoded using the same, A_s, and different, A_d, sensing matrix. (**d**) *MSE* obtained when classic music, with a bit rate of 88 kb/s, is decoded using the same, A_s, and different sensing matrix. (**e**) *MSE* using the same, A_s, and different, A_d, sensing matrix with a bit rate of 704 kb/s. (**f**) *MSE* using the same, A_s, and different, A_d, sensing matrix with a bit rate of 352 kb/s. (**g**) *MSE* using the same, A_s, and different, A_d, sensing matrix with a bit rate of 176 kb/s. (**h**) *MSE* using the same, A_s, and different, A_d, sensing matrix with a bit rate of 88 kb/s.

3.5. Spectral Similarity Analysis

Another metric that can be used for evaluating the security and reconstruction quality of proposed system is the spectral similarity *(SMSE)*, which is given by

$$SMSE(m) = \frac{\sum_{k=1}^{1024}\left(x_{fo}(1024(m-1)+k) - x_{fd}(1024(m-1)+k)\right)^2}{\sum_{k=1}^{1000} x_{fo}^2(1000(m-1)+k)}, \quad (77)$$

where $X_{fo}(1024(m-1)+k)$ and $X_{fd}(1024(m-1)+k)$ is the k-th component of the m-th frame of original and decoded signals, respectively. Figure 7a–d show the spectral similarity obtained when the sensing matrix used for decoding is equal and different to that used in the encoding stage. Figure 7a,b show the spectral similarity obtained when the sensing matrixes equal and different to that used for encoding used classical music signals with bit rates of 704 and 352 bits/s, respectively.

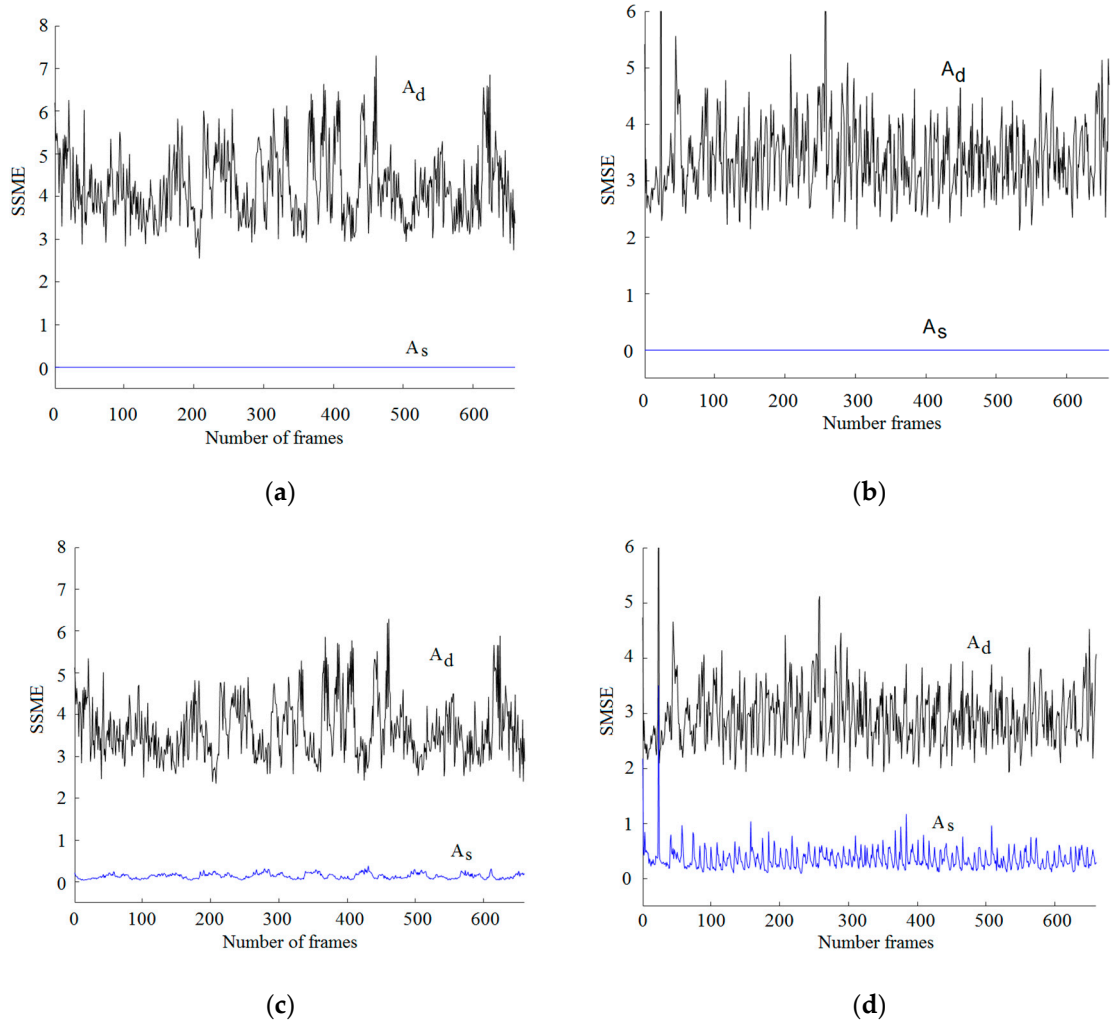


Figure 7. (**a**) Spectral similarity (*SMSE*) obtained when classic music is decoded using the same sensing matrix, with a sampling rate of 704 kb/s. (**b**) Spectral similarity (*SMSE*) obtained when classic music is decoded using the same sensing matrix, with a sampling rate of 352 kb/s. (**c**) Spectral similarity (*SMSE*) obtained when classic music is decoded using the same sensing matrix, with a sampling rate of 176 kb/s. (**d**) Spectral similarity (*SMSE*) obtained when classic music is decoded using the same sensing matrix, with a sampling rate of 88 kb/s.

Figure 7c,d show the spectral similarity obtained when the sensing matrix is equal and different to that used for encoding using a classical music signals with bit rates of 176 kb/s and 88 kb/s, respectively.

The evaluation results show that the *MSE* obtained when the signal is transmitted without and with compression rates of 50% is close to zero, providing secure communications with high quality decoded signals. Meanwhile, when the compression rate increases, the quality of the decoded signal becomes lower.

Tables 1 and 2 show the *MSE*, *SMSE*, and correlation coefficient of the proposed algorithm when it is used for encoding popular Mexican music. Table 2 shows the performance of the proposed scheme when it is used for compressing and encrypting classic music. Table 3 shows the NSCR and UACI parameters obtained when the incoming signal is classic music, popular music, and pop music with different bit rates.

Table 1. Similarity, spectral similarity, and Person correlation obtained using the proposed system when popular music is encoded using a different number of samples/frames. MSE, mean square error; SMSE, spectral similarity.

Samples/Frame	*MSE*		*SMSE*		Pearson-Correlation		kb/s
	Same Matrix	Different Matrix	Same Matrix	Different Matrix	Same Matrix	Different Matrix	
128	0.4723	1.452	1.0219	2.1325	0.8636	0.0605	88
256	0.1920	1.539	0.7180	2.5095	0.9409	0.0651	176
512	0.0333	1.744	0.3524	2.9474	0.9872	0.0730	352
700	2×10^{-6}	2.003	0.1696	3.1471	0.9968	0.0731	492
1024	2×10^{-6}	2.003	1.1×10^{-6}	3.4118	1.0000	0.0774	704

Table 2. Similarity spectral similarity and Person correlation provided by the proposed system when classic music is encoded using a different number of samples/frames.

Samples/Frame	*MSE*		*SMSE*		Pearson-Correlation		kb/s
	Same Matrix	Different Matrix	Same Matrix	Different Matrix	Same Matrix	Different Matrix	
128	0.2863	1.458	0.9343	2.5923	0.8952	0.0510	88
256	0.1114	1.579	0.4949	3.0590	0.9787	0.0530	176
512	0.0260	1.744	0.1394	3.6377	0.9978	0.0700	352
700	0.0066	1.846	0.0623	3.9016	0.9987	0.0699	492
1024	4×10^{-6}	1.994	4.7×10^{-6}	4.2401	1.0000	0.0778	704

3.6. NSCR and UACI Parameters

Other important parameters included in the NIST recommendations to determine the quality of speech encryption are the NSCR and UACI, which determine the number of changing samples and the number of average of changes in the intensity of the encrypted speech, respectively. The Number of Sample Change Rate (NSCR) and Unified Average Changing Intensity (UACI) are given by

$$\text{NSCR} = \frac{1}{N}\sum_{i=1}^{N} D_i \times 100\%, \tag{78}$$

where

$$D_i = \begin{cases} 1, & x_i \neq x_i' \\ 0, & x_i = x_i \end{cases}, \tag{79}$$

$$\text{UACI} = \frac{1}{N * \max(x'(n))}\sum_{i=1}^{N} \left|x_i - x_i'\right| \times 100\%, \tag{80}$$

where $x_i(n)$ and $x_i'(n)$ are the ith sample of two cyphered audio signals, whose original versions differ only in one sample, and N denotes the length of the audio frame. Table 3 provides the NSCR and UACI when the proposed algorithm is required to compress and encrypt several genders of audio signals.

Table 3 shows that the values of UACI and NCSR provided by the proposed scheme are close to the optimum ones reported in the literature [23].

Table 3. UACI and NSCR obtained using the proposed algorithm.

Type	704 kb/s		352 kb/s		176 kb/s	
	UACI	NSCR	UACI	NSCR	UACI	NSCR
Speech	31.91	99.02	34.60	99.02	32.36	99.20
Classic music	29.32	98.96	29.32	98.05	33.70	98.98
Popular music	32.54	99.09	39.13	99.03	33.70	99.06
Pop music	35.05	99.88	39.16	99.12	32.54	99.03

3.7. Comparison with Other Reported Schemes

An important evaluation of the proposed scheme is the comparison of its performance with the performance provided by other previously proposed schemes. Table 4 shows a comparison of the Pearson correlation coefficients and the mean square error provided by the proposed scheme and other previously proposed schemes when the sensing matrix used for decoding the encrypted signal is either the same or different sensing matrix to that used for encoding the audio signals.

Tables 4 and 5 show a comparison of the correlation coefficient and MSE provided by the proposed scheme together with the system proposed by G. Sudhish et al. [3], Sathiyamurthi [23,24] Kordov [25], and when the input signals are classic and popular music, with a bit rate of 704 kb/s. Table 6 shows a comparison of the correlation coefficient and MSE provided by the proposed scheme and the system proposed by G. Sudhish et al. [3], when the input signals are classic and popular music, with bit rates of 352 kb/s and 176 kb/s, respectively. Finally, Table 7 shows a comparison of the NSCR and UACI parameters provided by the proposed scheme together with the system proposed by Sathiyamurthi [23] and Kordov [25] when the input signals are classic and popular music, with a bit rate of 704 kb/s.

Table 4. Person correlation coefficient obtained using the proposed system and those proposed by G. Sudhish et al. [3], Kordov [25], and Sathiyamurthi [24], with a bit rate of 704 kb/s.

Scheme	Proposed		Ref. [7]		Ref. [25]		Ref. [24]	
	Classic	Popular	Classic	Popular	Classic	Popular	Classic	Popular
Same matrix	0.9999	0.9999	0.9998	0.9998	0.9997	0.9998	0.9999	0.9999
Different matrix	0.0774	0.0778	0.0004	0.0003	0.0169	0.0048	0.0384	0.0157

Table 5. Mean square error obtained using the proposed system and those proposed by G. Sudhish et al. [3], Kordov [25], and Sathiyamurthi [24], with a bit rate of 704 kb/s.

Scheme	Proposed (dB)		Ref. [3] (dB)		Ref. [25] (dB)		Ref. [24] (dB)	
	Classic	Popular	Classic	Popular	Classic	Popular	Classic	Popular
Same matrix	−53.27	−53.27	−31.308	−31.302	−20.655	−26.193	−32.596	−33.098
Different matrix	6.2738	5.3298	2.2713	2.7717	5.8798	6.4667	2.4294	2.2632

Table 6. Pearson-correlation coefficient and mean square error obtained using the proposed system and the system proposed by G. Sudhish et al. [3], with bit rates of 352 kb/s and 176 kb/s.

Scheme	Proposed				Sudhish et al. [3]				
	Correlation		*MSE* (dB)		Correlation		*MSE* (dB)		Bit Rate
	Classic	Popular	Classic	Popular	Classic	Popular	Classic	Popular	kb/s
Same matrix	0.9978	0.9872	−15.850	−14.775	0.9989	0.0044	−26.383	−21.307	352
Different matrix	0.0730	0.0530	5.6083	2.4165	0.9971	0.0043	2.0798	2.3970	352
Same matrix	0.9787	0.0530	−7.1670	−9.5311	0.9940	0.0015	−19.066	−16.778	176
Different matrix	0.0530	0.0651	1.9841	1.8730	0.9898	0.0010	2.0548	2.4157	176

Table 7. NSCR and UACI parameters obtained using the proposed system, and the systems proposed by G. Kordov [25] and Sathiyamurthi [24] with a bit rate of 704 kb/s.

Audio Signal	Proposed		Ref. [25]		Ref. [24]
	NSCR	UACI	NSCR	UACI	NSCR
Speech	99.02%	31.09%	99.24%	33.30%	99.99%
Classical	98.02%	29.52%	99.22%	33.26%	99.94%
Popular	99.09%	32.54%	99.08%	37.10%	99.72%

Tables 4–7 show that the proposed scheme provides results that are quite competitive compared with other previously proposed schemes. It provides the same correlation coefficients and smaller

MSE than other previously proposed schemes, when the audio signals are transmitted without compression [3,21,24]. On the other hand, when the audio signals are simultaneously compressed and encrypted, the proposed scheme is quite competitive with other previously proposed schemes [3].

4. Conclusions

This paper presents a CS-based encoding system for jointly encrypting and compressing audio signals. In proposed scheme, the audio signals are firstly segmented in frames of 1024 samples, which are then transformed using the DCT for generating a sparse frame. Each frame is then multiplied by a different sensing matrix for compression and encryption, which is constructed using a Gaussian random number generator and a chaotic mixing scheme. This assures that the sensing matrixes used in the proposed system are different in each frame, and then satisfies the EWS criterion.

The evaluation results obtained show that the proposed algorithm provides a rather secure transmission system with a very good quality of decoded signal, because when the same matrix is used for encoding and decoding; the correlation coefficient is close to one, while the *MSE* and *SMSE* are close to zero. Meanwhile, when the sensing matrixes used for encoding and decoding are different, the correlation coefficients for each frame are close to zero, and *MSE* and *SMSE* become larger than one, even when the bit rates used are relatively low. Besides that, the NSCR and UACI obtained are close to 100% and 33%, respectively. Thus, the proposed scheme allows the secure transmission of high-quality audio signals.

Finally, the evaluation results show that the proposed scheme provides results that are quite competitive compared with other previously proposed schemes. It also provides the same correlation coefficients and smaller *MSE* than other previously proposed schemes, when the audio signals are transmitted without compression, whereas when the audio signals are simultaneously compressed and encrypted, the proposed scheme is quite competitive compared with other previously proposed schemes.

Author Contributions: The authors contributions are as follows: conceptualization, R.M.-A. and H.P.-M.; methodology and software, E.R.-J.; validation, R.M.-A., H.P.-M., and M.N.; formal analysis, E.R.-J. and R.M.-A.; investigation, M.N.; data analysis, R.M.-A.; writing and original draft preparation, H.P.-M.; review, R.M.-A. and M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the National Science and Technology Council of Mexico and the National Polytechnic Institute of Mexico for the financial support provided during the realization of this research.

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

References

1. Gadanayak, B.; Prodhan, C. Selective Encryption of MP3 Compression. In Proceedings of the International Conference on Information Systems and Technology, Shanghai, China, 4–7 December 2011; pp. 23–26.
2. Kaur, M.; Kaur, S. Survey of various encryption techniques for audio data. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2014**, *4*, 1314–1317.
3. Sudhish, G.; Nishanth, A.; Deepthi, P. Audio security through compressive sensing and cellular automata. *Multimed. Tools Appl.* **2005**, *74*, 10291–10417.
4. Cambareri, V.; Mangia, M.; Plow, F.; Pareschi, C.; Rovatti, R.; Setti, G. Low complexity multiclass encryption by compressed sensing. *IEEE Trans. Signal Process.* **2015**, *63*, 2183–2195. [CrossRef]
5. Ramezani-Matimi, M.; Bafghi, H.; Seyfe, B. Compressive sensing encryption: Compresive sensing meets detection theory. *J. Commun.* **2018**, *13*, 82–87. [CrossRef]
6. Eldar, Y.; Kutiniok, G. *Compressive Sensing: Theory and Applications*; Cambridge University Press: Cambridge, MA, USA, 2012.
7. Parkale, Y.; Nalbalwar, S. *Application of 1-D Discrete Wavelet Transform based Compressed Sensing Matrices for Speech Compression*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 20, pp. 1–60.

8. Duta, C.; Gheorhe, L.; Tapus, N. Performance Comparison of Voice Encryption Algorithms Implemented in Blackfin Platform. In Proceedings of the International Conference on Information Systems, Security and Privacy, Rome, Italy, 19–21 February 2016; pp. 169–191.
9. Ponnanian, D.; Chandranbabu, K. Crypt Analysis Compression-encryption algorithm and a modified scheme using compressive sensing. *Optik* **2017**, *147*, 265–276.
10. Yang, Z.; Yan, W.; Xiang, Y. On the Security of Compressed Sensing-Based Signal Cryptosystem. *IEEE Trans. Emerg. Top. Comput.* **2015**, *3*, 363–371. [CrossRef]
11. Moreno-Alvarado, R.; Rivera-Jaramillo, E.; Nakano, M.; Perez-Meana, H. Joint Encryption and Compression of Audio Based on Compressive Sensing. In Proceedings of the International Conference on Telecommunications and Signal Processing, Budapest, Hungary, 1–3 July 2019; pp. 58–61.
12. Al-Azawi, M.M.; Gaze, A. Combined speech compression and encryption using chaotic compressive sensing with large key size. *IET Signal Process.* **2018**, *12*, 214–218. [CrossRef]
13. Reyes, R.; Cruz, C.; Nakano, M.; Perez-Meana, H. Digital Video Watermarking in DWT Domain Using Chaotic Mixtures. *IEEE Lat. Am. Trans.* **2010**, *8*, 304–310. [CrossRef]
14. Alvarez-Hernandez, M.; Shinbrot, T.; Zalc, J. Practical chaotic mixing. *Chem. Eng. Sci.* **2002**, *57*, 3749–3753. [CrossRef]
15. Candes, E. Compressive Sampling. In Proceedings of the International Congress of Mathematicians, Madrid, Spain, 20–30 August 2006; pp. 1–20.
16. Marsaglia, G.; Bray, T.A. A Convenient Method for Generating Normal Variables. *SIAM Rev.* **1964**, *6*, 260–264. [CrossRef]
17. Donoho, D. Compressed sensing. *Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]
18. Cohen, A.; Parhi, K. Architecture Optimizations for the ERSA Public Key Cryptosystems: A Tutorial. *IEEE Circuit Syst. Mag.* **2011**, *11*, 24–34. [CrossRef]
19. Cover, T.; Thomas, J. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 1991.
20. Van Tilborg, H.; Jalodia, S. *Encyclopedia of Cryptography and Security*; Springer: Boston, MA, USA, 2011.
21. Newman, D.; Omura, J.; Pickholtz, R. Public key management for network security. *Netw. Mag.* **1987**, *1*, 11–16. [CrossRef]
22. Candes, E.; Romberg, J.; Tao, T. Robust uncertainty principles; exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509. [CrossRef]
23. Sathiyamurthi, P.; Ramakrishnan, S. Speech encryption algorithm using FFT and 3D-Lorenz–logistic chaotic map. *Multimedia Tools Appl.* **2020**, 79. [CrossRef]
24. Sathiyamurthi, P.; Ramakrishnan, S. Speech encryption using chaotic shift keying for secure speech communications. *EURASIP J. Audio Speech Music. Process.* **2017**, *20*, 1–11. [CrossRef]
25. Kordov, K. A Novel Audio Encryption Algorithm with Permutation-Substitution Architecture. *Electronics* **2019**, *8*, 530. [CrossRef]