

# Efficient audio cryptosystem based on chaotic maps and double random phase encoding

Elsayed M. Elshamy<sup>1</sup> · El-Sayed M. El-Rabaie<sup>2</sup> · Osama S. Faragallah<sup>1,5</sup> ·  
Osama A. Elshakankiry<sup>1,5</sup> · Fathi E. Abd El-Samie<sup>2</sup> · Hala S. El-sayed<sup>3</sup> ·  
S. F. El-Zoghdy<sup>4,5</sup>

Received: 1 March 2014 / Accepted: 3 April 2015 / Published online: 8 October 2015  
© Springer Science+Business Media New York 2015

**Abstract** This paper presents an efficient audio cryptosystem based on combining chaotic maps with optical encryption. The proposed audio cryptosystem is designed with the potential of enforcing the security level for digital speech communication via shared networks. The proposed audio cryptosystem is built using two security phases. In the first phase, it utilizes a chaotic system using either Baker map or cat map for providing the first security level. In the second phase, it utilizes optical encryption using double random phase encoding (DRPE) for providing the second security level. This second security level represented with DRPE is a physical security that is very hard to attack. The proposed audio cryptosystem is implemented and its performance is evaluated using different audio encryption/decryption quality metrics. The results demonstrated that the proposed audio cryptosystem increases the level of voice security with high degree of confidentiality.

**Keywords** Audio encryption · Arnold cat map · DRPE

✉ Osama S. Faragallah  
osam\_sal@yahoo.com; o.salah@tu.edu.sa

<sup>1</sup> Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

<sup>2</sup> Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

<sup>3</sup> Department of Electrical Engineering, Faculty of Engineering, Menoufia University, Shebin El-Kom 32511, Egypt

<sup>4</sup> Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt

<sup>5</sup> Department of Information Technology, College of Computers and Information Technology, Taif University, Al-Hawiya 21974, Kingdom of Saudi Arabia

## 1 Introduction

The wide popularity of the Internet combined with the rapid advances in cell phones and computer technologies lead to explosive growth of electronic data exchange and digital communications. As a result, digital speech communication is almost used in all activities in our daily life. As examples, it is used in commerce, education, military, politics, e-learning, news telecasting phone, and banking. This means that, a massive amount of sensitive audio data is exchanged in a daily routine over open and shared networks. Because of the rapid growth of digital audio, and data communications, the importance of providing a high level of security becomes a major issue.

The voice information, which is different from text messages and images, has higher redundancy and stronger correlation between samples. In recent years, several image and video scrambling methods have been presented (Zhang et al. 2008; Lin and Chang 2001; Petitcolas et al. 1999; Langelaar et al. 2000; Chen and Lin 2003; Barni et al. 2001; Refregier and Javidi 1995; Hedelin et al. 1999; Yang et al. 1998; Sang et al. 2003; Kwon et al. 2006; Wu and Ng 2002; Shumei et al. 2009; Al Saad and Hato 2014), but the number of audio scrambling techniques is relatively less. Audio scrambling algorithms must disorganize the audio signal and eliminate the correlation between samples. The common audio scrambling methods use 1-D linear mapping which is based on simple algorithms, and this makes these methods easy to be attacked (Zeng et al. 2012).

The main problem in any audio encryption technique is that the audio signal is a slowly time varying signal so, it contains a large number of adjacent samples of similar values. To solve this problem, we need to do the following:

1. Destroy correlation between samples in order to remove any intelligibility.

2. Change formant, pitch and energy of original signal.
3. Fill silent periods within conversation without revealing the secret key.
4. Make processing time small to decrease delay.

Sound waves are characterized in terms of frequency (Hz), amplitude (dB) and phase (degree), whereas frequencies and amplitudes are perceived as pitch and loudness, respectively.

Traditional cryptographic techniques are efficient for the text data only. They cannot be used for securing sensitive audio data because of its huge capacity and high redundancy. Strong cryptographic algorithms are needed to encrypt sensitive audio data before transmission. Therefore, designing efficient audio encryption techniques that can provide high security level to the sensitive audio data are new challenges. This problem has been studied recently by many scientists, and they proposed a number of audio encryption schemes (Al Saad and Hato 2014; Li et al. 2009; Kohad et al. 2012; Sharma 2012; Zeng et al. 2012; Sheu 2011; Elshamy et al. 2013; Zhao et al. 2014; Mermoul and Belouchrani 2010; Maysaa and Qays 2013). These schemes can generally be classified into analog and digital (Al Saad and Hato 2014).

In a digital cryptosystem, the audio signal is sampled with a suitable sampling frequency to break the continuous audio signal into equal short time segments. This stream of discrete audio segments is quantized and encoded to generate a binary data stream. The output binary data stream can be compressed to generate a data signal at a suitable bit rate. At the transmitter side, the binary data stream is encrypted with an encryption algorithm, which changes the sequence of bits by means of block or stream ciphering. The changed sequence is then transmitted via digital modulation (Zhai et al. 2008). Digital encryption has a higher security level than analog, but it requires complex implementation and large bandwidth for transmission (Al Saad and Hato 2014).

The most efficient encryption techniques for dealing with redundant, and bulky audio data are the chaos-based and double random phase encoding (DRPE) techniques as they provide highly secure, and fast encryption (Maysaa and Qays 2013; Del Re et al. 1989).

This paper presents an efficient method for encrypting digital audio signals using a hybrid mixture of chaotic maps and optical encryption. The proposed audio cryptosystem aims to enhance the audio security during transmission on shared networks. The proposed audio cryptosystem is implemented using either Arnold's cat map with DRPE, or Baker map with DRPE. From their names, it is clear that each of them is composed of a chaotic cryptosystem and DRPE as an optical cryptosystem which means that they have two security layers. The first layer is

achieved by the chaotic map, while the second layer is achieved by DRPE which adds a physical security layer. This physical security layer enforces the security level of the proposed audio cryptosystem. The performance of the proposed audio cryptosystems is investigated using different encryption/decryption audio quality metrics including Spectral Distortion (SD), Log Likelihood Ratio (LLR), correlation, processing time, histogram and spectrogram. The results show that the proposed audio cryptosystem provides a high security level with high confidentiality. Also, the results show that the hybrid Baker map with DRPE cryptosystem outperforms the Arnold's map with DRPE system for different audio quality metrics.

The rest of the paper is organized as follows: Sect. 2 presents a general discussion of the chaotic cryptosystems embedded in the proposed systems. In Sect. 3, the details of DRPE are discussed. Section 4 presents the proposed audio cryptosystem. In Sect. 5, the key performance indicators used to evaluate the proposed audio cryptosystems are given. Section 6 presents the experimental results and discussion. Finally, Sect. 7 summarizes the main conclusions of the paper.

## 2 Chaotic system

Any dynamic and nonlinear deterministic system that exhibits pseudorandom behavior is a chaotic system. The output of chaotic systems depends on initial conditions, and specific parameters. That is, different initial conditions, and/or different parameter values yield different system's output (Del Re et al. 1989). Chaotic systems are motivated for applications in cryptography, and pseudo-random number generators because of their random-like behavior and sensitivity to initial conditions and parameter settings (Chen et al. 2014; Ahmad et al. 2012). Because of these unique characteristics, chaotic systems are able to fulfill the cryptographic properties such as confusion, diffusion and disorder. Chaotic systems are very sensitive to their inputs that is any simple change in the initial conditions and parameters setting can lead to a very big difference in the final system state over few iterations. Many researchers utilized chaotic systems in developing cryptographic algorithms such as (Al Saad and Hato 2014; Li et al. 2009; Kohad et al. 2012; Sharma 2012; Zeng et al. 2012; Sheu 2011; Elshamy et al. 2013; Zhao et al. 2014; Mermoul and Belouchrani 2010; Maysaa and Qays 2013; Del Re et al. 1989).

### 2.1 Chaotic Arnold's cat map

In mathematics, a chaotic function or map is any function that possesses some kind of chaotic behavior (Del Re et al. 1989). Arnold's cat map (ACM) or Arnold transform (AT),

proposed by Vladimir Arnold in 1960, is an example of a two-dimensional chaotic map (Zhang et al. 2008). If ACM is applied to a digital image, it randomizes the original organization of its pixels and the image becomes insignificant or noisy. However, it has a period  $p$  and if iterated  $p$  times, the original image reappears.

**Definition** The generalized schema of Arnold’s cat map can be given by the following transformation

$$\Gamma : T^2 \rightarrow T^2$$

such that:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \pmod{N} \tag{1}$$

where,  $x, y \in \{0, 1, 2 \dots N - 1\}$  and  $N$  is the size of a digital image. A new image is produced when all points in a given image are manipulated once by Eq. (1). ACM is easy and at the same time it is a strong transform (Lin and Chang 2001). Digital image encryption can be done by applying ACM in the following manner: Let  $p$  be the transform period of an  $N \times N$  digital image  $I$ . Placing ACM for a random iteration of times ( $t \in [1, p]$ ) to  $I$ , a scrambled image  $I'$  is acquired which is completely chaotic and is different from  $I$ . Now  $I'$  can be transmitted over the communication channels without revealing any data to the unauthorized receivers or sniffers (Petitcolas et al. 1999; Langelaar et al. 2000). The receiver repeats the operation for  $(p - t)$  times to obtain back the original image. The proposed scheme utilizes the ACM for encrypting audio. It first reshapes the audio from 1-D format to 2-D format, and then applies the same ACM procedure on the resulting 2-D matrix.

The original Arnold transformations given by Eq. (1) can be easily modified to produce a sequence of Arnold transformations as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} i & i + 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \pmod{N} \tag{2}$$

or

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} i + 1 & i \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \pmod{N} \tag{3}$$

where  $i \in \{1, 2, 3, \dots\}$

Transformations given by Eqs. (2) and (3) are periodic since the absolute value of  $\det(A)$  is always 1 in both cases, where  $A = [a, b; c, d]$  is the Arnold transform matrix. Equations (2) and (3) make a sequence of various Arnold transforms (Petitcolas et al. 1999; Langelaar et al. 2000) with various periodicity values  $P_k$  (Sang et al. 2003).

### 2.2 Chaotic Baker Map

In this section, the chaotic Baker map is explained (Pande and Zambreno 2011; Fridrich 1998). There are two types of the chaotic Baker map method, generalized map and discretized map. The discretized Baker map is an effective way to randomize the elements in a square matrix. 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 elements in one row, the secret key is chosen such that each integer  $n_i$  divides  $N$ , and  $n_1 + \dots + n_k = N$ .

Let  $N_i = n_1 + \dots + n_{i-1}$ . The data item at the indices  $(q, z)$ , is moved to the indices:

$$B_{(n_1, \dots, n_k)}(q, z) = \left( \frac{N}{n_i}(q - N_i) + z \pmod{\left(\frac{N}{n_i}\right)}, \frac{n_i}{N} \left( z - z \pmod{\left(\frac{N}{n_i}\right)} \right) + N_i \right) \tag{4}$$

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

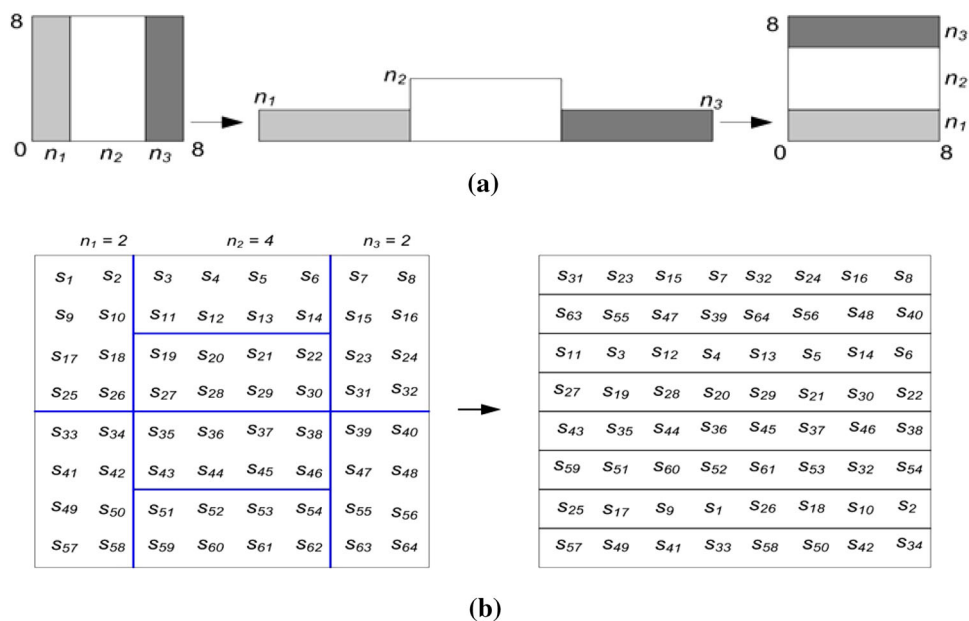
The following steps summarize the execution procedure of chaotic permutation:

- (1) An  $N \times N$  square matrix is divided into  $N$  rectangles of width  $n_i$  and number of elements  $N$ .
- (2) The elements in each rectangle are reshaped to a row in the permuted rectangle. Rectangles are taken from left to right beginning with upper rectangles then lower ones. Inside each rectangle, the scan begins from the bottom left corner towards upper elements. Figure 1 shows an example of the permutation of an  $8 \times 8$  matrix. The secret key is chosen to be (2, 4, 2), hence  $N = 8$ ,  $n_1 = 2$ ,  $n_2 = 4$ , and  $n_3 = 2$ . Figure 1a shows the generalized Baker map and Fig. 1b shows the discretized Baker map.

### 3 Double random phase encoding (DRPE)

In literature, many optical image encryption techniques have been proposed. The most widely applied and effective encryption technique among them is the DRPE which is presented by Refregier and Javidi (1995). DRPE is selected among the optical family of security algorithms to be used with a chaotic system in the proposed hybrid audio cryptosystems for many reasons. First one is that DRPE uses two keys in addition to optical transformation which increases the security level. Second one is that encryption/decryption time is very small compared with the other optical based encryption algorithms. Thirdly, it does not need any special requirements, and its implementation is very simple. Fourthly, it is applicable to 2-D, and 3-D formats. Finally, it could be optically and digitally applied.

**Fig. 1** Baker map.  
**a** Generalized baker map,  
**b** discretized baker map



**Table 1** Comparison between different optical cryptosystems

Optical encryption techniques	Keys of encryption	Encryption time	Implementation	Dimensions	Applied by
Double random phase encoding	Two	Very small (< 0.3 s)	Very simple (no requirements)	2D and 3D	Optical or digital
Holographic memory	one	Very high	More complex	3D	Optical or digital
Digital holography	one	High	Normal	3D	Digital
Polarization encryption	one	Normal	Normal	2D	Optical
Optical ID tags	one	High	More complex	2D	Optical

Table 1 presents a comparison between some optical encryption techniques using different parameters.

The DRPE is based on modifying the spectral distribution of the reformatted 2-D audio. The main idea for DRPE works by inserting two random phase masks (RPMs), one in the input plane and the other in the Fourier plane to encrypt the transformed 2-D audio into stationary noise in a setup called “4f”. The 4f setup is an optical system consisting of two cascaded lenses, separated by two focal lengths as shown in Fig. 2, with each of the input and output image planes having one focal length outside the lens system from different directions (i.e., so the total length is four focal lengths, hence ‘4f’).

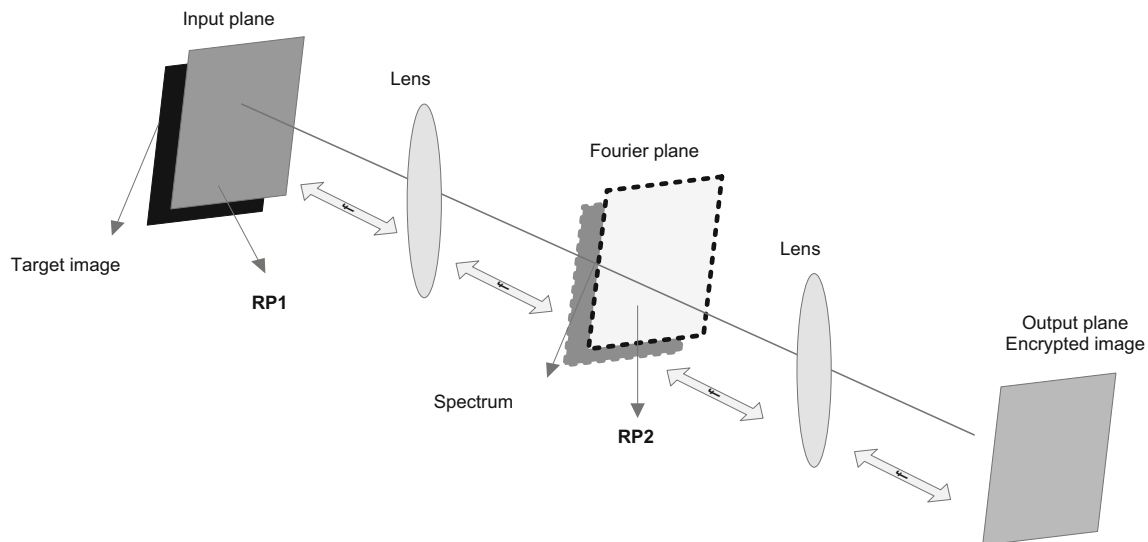
The decryption process must use the same RPM used in the encryption process. When applied in a 4f optical processor, the complex conjugate Fourier phase key is required to decrypt the encrypted 2-D audio in DRPE.

The normal DRPE is divided into three stages:

- (1) RPM1, the first key, is multiplied by the transformed 2-D audio to be encrypted. This procedure introduces the first modification for the spectrum of the target encrypted 2-D audio.

- (2) RPM2, the second key, is directly multiplied into the spectrum of transformed 2-D audio in the Fourier plane. Multiplication of the RPM2 by the spectrum obtained in the first stage will be the second modification in the spectrum of the target transformed 2-D audio.
- (3) The second optical inverse Fourier transform is carried out through a second lens to obtain an encrypted 2-D audio in the original audio 2-D space.

To explain the DRPE in detail, we consider a primary intensity 2-D transformed audio  $f(x, y)$  with positive values, where  $x$  and  $y$  denote the time domain coordinates. Also,  $v$  and  $\eta$  denote the Fourier domain coordinates. Let  $\psi(x, y)$  represent the encrypted 2-D audio, and  $n(x, y)$  and  $m(x, y)$ , represent two key independent white sequences uniformly distributed in  $[0, 2\pi]$ . To encode primary transformed 2-D audio  $f(x, y)$  into white stationary sequences, the two RPMs used are  $\varphi_n(x, y) = \exp [2i\pi n(x, y)]$  and  $\varphi_m(x, y) = \exp [2i\pi m(x, y)]$ .  $h(x, y) = m(x, y)$  is a phase function uniformly distributed in  $[0, 2\pi]$ . The RPM2 key  $\varphi_m(v, \eta)$ , is the Fourier transform for the function  $h(x, y)$ , that is,



**Fig. 2** Encryption process by double random phase encoding (DRPE)

$$FT\{h(x, y)\} = \hat{h}(v, \eta) = \varphi_m(v, \eta) = \exp[2i\pi m(v, \eta)] \quad (5)$$

The encryption process is composed of multiplying the primary transformed 2-D audio  $f(x, y)$  by the RPM1  $\varphi_n(x, y)$ . Encrypting 2-D audio is complex, with amplitude and phase, and is given by the following expression:

$$\psi(x, y) = \{f(x, y)\varphi_n(x, y)\} * FT^{-1}\{\varphi_m(v, \eta)\} \quad (6)$$

The symbol (\*) denotes convolution. The encrypted audio by Eq. (6) has a noise-like appearance that does not reveal any details of the original 2-D audio.

The decryption result output is

$$FT^{-1}\{FT[\psi(x, y)]\varphi_m^*(v, \eta)\} = FT^{-1}\{FT[f(x, y)\varphi_n(x, y)]\varphi_m(v, \eta)\varphi_m^*(v, \eta)\} = f(x, y)\varphi_n(x, y) \quad (7)$$

Absolute value for Eq. (7) turns out the decrypted 2-D Audio  $f(x, y)$ . The whole encryption and decryption mechanism can be implemented either digitally or optically. Optical hardware is very simple, it can be the classical 4f-processor (Goodman 1996; Javidi et al. 1997). Also in the encryption mechanism, the 4f-processor has the RPM1 as a first key stuck to the original transformed 2-D Audio in the time plane and the RPM2 as a second key in the Fourier plane.

The encrypted 2-D audio is given by Eq. (8):

$$\psi_p(x, y) = \{\exp[i\pi f(x, y)]\varphi_n(x, y)\} * h(x, y) = \{\exp[i\pi f(x, y)]\varphi_n(x, y)\} * FT^{-1}\{\varphi_m(v, \eta)\} \quad (8)$$

The encrypted 2-D audio can be generated either optically or digitally similar to that described in Eq. (6). Also the same optical setup shown in Fig. 2 is used for the decryption process, but in this process, the complex

conjugate of both RPMs  $\varphi_n^*(x, y) = \exp[-2i\pi n(x, y)]$  and  $\varphi_m^*(v, \eta) = \exp[-2i\pi m(v, \eta)]$ , referred to as first and second keys, are very important for decryption to retrieve the original 2-D audio. The Fourier transform second key  $\varphi_m^*(v, \eta)$  is placed in the Fourier plane, whereas the phase function first key  $\varphi_n^*(x, y)$  is placed at the input plane of the optical processor. The phase only version of the original 2-D audio  $\exp[i\pi f(x, y)]$  is recovered in the time domain. Original 2-D audio  $f(x, y)$  can be displayed as an intensity distribution by extracting the phase of  $\exp[i\pi f(x, y)]$  and dividing it by  $\pi$ .

### 4 Proposed hybrid audio encryption technique

To meet the requirements of modern cryptography applications with high security levels, the proposed hybrid cryptosystem is composed of digital and optical encryption techniques. It combines DRPE as an optical encryption technique with Arnold’s Cat map or Chaotic Baker map as a digital encryption technique. Figure 3 illustrates how the proposed hybrid cryptosystem works.

As it is shown from Fig. 3a, the processing steps of the proposed encryption system can be summarized as follows:

1. Segment original audio signal into segments, and then reshape them into 2-D format.
2. Mask with Arnold cat map or Baker map.
3. Add mask to the transformed 2-audio.
4. Clipping:
  - a. A value of 2 is subtracted from all values exceeding 1 resulting in negative values to make all samples between  $-1$  and  $1$ .

5. Apply DRPE.
  - a. Generate first Fourier RPM key RPM1 and multiply it by the target audio signal to be encrypted.
  - b. Generate second Fourier RPM key RPM2, and insert it into the audio signal in the Fourier plane. The insertion of the RPM2 in the audio obtained in the first phase introduces the second amendment into the target audio signal.
  - c. Perform the second optical Fourier transform using a second lens to obtain the encoded audio in the original 2-D space of audio signal.
6. Reshape the 2-D format to 1-D format which represents the encrypted audio signal.
7. Synthesize segments.
  - a. Add a value of 2 to negative values  $< -1$  in the resulting encrypted 2-D audio.
3. Inverse clipping.
  - a. Add a value of 2 to negative values  $< -1$  in the resulting encrypted 2-D audio.
4. Masking with Arnold cat map or Baker map.
5. Subtract mask from the encrypted 2-D audio.
6. Reshape the 2-D audio to 1-D format which represents the original audio signal.
7. Synthesize segments and reconstruct audio signal.

Also, as it is shown from Fig. 3b, the operation of the proposed decryption system can be summarized as follows:

1. Segment encrypted audio signal into segments and then reshape them into 2-D format.
2. Apply DRPE.
  - a. Generate first Fourier RPM key RPM1 and multiply it by the target audio signal to be encrypted.
  - b. Generate second Fourier RPM key RPM2, and insert it into the audio signal in the Fourier plane. The insertion of the RPM2 in the audio obtained in

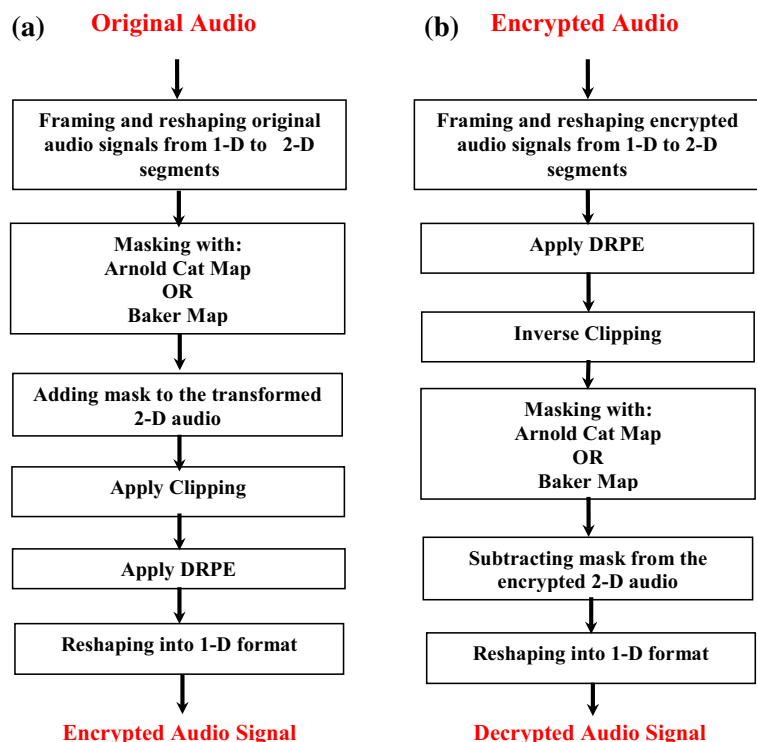
#### 4.1 Masking

The mask is created from the secret key. A specific number of ones is introduced to an all-zero block, then this block is permuted with the chaotic Baker map to create a mask of zeros and ones as shown in Fig. 4. The output mask is added to each block of the audio signal after the reshaping step. This step is necessary to hide silent periods within the audio signal to overcome known-plaintext attacks.

If, for example, the secret key of the Baker map is equal to  $\{4, 2, 2, 4\}$  then the sum of the sub-keys drives to  $12 \times 12$  blocks.

**Fig. 3** Flowchart of proposed audio cryptosystem.

**a** Encryption steps of proposed audio cryptosystem,  
**b** decryption steps of proposed audio cryptosystem



The number of sub-keys is 4, so the following steps can be executed:

- (a) First, fill the four rows with ones in a  $12 \times 12$  all-zero block based on the number of sub-keys that is equal to four.
- (b) Second, permute the block resulting from step (a) with the chaotic Baker map to spread ones in whole block.
- (c) Finally, add the output mask to each block and a clipping step is used in the output block. In the clipping step, a value of 2 is subtracted from all values exceeding 1 resulting in negative values to make all samples between  $-1$  and  $1$  as shown in Fig. 5.

### 5 Key performance metrics

Many metrics can be used for quality evaluation of audio cryptosystems. These metrics are classified into two categories; encryption quality metrics, and decryption quality metrics.

#### 5.1 Encryption quality metrics

Encryption quality metrics for audio cryptosystems have a great importance in the design of the encryption techniques. They are desirable for indicating the amount of distortion introduced by the audio cryptosystems, determining the parameter settings, and optimizing the audio cryptosystems structures. The more the amount of distortion, the better is the performance of the audio cryptosystem. Audio quality metrics are concerned with calculating how far the encrypted signal from the original signal is.

They are also concerned with determining the immunity of the encryption algorithm to cryptanalysis attacks.

A good encryption scheme should resist all types of known attacks. In this research paper, the security of the proposed audio encryption scheme is evaluated and compared with other audio encryption schemes using the following encryption quality metrics:

##### 5.1.1 Statistical analysis

Several kinds of ciphers might be solved by statistical analysis (Pascal 2005). Statistical analysis is applied on the proposed audio cryptosystem to demonstrate its confusion and diffusion properties. This is shown by a test on the histogram of the encrypted audio signal, the correlation between samples in the clear and encrypted signals, and the SD measures.

*Histogram* A histogram is a graphical display of the tabulated densities of data (Shumei et al. 2009). In this research, the histogram test is given to evaluate the success of the substitution step by indicating that new sample values are introduced into the encrypted signal instead of the original values.

*Correlation coefficient (CC)* A useful metric to determine the encryption quality of audio cryptosystem is the CC between similar samples in the clear and the encrypted signals. It can be computed as follows:

$$r_{xy} = \frac{c_v(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \tag{9}$$

where  $c_v(x,y)$  is the covariance between the original signal  $s$  and the encrypted signal  $y$ .  $D(x)$  and  $D(y)$  are the variances of the signals  $x$  and  $y$ , respectively. In numerical calculations, the following discrete formulas can be applied (Manjunath and Anand 2002):

**Fig. 4** Generation of the Mask.  
**a** Fill first four rows with ones,  
**b** permute the ones with the chaotic Baker map

1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	1	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

**Fig. 5** Steps for addition of the mask. **a** Original signal, **b** mask, **c** after addition of mask, **d** final block

0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.2	0	0	0

**(a)**

1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

**(b)**

1.1	1	1	1.5	1.4	1.4	1.4	1.4	0.8	1	1	1
0.1	0	2	0.5	0.4	1.4	0.4	0.4	0.7	0	0	1
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
1.1	1	1	1.5	1.4	1.4	1.4	1.4	0.8	1	1	1
0.1	0	2	0.5	0.4	1.4	0.4	0.4	0.7	0	0	1
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
1.1	1	1	1.5	1.4	1.4	1.4	1.4	0.8	1	1	1
0.1	0	2	0.5	0.4	1.4	0.4	0.4	0.7	0	0	1
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0

**(c)**

-0.9	1	1	-0.5	-0.6	-0.6	-0.6	-0.6	0.8	1	1	1
0.1	0	0	0.5	0.4	-0.6	0.4	0.4	0.7	0	0	1
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
-0.9	1	1	-0.5	-0.6	-0.6	-0.6	-0.6	0.8	1	1	1
0.1	0	0	0.5	0.4	-0.6	0.4	0.4	0.7	0	0	1
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
-0.9	1	1	-0.5	-0.6	-0.6	-0.6	-0.6	0.8	1	1	1
0.1	0	0	0.5	0.4	-0.6	0.4	0.4	0.7	0	0	1
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0
0.1	0	1	0.5	0.4	0.4	0.4	0.4	-0.3	0	0	0

**(d)**



$$E(x) = \frac{1}{N_x} \sum_{n=1}^{N_x} x(n) \tag{10}$$

$$D(x) = \frac{1}{N_x} \sum_{n=1}^{N_x} (x(n) - E(x))^2 \tag{11}$$

$$c_v(x, y) = \frac{1}{N_x} \sum_{n=1}^{N_x} (x(n) - E(x))(y(n) - E(y)) \tag{12}$$

where  $N_x$  is the number of audio samples involved in the computations. The low value of the correlation coefficient  $r_{xy}$  indicates a good encryption quality.

**Spectral distortion (SD)** The SD is a form of encryption metrics that is applied in frequency domain on the frequency spectra of the original and processed audio signals. It is computed in dB to give how far is the spectrum of the processed audio signal from that of the original audio signal. The SD can be computed as follows (Hedelin et al. 1999):

$$SD = \frac{1}{M} \sum_{m=0}^{M-1} \sum_{n=L_s m}^{L_s m + L_s - 1} |V_s(k) - V_y(k)| \tag{13}$$

where  $V_s(k)$  is the spectrum of the original audio signal in dB for a certain segment,  $V_y(k)$  is the spectrum of the processed audio signal in dB for the same segment,  $M$  is the number of segments and  $L_s$  is the segment length. SD is used as a quality metric for both encryption, and decryption. For encryption, a high value of SD between the original signals and encrypted signals indicates a good quality. In contrast, for decryption, a low value of SD between the original signals and decrypted signals indicates a good quality.

### 5.1.2 Processing time (PT)

The processing time is the time needed to encrypt/decrypt signal. The lower the processing time, the higher the encryption speed. Only, the decryption time of the proposed audio cryptosystem is estimated as both the encryption and decryption processes have approximately the same time.

## 5.2 Decryption quality metrics

Decryption quality metrics have a great importance in the design and maintenance of audio cryptosystems. They are desirable to indicate the amount of distortion introduced by any audio cryptosystem for determining the parameter settings, and optimizing the encryption algorithm. These metrics determine the immunity of the audio cryptosystem to distortion and attacks. They are performed on the

decrypted signal to verify its quality. There are two approaches that are used to determine the quality of decrypted audio signals; subjective and objective (Yang et al. 1998). Subjective metrics determine the quality depending on the perceptual ratings by a group of listeners. Objective metrics determine the quality using the physical parameters and computational models, and they are less expensive. They save time and give more consistent results. Therefore, objective audio metrics are desirable in practical applications. Current objective audio quality metrics base their estimates on the use of both original and decrypted audio signals. It is useful to measure the quality of the decrypted audio signal compared to the original audio signal to show the effect of the audio cryptosystem.

### 5.2.1 Log likelihood ratio (LLR)

The LLR metric for audio signal depends on the assumption that each segment can be represented by an all-pole linear predictive coding model of the form (Sang et al. 2003; Kwon et al. 2006):

$$s(n) = \sum_{m=1}^{m_p} a_m s(n - m) + G_s u(n) \tag{14}$$

where  $a_m$  (for  $m = 1, 2, \dots, m_p$ ) are the coefficients of the all-pole filter,  $G_s$  is the gain of the filter and  $u(n)$  is an appropriate excitation source for the filter. The audio signal is windowed to form frames of 15–30 ms length. The LLR metric is then defined as (Sang et al. 2003):

$$LLR = \left| \log \left( \frac{\vec{a}_s \mathbf{R}_y \vec{a}_s^T}{\vec{a}_y \mathbf{R}_y \vec{a}_y^T} \right) \right| \tag{15}$$

where  $\vec{a}_s$  is the LPCs coefficient vector  $[1, a_s(1), a_s(2), \dots, a_s(m_p)]$  for the original clear audio signal,  $\vec{a}_y$  is the LPCs coefficient vector  $[1, a_y(1), a_y(2), \dots, a_y(m_p)]$  for the decrypted audio signal, and  $\mathbf{R}_y$  is the autocorrelation matrix of the decrypted audio signal. The closer the LLR to zero, the higher is the quality of the output audio signal.

## 6 Experimental results and discussions

During a course of experiments, the two proposed hybrid audio cryptosystems namely; Arnold cat map with DRPE, and Baker map with DRPE are implemented, and their performance is evaluated and compared with each other.

### 6.1 Experiment 1

In this experiment, using CC, SD, LLR, and PT as key performance indicators, the performance of the two

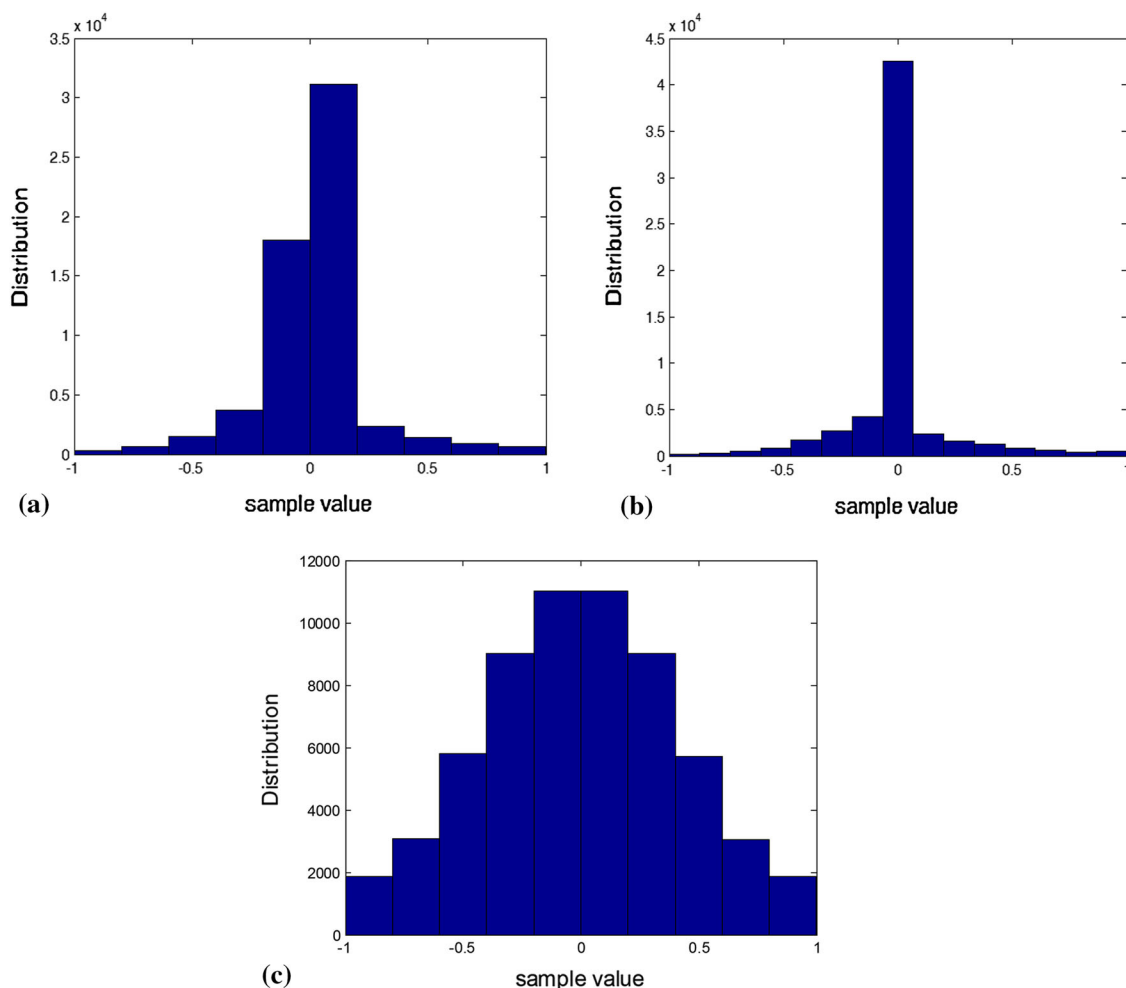
proposed audio cryptosystems is evaluated and compared. The results are tabulated in Table 2.

From that table; one can easily notice that the Hybrid Baker map with DRPE audio cryptosystem outperforms the Arnold map with DRPE. As in the encryption phase, it provides a lower value for CC, and a higher value for SD between the original and encrypted audio signals than the Arnold map with DRPE which indicates a better encryp-

tion quality. Also, in the decryption phase, it provides lower values for LLR, SD, and PT for the decrypted signals than the Arnold map with DRPE. This means that the quality of encrypted signal resulting from the hybrid Baker map with DRPE audio cryptosystem is better than that obtained by the hybrid Arnold map with DRPE audio cryptosystem.

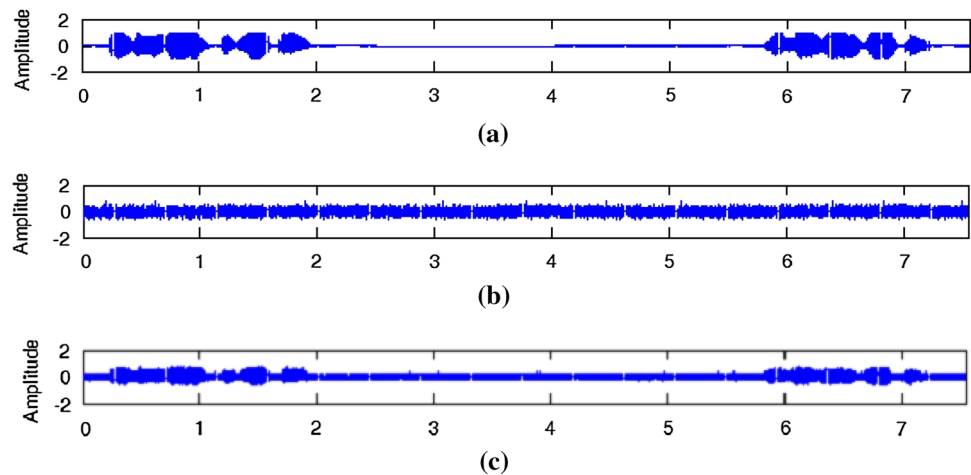
**Table 2** Quality metrics values for hybrid arnold cat map with DRPE and baker map with DRPE audio cryptosystems

		Correlation coefficient (CC)	Spectral distortion (SD)	Likelihood ratio (LLR)	Processing time (PT) in Sec
Arnold and DRPE	Encrypted	0.0051	20.36547	–	1.4343
	Decrypted	–	6.0072e – 003	1.1644e-006	
Baker and DRPE	Encrypted	0.0049	21.9829	–	0.3343
	Decrypted	–	5.8263e – 003	9.1354e-007	



**Fig. 6** Histograms of the audio signal. **a** Histogram of original audio, **b** histogram of encrypted audio signals using chaotic Arnold map with DRPE encryption, **c** histogram of encrypted audio signal using chaotic Baker map with DRPE encryption

**Fig. 7** Waveform of original, encrypted, and decrypted audio signal. **a** Waveform of the original audio signal, **b** waveform of the encrypted audio signal, **c** waveform of the decrypted audio signal



## 6.2 Experiment 2

In this experiment, the histogram check is used to evaluate the success of the substitution step by indicating the new sample values into the encrypted signal instead of the original values. Figure 6a presents the histogram of the original audio signal, while Fig. 6b presents the histogram of the encrypted audio signal using chaotic Arnold map with DRPE audio cryptosystem, and Fig. 6c presents the histogram of the of encrypted audio signal chaotic Baker map with DRPE audio cryptosystem.

From the Fig. 6b, it is observed that the histogram of the encrypted audio signal using Arnold map with DRPE audio cryptosystem is completely different from that of the original audio signal presented in the histogram at Fig. 6a which means that the security level of the first proposed audio cryptosystem (Arnold map with DRPE) is high. Also, from Fig. 6c, it is easy to notice that the histogram of the encrypted audio signal using Baker map with DRPE audio cryptosystem is fairly uniform and is significantly different from that of the original audio signal presented in the histogram at Fig. 6a which means that the security level of the second proposed audio cryptosystem (Baker map with DRPE) is also high.

## 6.3 Experiment 3

A waveform is an image that clarify an audio signal. It shows the changes in amplitude over a certain amount of time. In Fig. 7a, b, and c, the waveforms of the original audio signal, encrypted audio signal using Baker map with DRPE audio cryptosystem, and decrypted audio signal are presented. As it could be seen from that figure, the waveform of encrypted audio signal is completely different from that of the original audio signal which means that the encryption quality of proposed hybrid chaotic optical

cryptosystems is high. Also, by comparing the waveforms presented at Fig. 7a, and c one can notice that they are very close to each other which indicates the decryption quality of the proposed audio cryptosystems is high.

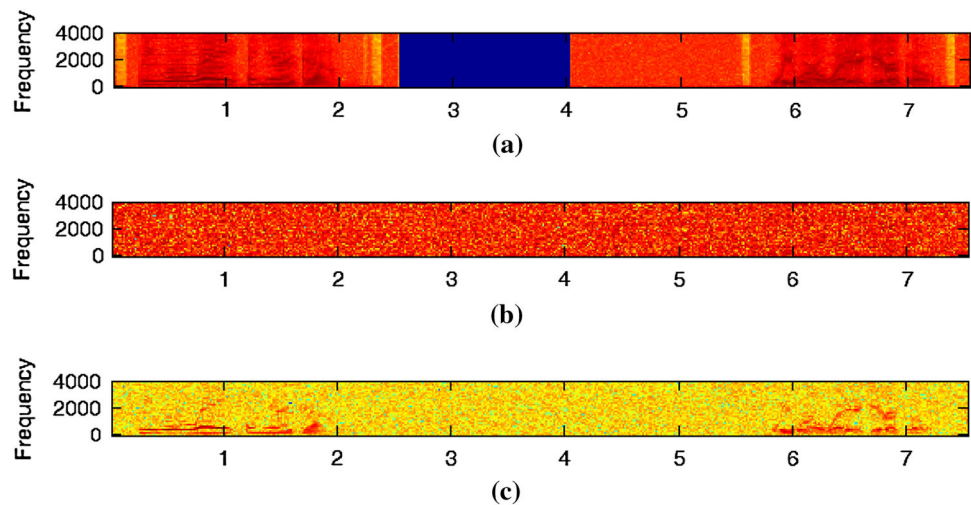
## 6.4 Experiment 4

A spectrogram is a visual clarification of the spectrum of frequencies in an audio signal as they change with time or some other variable. In Fig. 8a, b, and c, the spectrograms of the original audio signal, encrypted audio signal using Baker map with DRPE audio cryptosystem, and decrypted audio signal are presented. By comparing the Fig. 8a, and b, one can easily notice that spectrogram of encrypted signals is completely different from that of the original signal. This implies higher encryption quality. Also, by comparing the Fig. 8a, and c, it is easy to notice that they are very close to each other which implies higher decryption quality.

## 7 Conclusion

Audio security is concerned with insuring the secrecy, reliability, accessibility and confidentiality of data. The main target of voice security is to protect audio systems from unauthorized access, disruption, alteration, annihilation and use. This paper proposed two efficient hybrid chaotic optical audio cryptosystems. They are Arnold cat map with DRPE, and Baker map with DRPE. These proposed cryptosystems are multilayer security systems. They utilize the chaotic systems to provide the first security layer. In addition to this security layer, the proposed audio cryptosystems utilize the DRPE for providing the second security layer. This layer represents a physical security layer one that it is immune to attacks. The two proposed

**Fig. 8** Spectrograms of the original, encrypted and decrypted signal. **a** Spectrogram of the original audio signal. **b** spectrogram of the encrypted audio signal, **c** spectrogram of the decrypted audio signal



hybrid chaotic optical audio cryptosystems are implemented and their performances are evaluated and compared with each other using different performance evaluation metrics. The results showed that the two proposed audio cryptosystems have high security degree with high confidentiality. Also, the results showed that the hybrid Baker map with DRPE audio cryptosystem outperforms the Arnold map with DRPE using various encryption/decryption audio quality metrics.

## References

- Ahmad, M., Alam, B., & Farooq, O. (2012). Chaos based mixed keystream generation for voice data encryption. *International Journal on Cryptography and Information Security (IJCIS)*, 2(1), 39–48.
- Al Saad, S.N., Hato, E. (2014). A speech encryption based on chaotic maps. In *Int. J. of Computer Applications* (Vol. 93, pp. 19–28).
- Barni, M., Bartolini, F., & Piva, A. (2001). Improved wavelet based watermarking through pixel-wise masking. *IEEE Transactions on Image Processing*, 10, 783–791.
- Chen, L. H., & Lin, J. J. (2003). Mean quantization based image watermarking. *Image and Vision Computing*, 21, 717–727.
- Chen, J.-X., Zhu, Z.-L., Fu, C., Zhang, L.-B., & Zhang, Y. (2014). Cryptanalysis and improvement of an optical image encryption scheme using a chaotic Baker map and double random phase encoding. *Journal of Optics*, 16, 125403.
- Del Re, E., Fantacci, R., & Maffucci, D. (1989). A new speech signal scrambling method for secure communications: theory, implementation, and security evaluation. *IEEE Journal*, 7, 474–480.
- Elshamy, A. M., Rashed, A. N. Z., Mohamed, A. A., Faragalla, O. S., Mu, Y., Alshebeili, S. A., & El-Samie, F. A. (2013). Optical image encryption based on chaotic baker map and double random phase encoding. *Journal of Lightwave Technology*, 31(15), 2533–2539.
- Fridrich, J. (1998). Symmetric ciphers based on two-dimensional chaotic maps. *International Journal of Bifurcation and Chaos*, 8(6), 1259–1284.
- Goodman, J. W. (1996). *Introduction to Fourier optics* (2nd ed.). New York: McGraw-Hill.
- Hedelin, P., Norden, F., Skoglund, J. (1999). SD optimization of spectral coders. In *IEEE Workshop on Speech Coding Proc* (pp. 28–30).
- Javidi, B., Sergent, A., Zhang, G., & Guibert, L. (1997). Fault tolerance properties of a double phase encoding encryption technique. *Optical Engineering*, 36, 992–998.
- Kohad, H., Ingle, V.R., Gaikwad, M.A. (2012). An overview of speech encryption techniques. In *Int. J. of Engineering Research and Development* (Vol. 3, pp. 29–32).
- Kwon, J.S.K, Park, S., Sung, D.K. (2006) Collision mitigation by log-likelihood ratio (LLR) conversion in orthogonal code-hopping multiplexing. In *IEEE Transactions on Vehicular Technology* (Vol. 55, pp. 709–717).
- Langelaar, G. C., Setyawan, I., & Lagendijk, R. I. (2000). Watermarking digital image and video data. *IEEE Signal Processing Magazine*, 17(5), 20–46.
- Li, H., Qin, Z., Shao, L., Wang, B. (2009). A novel audio scrambling algorithm in variable dimension space. In *11th Int. Con. on Advanced Communication Technology ICACT 2009* (Vol. 03, pp. 1647–1651).
- Lin, C. Y., & Chang, S. F. (2001). A robust image authentication method distinguishing jpeg compression from malicious manipulation. *IEEE Transactions on Circuits and Systems of Video Technology*, 11, 153–168.
- Manjunath, G., & Anand, G. V. (2002) Speech encryption using circulant transformations, In *Proc. IEEE, Int. Conf. Multimedia and Exp* (Vol. 1, pp. 553–556).
- Maysaa, A. A. J., & Qays, I. (2013). Speech encryption using chaotic map and blow fish algorithms. *Journal of Basrah Researches*, 39(2), 68–76.
- Mermoul, A., Belouchrani, A. (2010). A Subspace-based method for speech encryption. In *Proc. of IEEE 10th Int. Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)* (pp. 351–541).
- Pande, A., & Zambreno, J. (2011). *A chaotic encryption scheme for real-time embedded systems: design and implementation*. New York: Springer.
- Pascal, J. (2005). Statistical cryptanalysis of block ciphers, *Ph. D. Thesis*.
- Petitcolas, F. A., Anderson, R., & Kuhn, M. (1999). Information hiding: A survey. *Proceedings of the IEEE*, 87, 1062–1078.
- Refregier, P., & Javidi, B. (1995). Optical image encryption based on input plane and Fourier plane random encoding. *Optics Letters*, 20, 767–769.

- Sang, W.K., Young, G.K., Simon, M.K. (2003). Generalized selection combining based on the log-likelihood ratio. In *IEEE Int. Con. on Communications ICC 2003* (Vol. 4, pp. 2789–2794).
- Sharma, D. (2012). Five level cryptography in speech processing using multi hash and repositioning of speech elements. In *Int. J. of Emerging Technology and Advanced Engineering* (Vol. 2).
- Sheu, L.-J. (2011). A speech encryption using fractional chaotic systems”. *Nonlinear Dynamics*, 65, 103–108.
- Shumei, W., Yong Fan, F., Ping, Y. (2009). A watermarking algorithm of gray image based on histogram. In *2nd International Congress on Image and Signal Processing CISP 2009* (pp. 1–5).
- Wu, Y., Ng, B.P. (2002). Speech scrambling with Hadamard transform in frequency domain. In *Proc. 6th Int. Conf. on Signal Processing* (Vol. 2, pp. 1560–1563).
- Yang, W., Benbouchta, M., Yantorno, R. (1998). Performance of the modified bark spectral distortion as an objective speech quality measure. In *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing* (Vol. 1, pp. 541–544), Washington, USA.
- Zeng, L., Zhang, X., Chen, L., Fan, Z., & Wang, Y. (2012). Scrambling-based speech encryption via compressed sensing. *EURASIP Journal on Advances in Signal Processing*, 257, 1–12.
- Zhai, Y., Lin, S., Zhang, Q. (2008). Improving image encryption using multi-chaotic map. In *Workshop on Power Electronics and Intelligent Transportation System, 2008. PEITS '08* (pp. 143–148), August 2–3, 2008.
- Zhang, C. et al. (2008). Digital image watermarking with double encryption by Arnold transform and logistic. In *4th Int. Con. on networked computing & advanced information management* (pp. 329–334).
- Zhao, H., He, S., Chen, Z., Zhang, X. (2014). Dual key speech encryption algorithm based underdetermined BSS. In *The Scientific World Journal* (pp. 1–7).