A novel approach based on stream cipher for selective speech encryption

Aissa Belmeguenai[1] · Zahir Ahmida[1] · Salim Ouchtati[1] · Rafik Djemii[1]

Received: 11 January 2017 / Accepted: 5 July 2017 / Published online: 13 July 2017

Abstract This paper proposes a new approach to the encryption of speech intended to be transferred on an insecure channel. The proposed technique is based on stream encryption where the original speech is pre-processed in order to select the relevant data for encryption by removing the unvoiced parts of the signal. In addition to reducing the size of the speech signal, the selection process also reduces the encryption and decryption times. The selected data is encrypted using a pseudo-random generator to make the information unable to understand by an intruder. The used pseudo-random generator is based on two 256-bits shift registers, one is linear and the other non-linear. This generator allows encryption in higher dimensional space which expands the key space and, therefore, improves security against brute force attack. Various analyzis such as histograms, key space, key sensitivity, correlation, brute force attack, signal to noise (SNR) and peak signal to noise ratio (PSNR) have been performed. The analysis shows that the proposed scheme is computationally simple, efficient with superior performance compared to other systems presented in the literature.

Keywords Stream ciphers · Selective speech encryption · Pseudorandom generator

1 Introduction

Nowadays, we witness an unprecedented development and use of information and technology systems coupled with an expansion of complex communication networks. These interconnected networks which vehicle huge and diverse amounts of information are also connected to internet. Consequently, the information which circulates freely in several forms: audio, images, textual and otheris potentially subject to interception by unauthorized persons or systems for whom it is not intended and can be, therefore, falsified, exchanged, stored, diverted, copied, etc. This situation can raise major security concerns when the information is confidential such as in military, government or corporate speech communication. It is, thus, necessary to ensure confidentiality and integrity of the informationin face of any undesirable interception by adversaries.

Confidentiality of speech communication can be achieved by making the information incomprehensible for all intruders through modern cipher of data which is an effective means to resolve this security problem. The contribution of this work is focused on two major problems: the first problem is closely related to the very nature of the type of encryption chosen, namely stream cipher. Stream cipher is one of the two types of secret key encryption which, unlike block cipher, consists of adding bit to bit to the clear text a sequence of bits of the same length as the text to be encrypted. The added bit sequence is called the keystream and, it also represents the secret key. This system ensures perfect security under the condition that the keystream is a completely random sequence of the same size as the text to be encrypted and that it can be exchanged in a safe manner between the sender and the receiver.

However, this kind of stream encryption method presents a great disadvantage because it is neither reasonable nor practical to share a secret key whose length equals that of the text to be transmitted. In the case of speech encryption, the secret key must have the same size as that of the speech. The second problem is specific to speech signal

✉ Aissa Belmeguenai
belmeguenaiaissa@yahoo.fr

[1] Universite du 20 aout 1955, Skikda, Algeria

encryption. In many cases, total encryption of a speech file is not necessary and the encryption effort can be targeted to those parts of the speech file containing useful and confidential information. Knowing that the time required for the encryption of a speech signal is directly proportional to its size (The larger the speech file, the longer the time required for its encryption), selective encryption of useful parts of the speech file can significantly reduce the encryption time and effort.

Several methods that deal with speech encryption are available in the literature Rahman et al. (2012), Musheer et al. (2012), Ashtiyani et al. (2012), Maysaa and Iman Qays (2013), Abd Elzaher et al. (2016), Farsana and Gopakumar (2016), Belmeguenai et al. (2015) and Belmeguenai et al. (2015). While some techniques use chaotic signals, others methods are based on registers Belmeguenai et al. (2015) and Belmeguenai et al. (2015). However, all of these contributions only address the first problem. In this work we propose a new approach by which both problems discussed above are addressed simultaneously. In fact, several schemes have been proposed to solve the problem related to the length of the key but these are rather complex and difficult to implement.

In this work the problem of the secret key is approached by using pseudo-random sequences generated in a deterministic way from a short secret key which can be exchanged more easily between sender and receiver. The mechanism for generating pseudo-random sequences relies on a set of linear feedback shift registers (LFSR) and nonlinear feedback shift registers (NLFSRs) combined with a Boolean function. On the other hand, the proposed method considerably optimizes the time required for encryption by focusing the encryption effort on a portion of the speech file which we call the useful information. This is achieved by removing silence/unvoiced parts of the speech based on some appropriate energy level threshold of the speech samples. The proposed method is simple, fast and easy to implement with the advantage of significant reduction of the computation time of the encryption compared to other stream cipher systems Gammel et al. (2006), Hell et al. (2006), Cid et al. (2009).

The paper is organized as follows. The Sect. 2 gives the notion of partial speech data encryption. In Sect. 3 the proposed pseudo-random generator criteria and choices are presented. Section 4 presents the proposed approach. The analysis and the results are given in Sect. 5. The concludes is presented in Sect. 6.

2 Partial speech data encryption

In partial encryption, the encryption process is not applied to the entire information set but, is limited to a selected subset of the original information. By limiting encryption to only a portion of the information, partial encryption allows significant reduction in required computing resources and encryption time. Selection of useful information is done manually or automatically depending on the application at hand Lamel et al. (1981) and Ramirez et al. (2007). This work adopts automatic detection and selection of voiced parts from silence/unvoiced parts based on the energy level of the speech samples. Very low energy samples are considered silence/unvoiced and thus removed. In order to achieve selection of useful speech information, the algorithm described in Burileanu et al. (2000) is implemented. Figure 1 shows autamatic selection of useful speech information.

3 Pseudo-random generator

The production of a pseudo-random generator (PRG), which is as resilient as possible to known attacks, requires an adequate mathematical base which makes it possible to generate robust and unpredictable pseudo-random sequences. An illustration of the different blocks used in the proposed pseudo random generator is depicted in Fig. 2.

The pseudo-random generator consists of one 256-bit linear Feedback Shift Register (LFSR), one 256-bit Non-Linear Feedback Shift Register (NLFSR) and a nonlinear Boolean function g. The content of both registers are denoted by $s_{i+1}, s_{i+2}, ..., s_{i+256}$ for the LSFR and $u_{i+1}, u_{i+2}, ..., u_{i+256}$ for NLSFR. The recurrence relationship for output of LFSR is given by:

$$\begin{aligned} s_{i+256} = s_{i+212} \oplus s_{i+194} \oplus s_{i+192} \oplus s_{i+187} \oplus \\ s_{i+163} \oplus s_{i+151} \oplus s_{i+125} \oplus s_{i+115} \oplus \\ s_{i+107} \oplus s_{i+85} \oplus s_{i+66} \oplus s_{i+64} \oplus \\ s_{i+52} \oplus s_{i+48} \oplus s_{i+14} \oplus s_{i}. \end{aligned} \tag{1}$$

The NLFSR uses the following recurrence relation:

$$\begin{aligned} u_{i+256} = u_i \oplus u_{i+2} \oplus u_{i+7} \oplus u_{i+13} \oplus \\ u_{i+14} \oplus u_{i+19} \oplus u_{i+27} \oplus u_{i+29} \oplus u_{i+34} \oplus \\ u_{i+12}u_{i+19} \oplus u_{i+16}u_{i+20} \oplus u_{i+17}u_{i+19} \oplus \\ u_{i+17}u_{i+29} \oplus u_{i+1}u_{i+12}u_{i+18} \oplus \\ u_{i+1}u_{i+12}u_{i+22} \oplus u_{i+12}u_{i+16}u_{i+20} \oplus \\ u_{i+12}u_{i+19}u_{i+21} \oplus u_{i+19}u_{i+21}u_{i+28} \oplus \\ u_{i+1}u_{i+12}u_{i+18}u_{i+21} \oplus u_{i+1}u_{i+12}u_{i+21}u_{i+22} \oplus \\ u_{i+1}u_{i+18}u_{i+21}u_{i+28} \oplus u_{i+1}u_{i+21}u_{i+22}u_{i+28} \oplus \\ u_{i+12}u_{i+16}u_{i+20}u_{i+21} \oplus u_{i+12}u_{i+18}u_{i+21}u_{i+22} \oplus \\ u_{i+12}u_{i+19}u_{i+20}u_{i+21} \oplus u_{i+16}u_{i+20}u_{i+21}u_{i+28} \oplus \\ u_{i+18}u_{i+21}u_{i+22}u_{i+28} \oplus u_{i+11}u_{i+19}u_{i+20}u_{i+28}. \end{aligned} \tag{2}$$

The 512 contents in the LFSR and NLFSR represent the state of the PRG. From this state, 14 variables are taken as

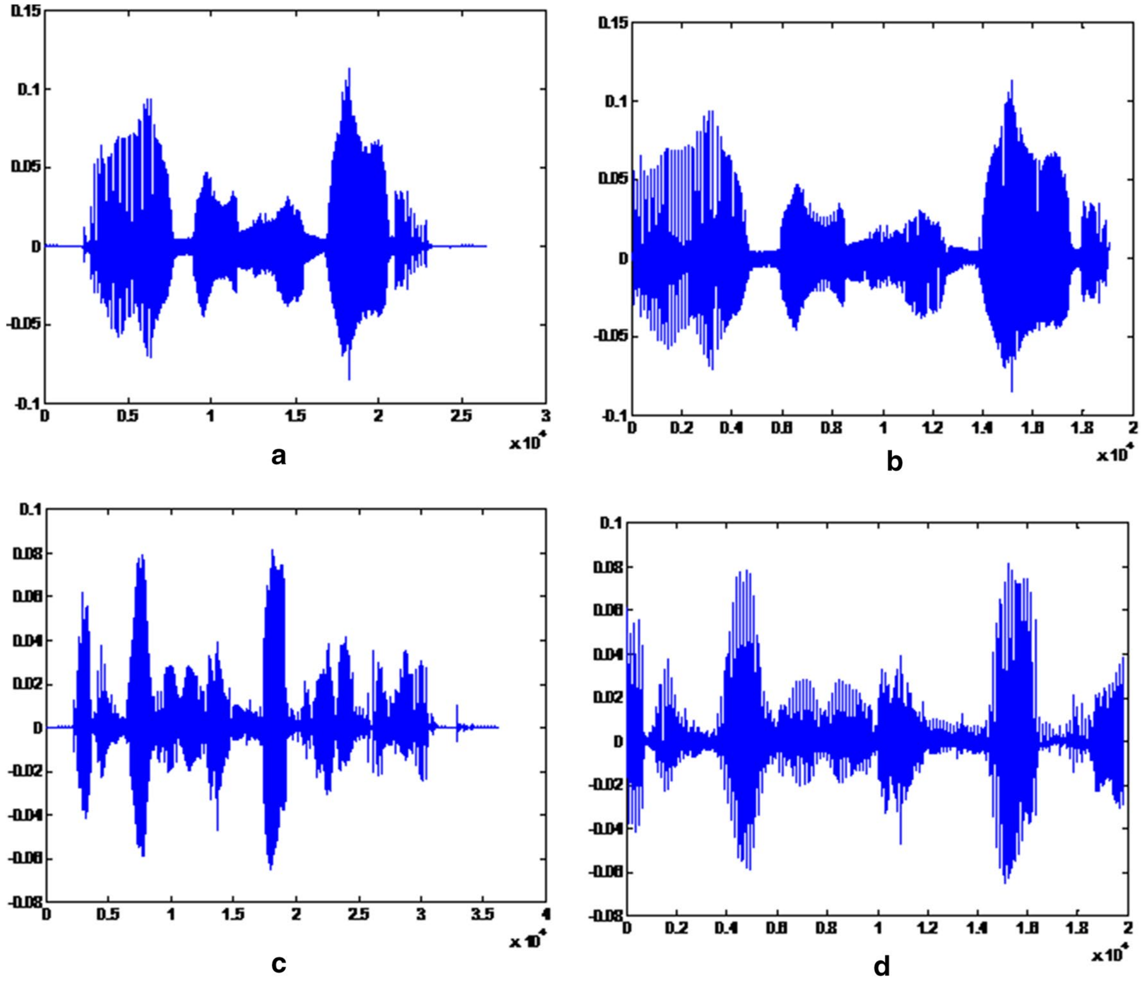

Fig. 1 The process of automatic selection of useful speech information: **a** original speech 1, **b** selected part to be encrypted, **c** original speech 2 and **d** selected part to be encrypted

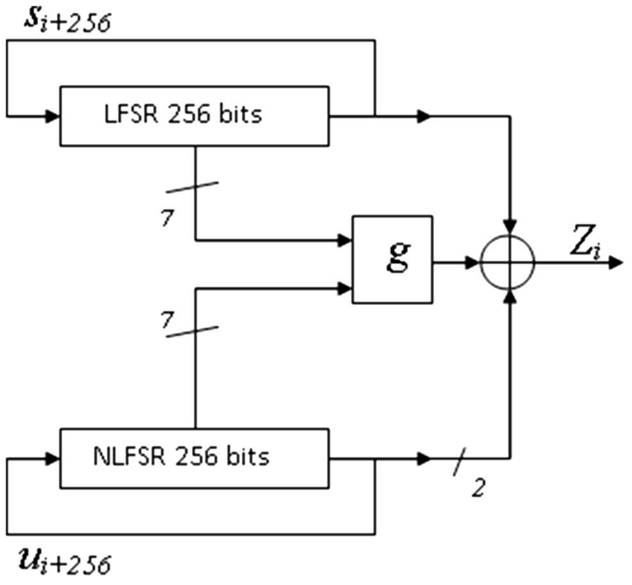

Fig. 2 Pseudo-random generator

input to a Boolean function, g. seven inputs to g are taken from the NLFSR and seven are taken from the LFSR. The Boolean function g is chosen to be balanced, correlation immune of the thirth order, has algebraic degree 4 and has nonlinearity 7936. Its output is defined as

$$\begin{aligned} g = u(i+2) \oplus u(i+15) \oplus u(i+45) \oplus s(i+95)u(i+64) \oplus \\ u(i+12)s(i+8) \oplus s(i+13)s(i+20) \oplus u(i+95)s(i+42) \oplus \\ u(i+36)u(i+64) \oplus s(i+60)s(i+79) \oplus \\ u(i+12)u(i+95)s(i+95)u(i+36) \oplus \\ u(i+12)u(i+95)u(i+64)(s(i+95) \oplus u(i+36) \oplus u(i+45)). \end{aligned} \tag{3}$$

The Keystream bits $Z(i)$ is produced by combining the two bits from the state of NLFSR and one bit from the state of LFSR with the output of a non-linear filtering function g, as:

$$Z(i) = u(i+73) \oplus u(i+89) \oplus s(i+93) \oplus g(i). \tag{4}$$

4 Proposed approach

The fundamental object of this contribution is to propose an approach that allows a sender and receiver to exchange speech signals through insecure channel in such a way that the exchanged speeches are totally unintelligible to any unauthorized third party hijacker. The main idea of this work is inspired from the references Kulkarni et al. (2008), Bhatnagar and onathan Wu (2012) and Belmeguenai et al. (2015). The block diagram of the overall proposed approach is shown in Fig. 3.

4.1 Input speech

The input speech signals used to validate the proposed approach are three files from the TIMIT database Garofolo (1993). The files denoted si1699, si2257 and si2329 are available in .sph file format. Before any processing is initiated, these speech files are to be converted into .wav file format using the 'readsph.m' Matlab function.

4.2 Silence removal and normalization

The first step in the encryption process, at the sender side, is the pre-processing of the original speech with the purpose of ridding the signal of its silent/unvoiced parts. The remaining voiced parts are concatenated to form the useful speech signal to be encrypted. This pre-processing step is implemented automatically using the silence removal algorithm presented in Burileanu et al. (2000). Figures 1b and d illustrate the results of useful speech data selection carried out on two different speech signals. The selected signal is normalized by applying the following relationship

$$X = 255 \times \frac{U - Min(U)}{Max(U) - Min(U)}. \tag{5}$$

where U is the useful speech signal, $Max(U)$ and $Min(U)$ are the maximum and minimum values of U, respectively, and X is the resulting normalized useful information to be encrypted. Subsequent to normalization, the useful data is converted into useful information digits using 8-bit quantization to obtain digital values in the range 0 to 255.

4.3 Encryption

Starting from an initial secret key k (initial values of LFSR and NLFSR), the sender generates a pseudo-random sequence (keystream) equal in size to the selected useful information digits to be encrypted. The encrypted useful information digits are obtained by XOR-ing each selected useful information digit with the keystream before being transferred to the receiver as shown in Fig. 3. The encryption process of the digital useful data is summarized in the following steps:

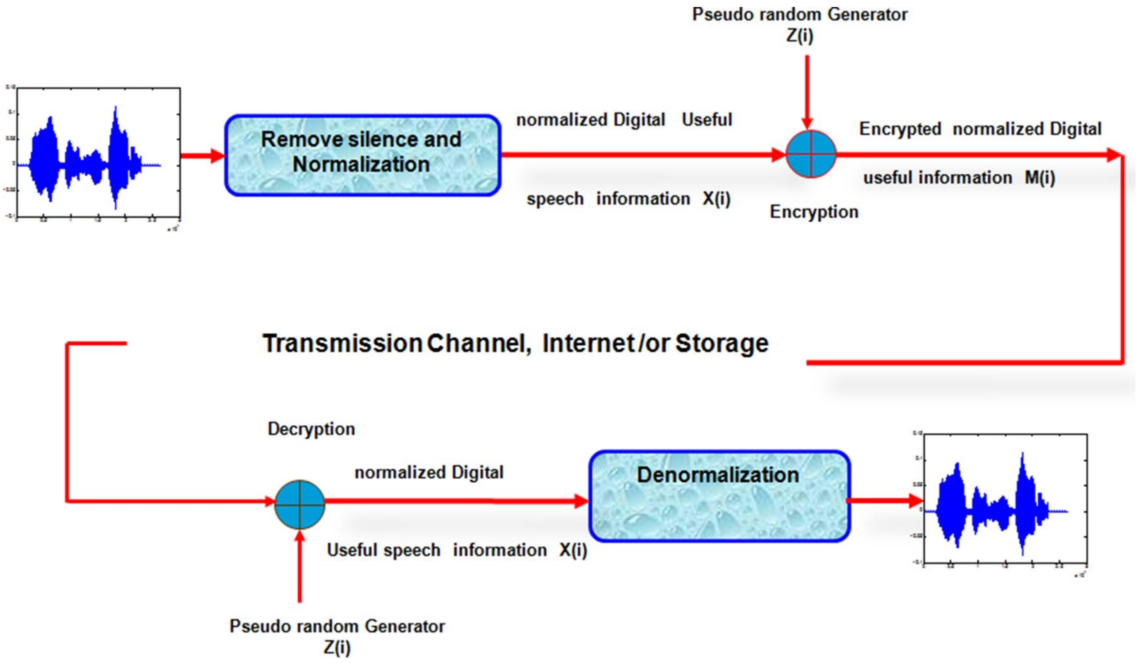
Pseudo random Generator
Z(i)
Remove silence and Normalization
normalized Digital Useful
speech information X(i)
Encryption
Encrypted normalized Digital
useful information M(i)
Transmission Channel, Internet/or Storage
Decryption
normalized Digital
Useful speech information X(i)
Denormalization
Pseudo random Generator
Z(i)

Fig. 3 Block diagram of the encryption-decryption process

- Generation of the keystream Z_i according to Eq. 4.
- XOR-ing the normalized digital information X_i with keystream Z_i to produce the normalized encrypted useful information digits M_i such that; $M_i = X_i \oplus Z_i$.
- Transmission of the normalized encrypted information Mi through the insecure communication channel.

4.4 Decryption

Having exchanged the secret key k with the sender, the receiver uses this key to restore of the normalized useful digital information by carring out the following the procedure:

- Reception of the normalized ecrypted digital information M_i.
- Generation of the keystream Z_i using the exchanged secret key k in Eq. 4.
- Compute the normalized decrypted digital information X_i by XOR-ing the normalized ecrypted digital information M_i with the keystream Z_i: $X_i = M_i \oplus Z_i$.

4.5 Denormalization

The final step in the ecryption-decryption process to ultimately recover the original selected speech signal is denormalization of the decrypted speech digits X_i according to the following relation:

$$U = \frac{Max(U) - Min(U)}{255} X + Min(U). \tag{6}$$

5 Analysis and discussion

In the sequel, the performance of the proposed scheme is demonstrated and analyzed through simulation experiments carried out on a MATLAB.7.5 platform. For this purpose, three speech files denoted si1699, si2257 and si2329 rare used as input speeches to the proposed encryption-decryption algorithm. Pre-processing and encryption results of these speech files sampled at a frequency of 16 Khz are shown in Figs. 4, 5 and 6. Figures 7, 8 and 9 illustrate the spectrums and histograms of encryption using Grain-128, Achterbahn-128 and Rakaposhi systems, respectively.

5.1 Histogram analysis

It is clear that the histogram of the encrypted speech illustrated in Fig. 10 is almost uniformly distributed, and different from the histogram of the original speech. Thus, the encrypted speech provides no index that would facilitate the use of a statistical attack on the proposed scheme, making statistical attacks difficult. These properties indicate that the proposed scheme provides high security against statistical attacks. Figure 10 shows the histograms of original speeches and their corresponding encrypted speeches.

5.2 Key space

To further strengthen security of the system, the secret key of the proposed scheme is constituted not only of the initialization of a single register, but of both registers (LFSR and NLFSR) concatenated as a single string of 512-bit length. This large string of bits can ensure maximum security, and can reasonably prevent any attempts of attack by adversaries. Furthermore, the larger the size of the secret key, the less effective are exhaustive attacks. In the proposed scheme, 2^{512} initializations are necessary for an exhaustive search which is large enough to prevent an exhaustive attack.

5.3 Key sensitivity

The most important feature of a cryptosystem based on stream cipher is key sensitivity. For example, a small change in the secret key leads to different results during the encryption and decryption of data. Correct decryption needs complete and precise knowledge of all of the secret key bits; otherwise the data cannot be decrypted. Figure 11 shows the decryption using a wrong key (one bit has been changed).

5.4 Correlation coefficient

The correlation coefficient analysis is a statistical measure to evaluate the mutual relationship between two variables. While a correlation coefficient that is equal to unity is an indication of a strong correlation between the original data and its corresponding encrypted data, a zero valued correlation coefficient points to a weak relationship between the signals. Correlation can be calculated as follows:

$$Cor(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y}. \tag{7}$$

where Cov is the covariance, σ_x and σ_y are the standard deviations of x any, respectively. Table 1 summarizes and compares correlation coefficients for the test speeches. It

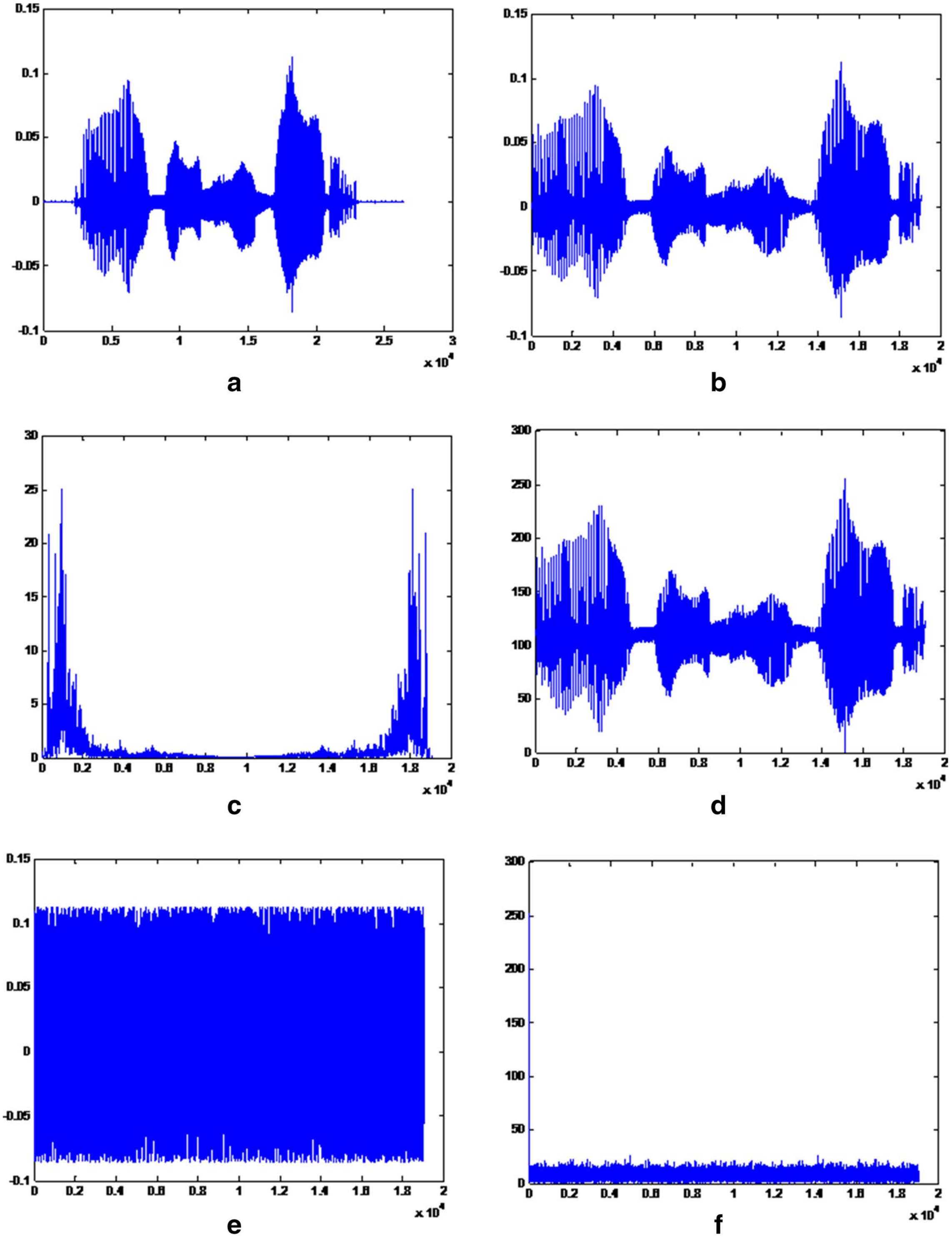

Fig. 4 Test for si1699: **a** original speech, **b** selected speech samples for encryption, **c** spectrum of selected speech, **d** normalized selected speech, **e** encrypted normalized selected speech, **f** spectrum of encrypted normalized selected speech

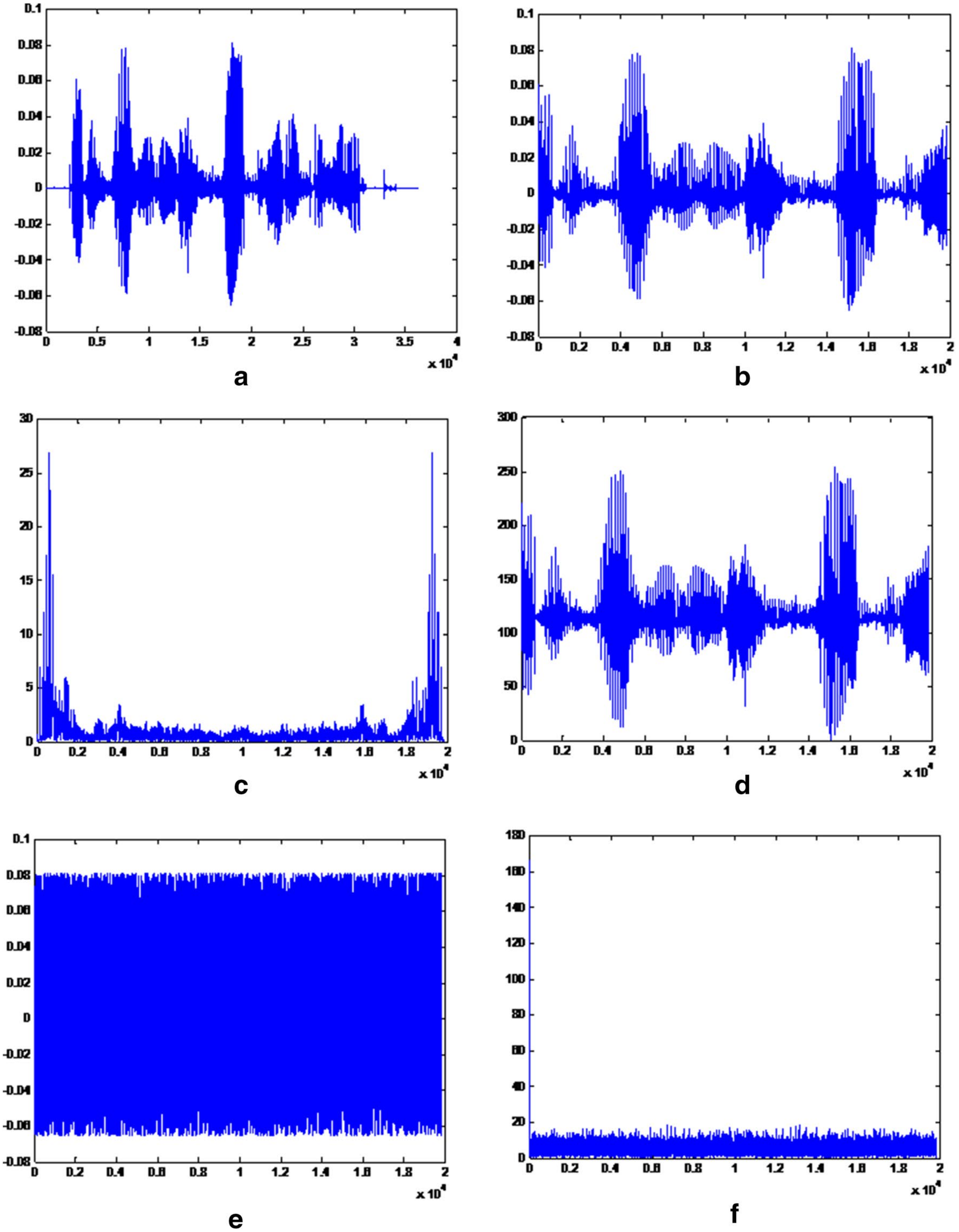

Fig. 5 Test for si2257: **a** original speech, **b** selected speech samples for encryption, **c** spectrum of selected speech, **d** normalized selected speech, **e** encrypted normalized selected speech, **f** spectrum of encrypted normalized selected speech

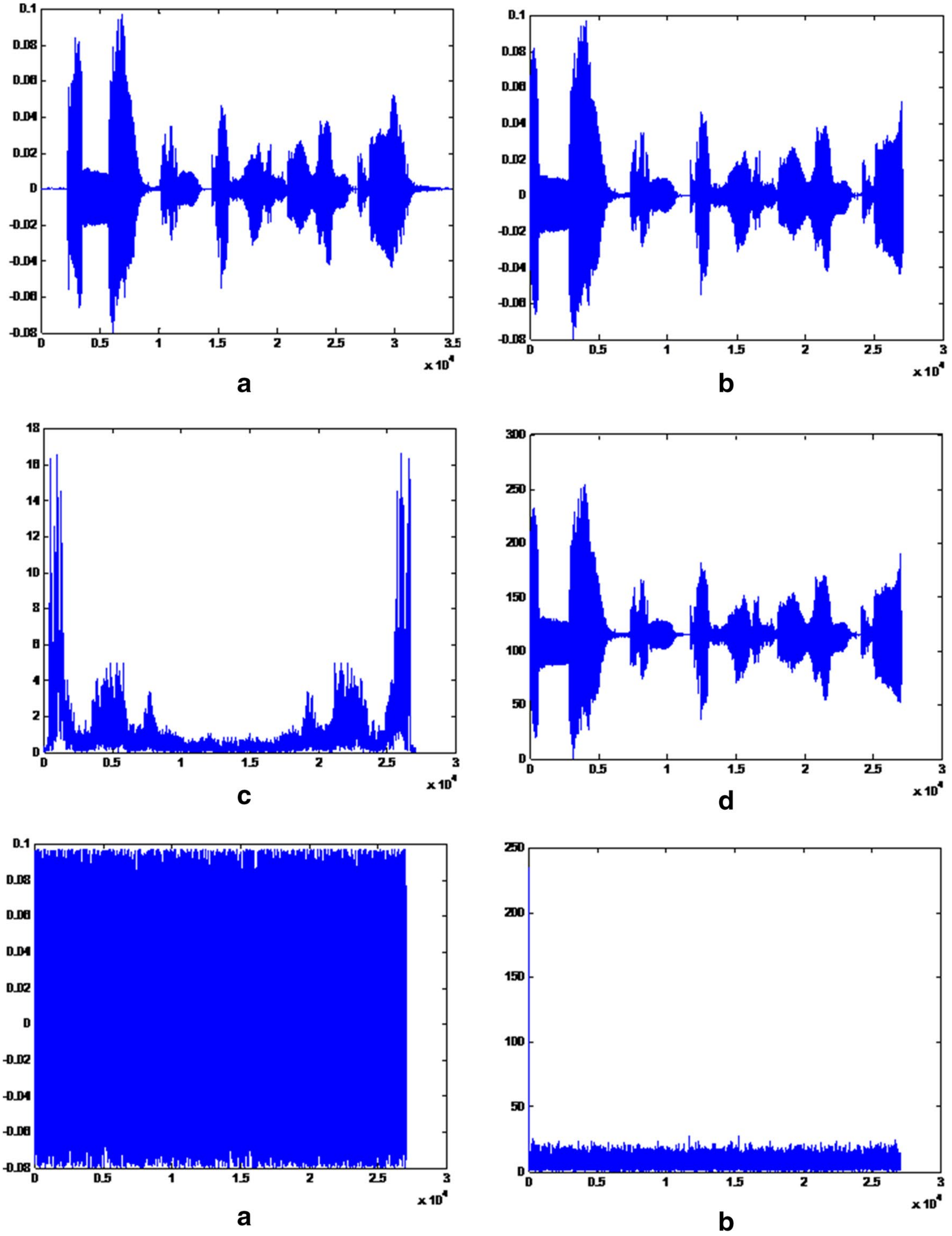

Fig. 6 Test for si2329: **a** original speech, **b** selected speech samples for encryption, **c** spectrum of selected speech, **d** normalized selected speech, **e** encrypted normalized selected speech, **f** spectrum of encrypted normalized selected speech

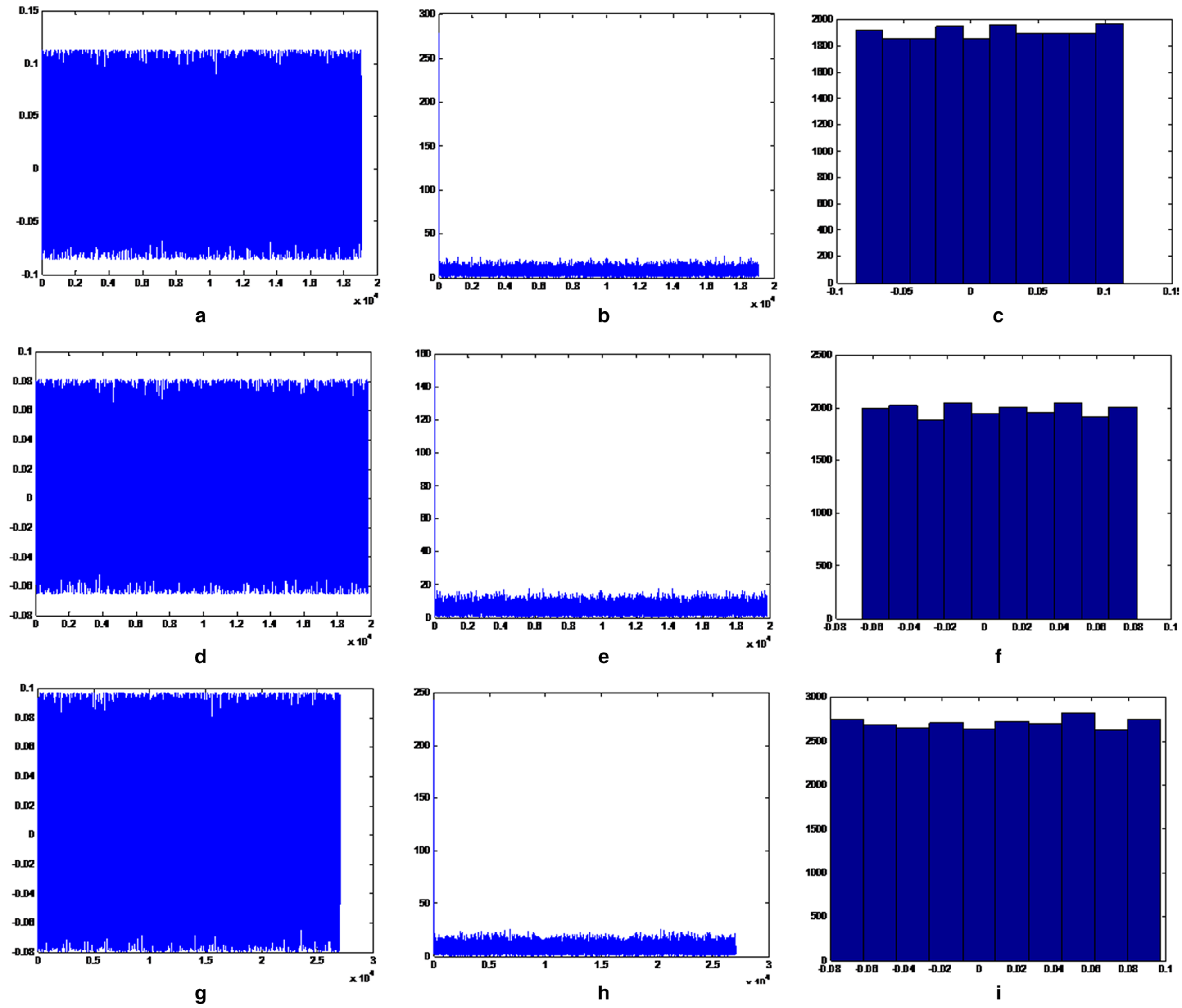

Fig. 7 Test using Graine-128: **a** encrypted normalized selected speech of signal si1699, **b** spectrum of signal shown by **a** and **c** histograms of signal shown by **a**, **d** encrypted normalized selected speech of signal si2257, **e** spectrum of signal shown by **d** and **f** histograms of signal shown by **d**, **g** encrypted normalized selected speech of signal si2329, **h** spectrum of signal shown by **g** and **j** histograms of signal shown by **g**

indicates that the proposed algorithm has correlation values close to zero which confirms its excellent performance.

5.5 Signals to noise ratio (SNR)

The signal to noise ratio is a statistical metric to evaluate the performance of any cryptosystem; it measures the information content in the encrypted data. It can be calculated as follows:

$$SNR = 10 \times \log_{10} \frac{E(x)}{E(d)}. \tag{8}$$

where $E(x)$ is the mean square of the original data and $E(d)$ is the mean square difference between the original and reconstructed data. In Table 2 the proposed scheme gives better negative signal to noise ratio.

5.6 Peak signal to noise ratio (PSNR)

Peak signal to noise ratio can be used to evaluate an encryption scheme. PSNR reflects the encryption quality. It is a measurement which indicates the changes in values between the original speech and itÃ¢â,¬â„¢s encrypted form. The formula to calculate $PSNR(dB)$ is:

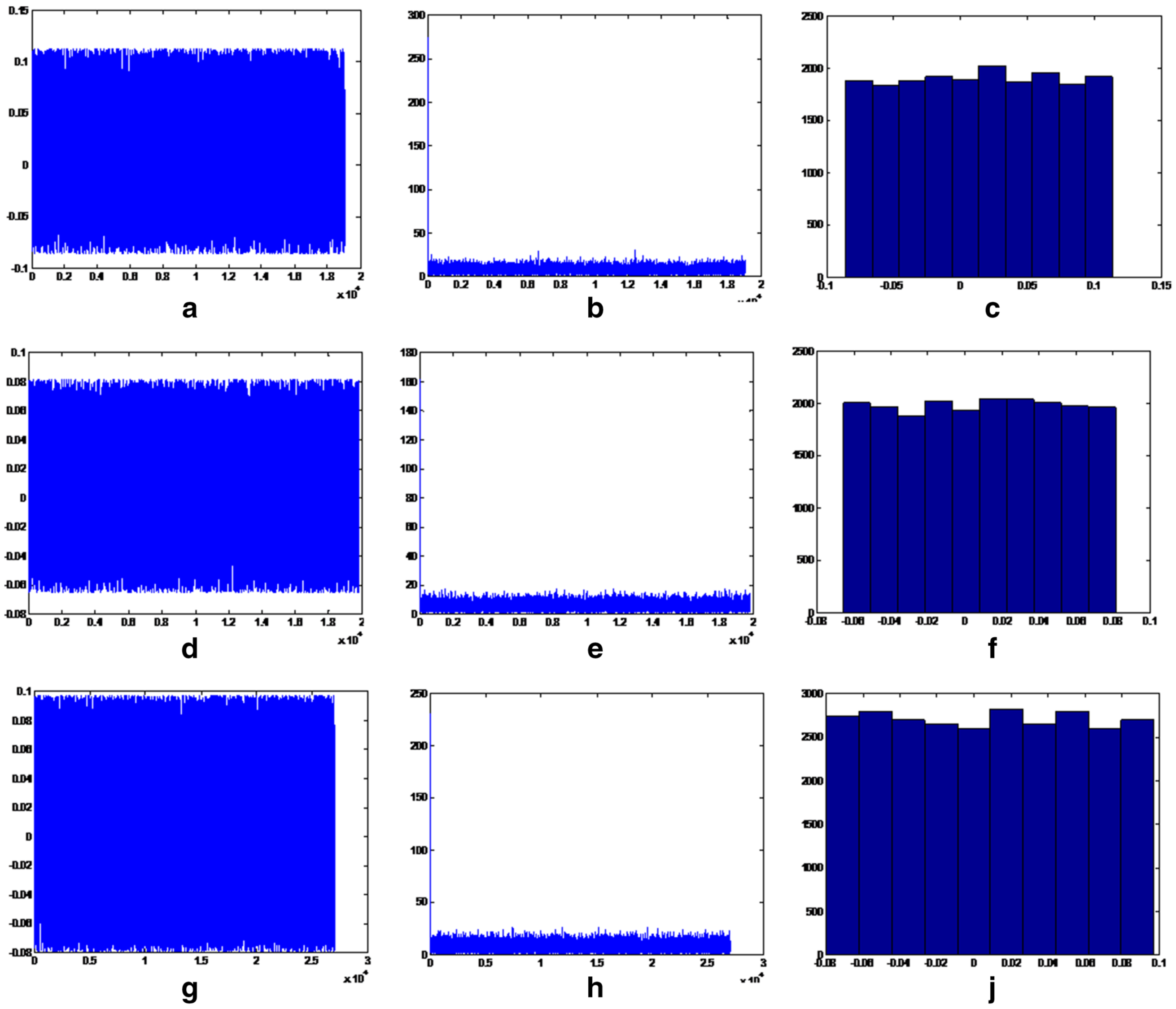

Fig. 8 Test using Achterbahn-128: **a** encrypted normalized selected speech of signal si1699, **b** spectrum of signal shown by **a** and **c** histograms of signal shown by **a**, **d** encrypted normalized selected speech of signal si2257, **e** spectrum of signal shown by **d** and **f** histograms of signal shown by **d**, **g** encrypted normalized selected speech of signal si2329, **h** spectrum of signal shown by **g** and **j** histograms of signal shown by **g**

$$PSNR = 10 \times \log_{10}\left[\frac{n \times (Max(U))^2}{E(d)}\right] \tag{9}$$

where n is the length of the original speech. The lower values of PSNR, the higher is the encryption quality. Table 3 summarizes the PSNR values.

6 Conclusion

In this work, a novel approach for selective speech encryption based on stream cipher is elaborated. The speech signal is pre-processed to remove automatically the silent/unvoiced parts and produce the selected useful speech to be encrypted. This allows a drastic reduction in computing time and resources. The proposed pseudo-random generator allows a large key space which makes the system secure from brute force attack. The algorithm is computationally simple and highly secure and easy to implement for speech encryption and decryption. The experimental results show that the proposed scheme gives large key space, high key sensitivity, lower correlation, good SNR and a more uniform histogram. Analysis of the proposed approach shows superior performance in comparison with Grain-128, Achterbahn-128 and Rakaposhi algorithms.

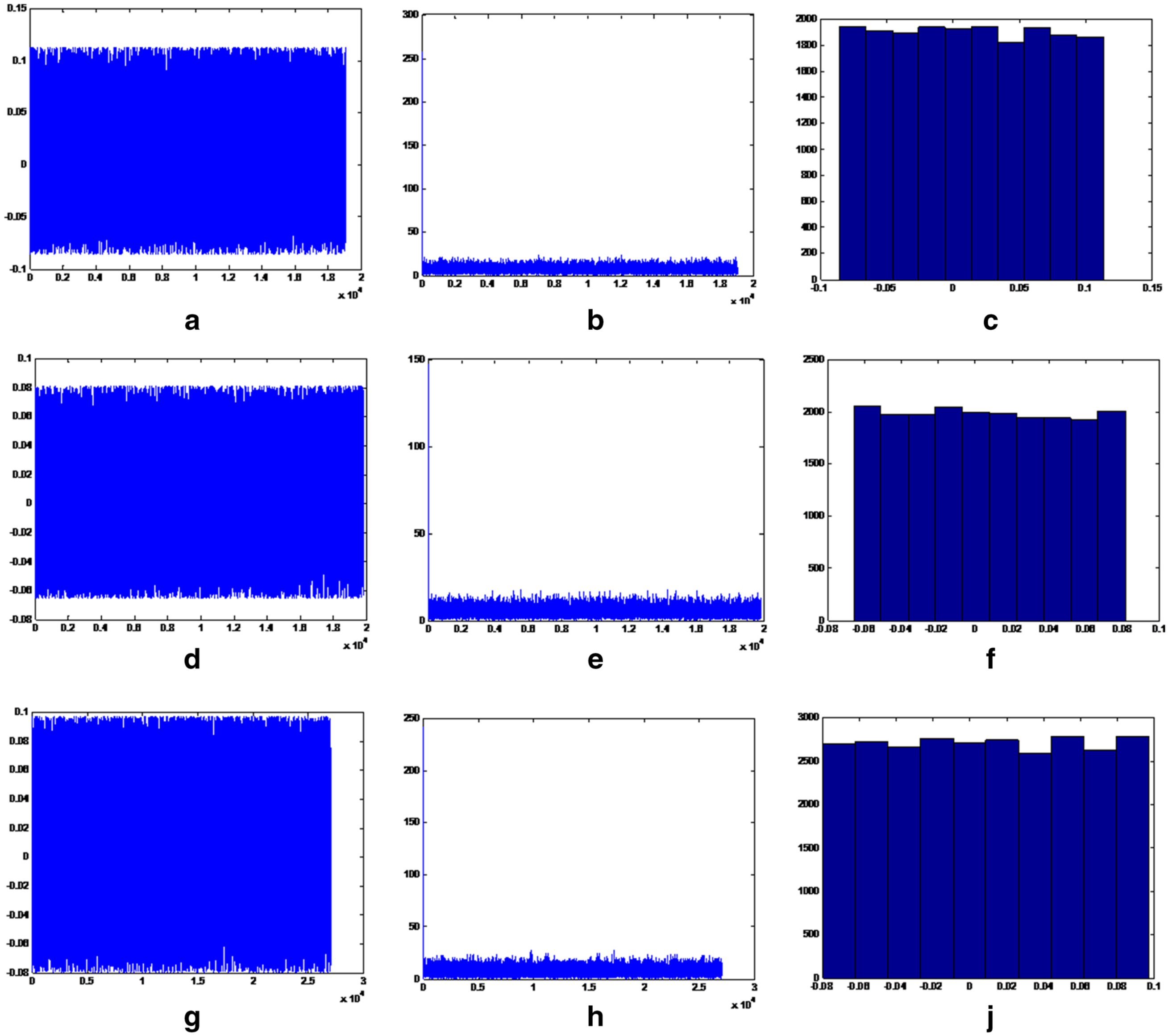

Fig. 9 Test using Rakaposhi: **a** encrypted normalized selected speech of signal si1699, **b** spectrum of signal showen by **a** and **c** histograms of signal showen by **a**, **d** encrypted normalized selected speech of signal si2257, **e** spectrum of signal showen by **d** and **f** histograms of signal showen by **d**, **g** encrypted normalized selected speech of signal si2329, **h** spectrum of signal showen by **g** and **j** histograms of signal showen by **g**

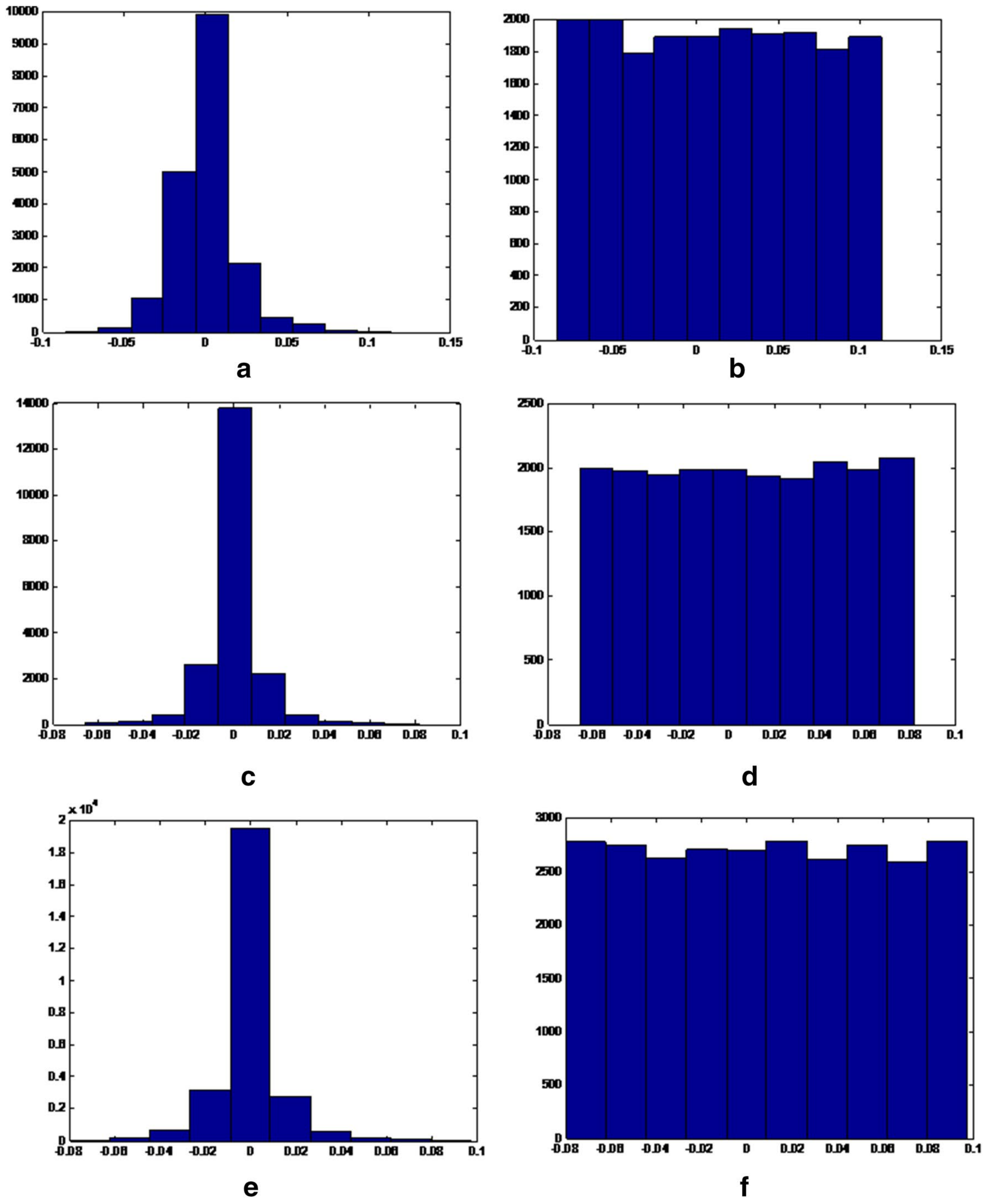

Fig. 10 Frames **a**, **c** and **e** show histograms of selected useful speeches in 4b, 5b and 6b, respectively. Frames **b**, **d** and **f** show histograms of their corresponding encrypted speeches, respectively

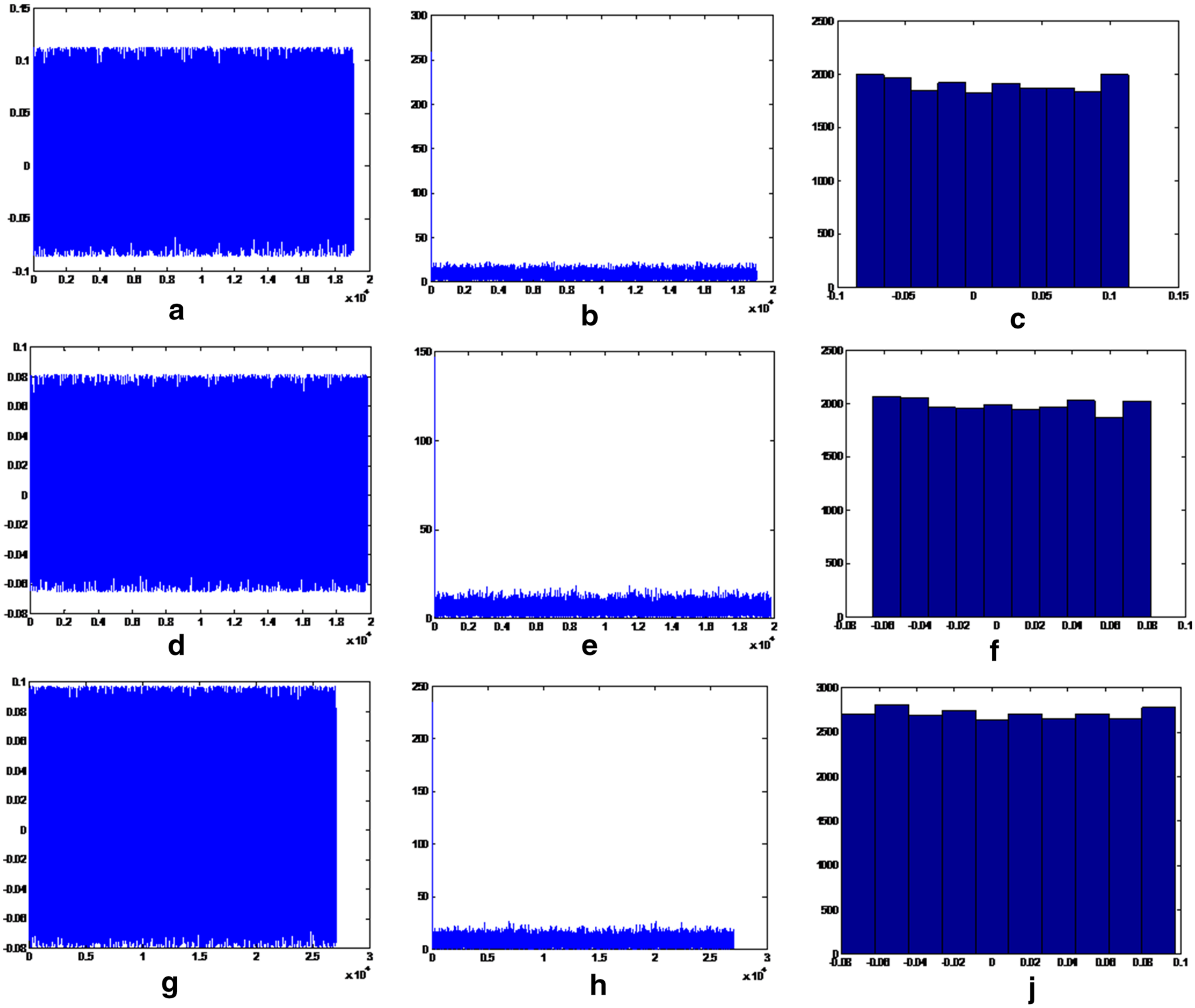

Fig. 11 Sensitivity analysis: **a** decrypted speech 1 with wrong key, **b** the spectrum, **c** the histogram, **d** decrypted speech 2 with wrong key, **e** the spectrum and **f** the histogram

Table 1 Correlation coefficient

Speech	Proposed generator	Grain-128	Achterbahn-128	Rakaposhi
si1699	−0.0011	0.017	0.0013	0.0093
si2257	−0.0033	−0.0100	0.0061	0.0050
si2329	−0.0059	−0.0062	0.00045	0.0051

Table 2 Signals to noise ratio SNR(dB)

Speech	Proposed generator	Grain-128	Achterbahn-128	Rakaposhi
si1699	−10.6357	−10.5807	−10.5967	−10.5371
si2257	−12.0187	−11.9722	−11.9857	−11.9379
si2329	−12.2879	−12.2740	−12.2747	−12.2463

Table 3 Peak signal to noise ratio PSNR(dB)

Speech	Proposed generator	Grain-128	Achterbahn-128	Rakaposhi
si1699	47.9830	48.0379	48.0219	48.0816
si2257	48.1130	48.1595	48.1460	48.1938
si2329	49.4650	49.4788	49.4782	49.5066

References

Abd Elzaher, M. F., Shalaby, M., & El Ramly, S. H. (2016). Securing modern voice communication systems using multilevel chaotic approach. *International Journal of Computer Applications*, *135*(9), 0975–8887.

Ashtiyani, M., Moradi Birgani, P., & Karimi Madahi, S. S. (2012). Speech signal encryption using chaotic symmetric cryptography. *Journal of Basic and Applied Scientific Research*, *2*(2), 1668–1674.

Belmeguenai, A., Mansouri, K., & Lashab, M. (2015). Speech encryption using stream cipher. *British Journal of Applied Science and Technology*, *8*(1), 107–125.

Belmeguenai, A., Berrak, O. & Mansouri, K. (2015). Selective encryption of medical images, In Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP-2015), pp. 93-99, ISBN: 978-989-758-091-8, Berlin.

Belmeguenai, A., Mansouri, K., & Medoued, A. (2015). Speech encryption using the nonlinear filter generator, 16th international Conference on Sciences and Techniques of Automatic control and computer engineering STA'2015, pp. 58–62, IEEE conference (STA conference is indexed scopus), Tunisia.

Bhatnagar, G., & Jonathan Wu, Q. M. (2012). Selective image encryption based on pixels of interest and singular value decomposition. *Digital Signal Processing*, *22*, 648–663.

Burileanu, D., Pascalin, L., Burileanu, C., & Puchiu, M. (2000). An adaptive and fast speech detection algorithm. Proceedings of the 3rd International Workshop on Text, Speech and Dialogue, p. 177182.

Cid, C., Kiyomoto, S., & Kurihara, J. (2009). The rakaposhi stream cipher, Proceedings of the 11th international conference on Information and Communications Security, ICICS'09, Berlin, Heidelberg, Springer, pp. 32–46.

Farsana, F. J., & Gopakumar, K. (2016). A novel approach for speech encryption: Zaslavsky map as pseudorandom number generator, 6th International Conference on Advances In Computing and Communications, ICACC 2016. *Procedia Computer Science*, *93*, 816–823.

Gammel, B. M., Göttfert, R., & Kniffler, O. (2006). Achterbahn-128/80, eSTREAM, ECRYPT Stream Cipher Project, Report 2006/001.

Garofolo, John, et al. (1993). TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1, Web Download. Philadelphia: Linguistic Data Consortium.

Hell, M., Johansson, T., & Meier, W. (2006). A stream cipher proposal: Grain-128, In IEEE International Symposium on Information Theory (ISIT 2006).

Kulkarni, N. S., Raman, B., & Gupta, I. (2008). Selective encryption of multimedia images. *NSC*, *2008*, 17–19.

Lamel, L., Labiner, L., Rosenberg, A., & Wilpon, J. (1981). An improved endpoint detector for isolated word recognition. *IEEE ASSP Magazine*, *29*, 777–785.

Maysaa, A., & Iman Qays, A. (2013). Speech encryption using chaotic map and blowfish algorithms. *Journal of Basrah Researches Sciences*, *39*(2), 68–76.

Musheer, A., Bashir, A., & Omar, F. (2012). Chaos based mixed key stream generator for voice data encryption. *International Journal on Cryptography and Information Security (IJCIS)*, *2*(1), 36–45.

Rahman, Md. M., Saha, T. K., & Bhuiyan, Md. A. (2012). Implementation of RSA algorithm for speech data encryption and decryption. *IJCSNS International Journal of Computer Science and Network Security*, *12*(3), 74–82.

Ramirez, J., Gorriz, J. M., & Segura, J. C. (2007). Voice activity detection, fundamentals and speech recognition system robustness, robust speech recognition and understanding, In M. Grimm & K. Kroschel (Eds.), ISBN: 978-3-902613-08-0.