A Secure Speech Communication Framework for the Embedded System Powered by the High-Frequency Technique – a Study

Prashnatita Pal, Bikash Chandra Sahana and Jayanta Poray

February 17, 2021

A Secure Speech Communication framework for the Embedded System powered by the High-Frequency Technique – A Study

Prashnatita Pal[1*], Bikash Chandra Sahana[2], Jayanta Poray[3],
[1,2]Electronics & Communication Engineering, National Institute of Technology, Patna, India
[3]Computer Science & Engineering, Techno India University, Kolkata, India

Abstract
Cryptography is the art of information protection so that only those for whom the information is meant can able to read and process. It is the science or technique of transforming a message (Encryption) and then re-transform the message back to its original form (Decryption). Cryptography is of two types, namely Symmetric Key or Private Key and Asymmetric key or Public-key Cryptography. For the development of this paper, we have used private key cryptography with eight sound signals. The signals are recognized using Python script and then the matched wave is encrypted using the ElGamal algorithm, a type of Public Key encryption algorithm. The encrypted message is digitized and sent through a channel towards the receiver end after FSK modulation is performed. At the receiver end, the signal is recovered and decrypted to obtain the resultant signal. This signal aimed to match with the original voice signal that was encrypted and sent. The equality ratio turns out to be around 70%. Thus, a voice signal can be easily verified and communicated from one end to another using a secure communication framework.

Keywords: Speech Recognition, Cryptography, ElGamal, Frequency Shift Keying, Reflex Klystron

****Author for Correspondence*** E-mail: prashnatitap@gmail.com, Tel: 9038905200

INTRODUCTION

Speech communication is an important aspect of our life. Security of speech to maintain its confidentiality, proper access control integrity, and availability has been a major issue in speech communication. Therefore, protection of speech passwords or data from misuse is essential. Today in the generation of electronic gadgets, the necessity to prevent data from miscreants is increasing day by day. Cryptography is the process of utilization of codes to prevent anyone from violating speech security. Speech protection can be accomplished by changing the original speech by any means to some other speech codes, so that if someone gets that speech employing hacking then also it must remain in useless bits of speech for that person. This process can be achieved by encrypting that speech by some means of algorithms that are known to the sender and on the other side, similar decryption algorithms must be known to only the desired receiver such that it can convert that encrypted speech back to the user understandable data or signal. To improve the protection mechanism, the ElGamal algorithm [1] (named after its author Taher ElGamal in 1985) is one of the most popular public-key cryptographic algorithms that is used to ensure speech communication security. It consists of two main cryptographic processes. Firstly, a public key is used to convert an input speech into an unrecognizable encrypted output called cipher speech (encryption process), which makes it practically infeasible to recover the original speech without the encryption key. Secondly, a private key is used which converts the unrecognizable speech back to its original form (decryption process). We can use any other pre cryptography algorithms. Its security depends upon the ability of a hacker to compute discrete logarithms. Encrypted data can be transmitted using FSK digital modulation technique [2]. Here FSK is generated using reflex klystron [3]. Frequency Shift Keying (FSK) is one of the popular digital modulation techniques in which the frequency of the carrier signal changes according to the variation digital message signal. The frequency of the output of an FSK modulated wave is high for binary high input and is low for binary low input. It plays an important role in long-distance communication. Reflex klystron is a microwave generator where velocity modulation technique has been utilized to form a high energy density bunch of electrons which suitably reflect to generate high-frequency RF oscillation in a re-entered cavity; It was used as a local oscillator in some radar receivers and a modulator in microwave transmitters the 1950s and 1960s. Demodulation and decryption are done at the receiver.

The conventional MFCC used in [4], [5], [6], and [8] has the disadvantage in removing the specific band noise, has inherently low recognition rates. We have used Noise reduction using spectral gating filtering. This algorithm is based on the one outlined by Audacity for the noise reduction effect. The disadvantage of [7] and [9] is that it can recognize voice signals of a very short period containing at the most one word. Our study deals not only with voice clips containing one word but are also able to deal with voice clips containing several words. This is an enormous advantage over the above-mentioned literature. Since most of the voice recognition systems required in the present world need to handle voice clips containing multiple words spanned over a long interval. In [5] the accuracy of voice recognition was not sufficient; we have improved the recognition accuracy by employing a digital filter. In [3] and [4], the degree of effect of voice disguise on the recognition rate varies with different disguising types. So, it is not easy to understand if a voice is disguised. Our system can easily recognize a disguised voice because it compares the input speech signal and therefore the reference speech signal and allows the communication to happen when the equality ratio is over 0.6. Experimental results in our system have shown that a disguised voice could not achieve an equality ratio of a minimum of 0.6. [5] Has used the Lab View Programming Model which has several disadvantages like a lot of memory is needed and also time-consuming. The developer edition is very costly. It also has debugging issues. We have used Python which is an open-source language and resource-efficient. We used Jupyter Notebooks embedded in Anaconda which is a well-known software for executing Python programs. Several other IDEs are available at which open source.

METHODOLOGY

Speech samples are recorded using a mobile recorder in .mp3 format. But .wav format is desirable for working on the speech sample. Therefore, speech samples in .mp3 format are converted to .wav format using one converter application. Plotted the amplitude vs. time graph for each of the speech samples in MATLAB, which are uploaded into a python-based voice identification system. Jupyter notebooks are embedded in Anaconda which is a very well-known software for executing Python programs. Now, when a person speaks his speech is compared with the recorded speech samples. If it is matching with one of the speech samples, then he is an authenticated speaker and his speech is processed for transmission to the receiver. Otherwise, he is an unauthenticated speaker and transmission to the receiver will not take place. After a match is found, the spectrogram of the corresponding speech sample is plotted after eliminating the noise. ElGamal algorithm is applied on noise eliminated signal for encryption of speech signal. Digital signal was then passed through Reflex Klystron to convert this digital signal into the modulated signal using FSK (Frequency Shift Keying) and transmitted to the receiver as shown in Fig.1. Reflex Klystron will assign two frequencies where high frequency (f_1) is assigned for binary high input and low frequency (f_2) is assigned for binary low input. At the receiver, to identify f_1 and f_2, the FSK signal is passed through a coupler which divides the corresponding signal into two parts. These two parts contain both f_1 and f_2 frequencies. The resulting two signals are passed through two different resonating cavities of frequencies of f_1 and f_2 to identify them. The resulting two signals are then summed up using an adder circuit to get the original speech signal. This signal is amplified and applied to DAC (Digital to Analog Converter) to get back the analog signal. ElGamal decryption algorithm or any other decryption algorithm (ex- like RSA) is applied on analog signal for decryption and get back original speech signal which is spoken as shown in Fig.2. We know the characteristic curves for the reflex klystron (2K25) as used in our paper. Both repeller voltage vs output power and frequency characteristic curves combine in a single page. After observation of combined characteristics select maximum efficiency mode.

Frequency Shift Keying is a well-known digital modulation technique. Here the frequency of the carrier signal varies according to the digital signal changes. FSK is a type or scheme of frequency modulation. The output of an FSK modulated wave is high in frequency for a '1' i.e., High input and is low in frequency for a '0'i.e. Low input. The binary 1s and 0s are called Mark and Space frequencies. Fig.4 is the diagrammatic representation of the FSK modulated waveform along with its input.

Encryption and Decryption Technique

ElGamal encryption system [1] is an asymmetric key encryption algorithm used as public-key cryptography. The basic of this algorithm is the Diffie–Hellman key exchange. It was described by Taher ElGamal in 1985. ElGamal encryption is used in the free GNU Privacy Guard software, recent versions of PGP, and other cryptosystems. The Digital Signature Algorithm (DSA) is one variant of the ElGamal signature scheme, which should not be confused with ElGamal encryption. ElGamal encryption can be defined over any cyclic group G, as a multiplicative group of integers modulo *n*. Its security depends upon the difficulty of a certain problem related to computing discrete logarithms. Like other cryptographic algorithms, **ElGamal** Algorithm has three steps, Key generation, Encryption, Decryption as in Fig.5

Proposed Methodology:

1. Start
2. Speech samples are recorded using the mobile recorder.
3. Conversion into.wav format for ease.
4. Plotting the amplitude vs. time graph for each of the speech samples in MATLAB/Python.
5. Uploading it into Python for ste13c1t further analysis
6. Voice recognition is done by matching with a reference stored speech keyword database.
7. After a match is found, the spectrograph of the corresponding speech sample is plotted in Python after eliminating the noise.
8. Elgamal algorithm or any other algorithm is applied on noise eliminated signal for encryption of speech signal.
9. Digital signal was then passed through Reflex Klystron to convert this digital signal into FSK (Frequency Shift Keying) and transmitted to the receiver. [3]
10. At the receiver, to identify f_1 and f_2, the FSK signal is passed through a coupler which divides the corresponding signal into two parts. These two parts contain frequencies both frequencies f_1 and f_2 in the same phase. [18]
11. The resulting two signals are passed through two different resonating cavities of frequencies of f_1 and f_2 to identify them and then summed up using a circuit to get the original speech signal.

12. ElGamal decryption algorithm or any other decryption algorithm (whichever is applied in Step 8) is applied on analog signal for decryption and get back original speech signal which is authenticated in Step 6.
13. Stop

RESULTS AND DISCUSSIONS

The relevant spectrograms obtained from the voice signal preceded by the plotting of wave plots and recognition and implementation of the ElGamal algorithm on this sound signal are shown respectively in Fig 6 and Fig.7. It shows two such samples used for recognition. The spectrograms after recognition of the correct voice signal and distinguishing into its components is shown below in Fig.8
The sound wave is further processed to plot its harmonic, percussive, and full power spectrogram. This is depicted in Fig.9. These breakdowns are suitable when the sound analysis is done at higher

levels of processing. The ultimate stage in the encryption stage involves implementing the ElGamal algorithm. The waveforms are shown in Fig.10.
The recovered waveform after applying the decryption algorithm is shown below in Fig.11. This waveform is obtained based on software simulation. From figure no 12 [18] calculate two frequencies (F_1 and F_2). We can show the voltage level at these two points (F1 and F2) are the same. In these points, one represents binary '1' (marked here V_x) and another binary '0' (marked as V_y). The experimental setup for obtaining our desired result is shown below in Fig.14. [18]

Comments on results

The results were obtained using python with the help of librosa, matplolib, numpy, PIL, and glob. As the results are based on software simulation hence the original and reconstructed spectrograms show almost match. The DAC was implemented and the output was brought into effect in the form of a LED signal.

FIGURES AND TABLES

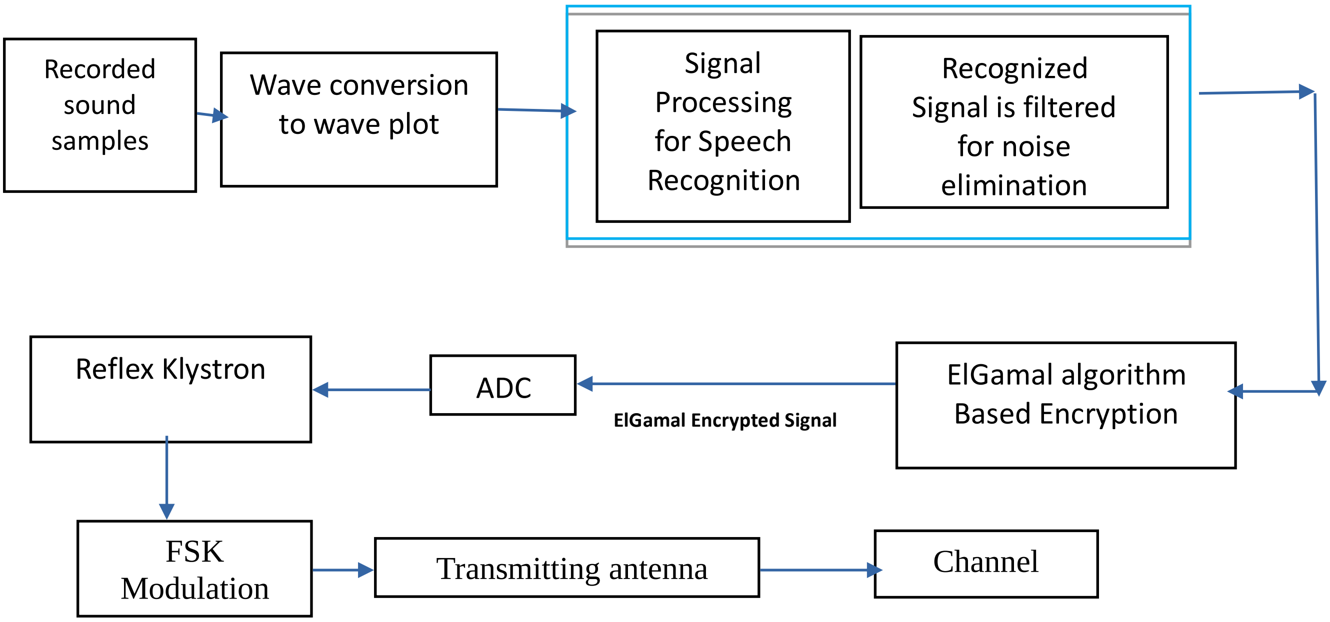

Fig.1. voice authentication, Encryption, and modulation Block at the transmitter

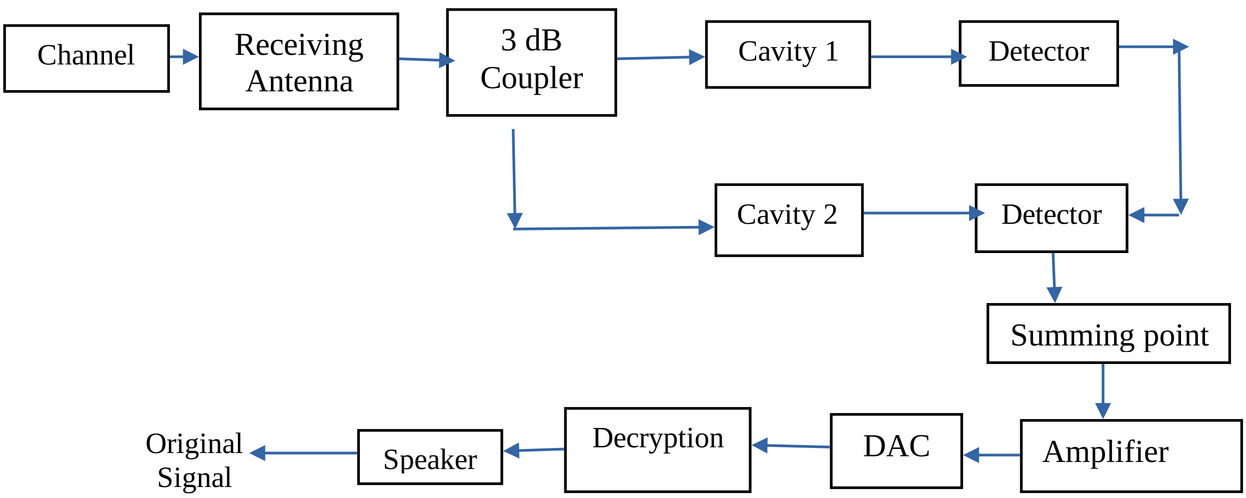

Fig.2.Demodulation and decryption at receiver

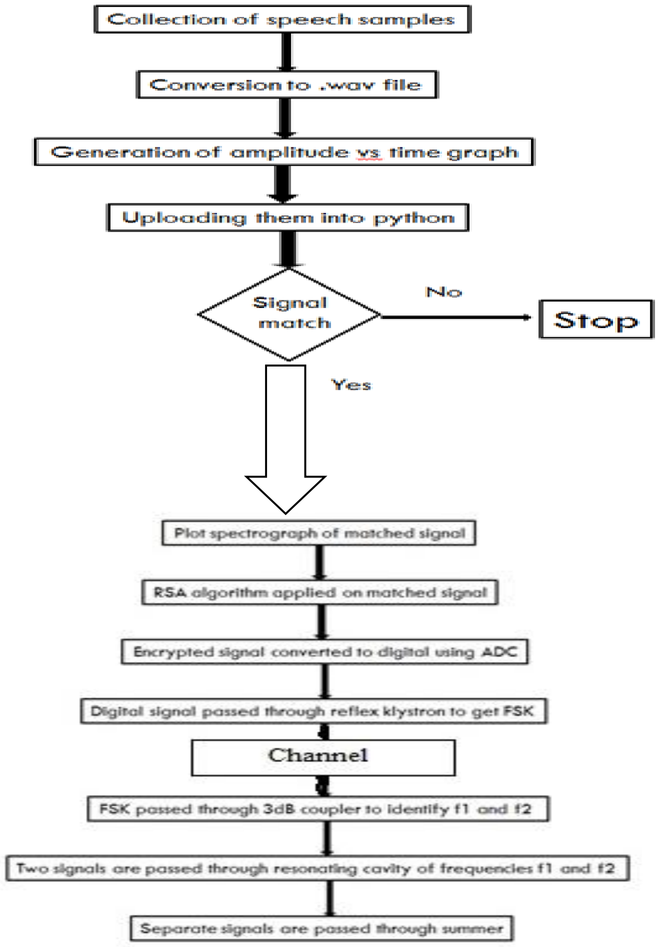

Fig.3 Flow chart of voice password authentication and Modulation and demodulation of the speech signal

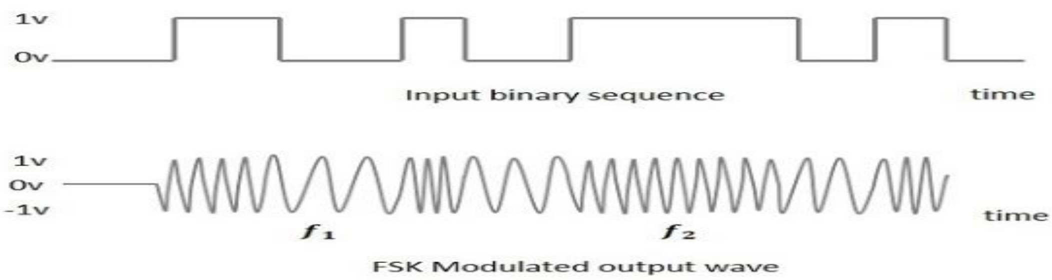

Fig.4. Frequency Shift keying [2]

- **Key Generation**

  Select a large prime as a q
  Select x *to* be a member of the group **G** = < Zq^*, X >, x must be "$1 \leq x \leq q-1$"
  Select g *to* be a primitive root (generator) in the group **G** = < Zq^*, X >
  $y = g^{x} \bmod q$
  Public key ← (g, y, q)
  Private key ← x

- **Encryption**

  Select a random integer r in the group **G** = < Zq^*, X >, r must be "$1 \leq r \leq q-1$"
  $C_1 = g^{r} \bmod q$
  $C_2 = (p \,.\, y^{r}) \bmod q$ // p is the plaintext

- **Decryption**

  $P = [\, C_2 (C_1{}^{x})^{-1} \,] \bmod q$

Fig.5 ElGamal algorithm [1]

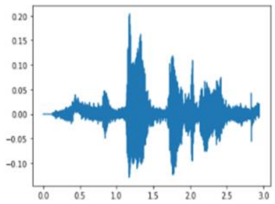

Fig6. Sample-1

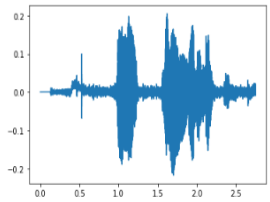

Fig7. Sample-2

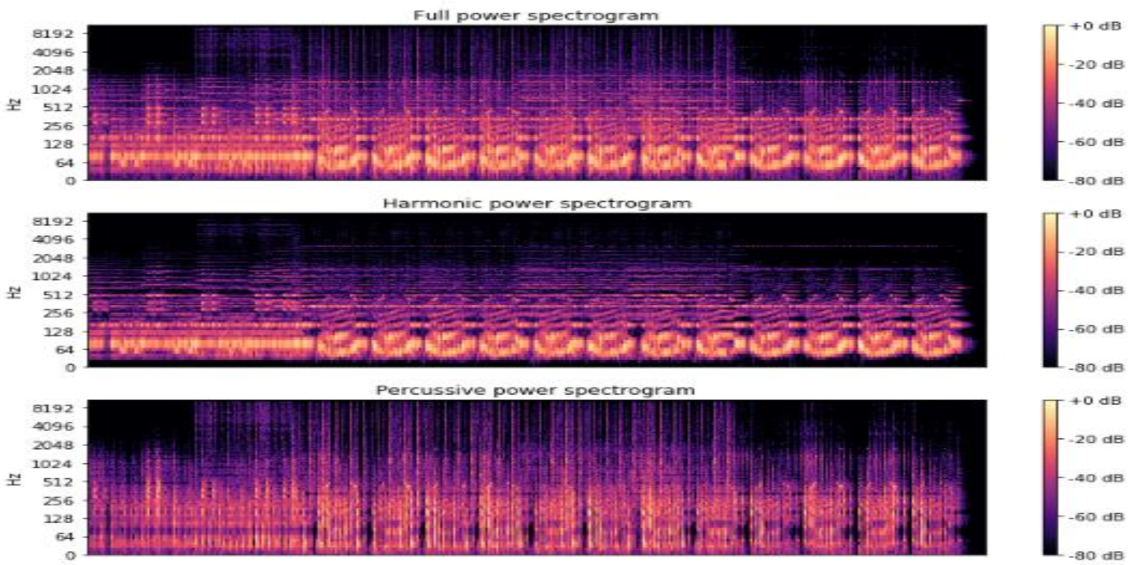

Fig.9. the three corresponding spectrograms

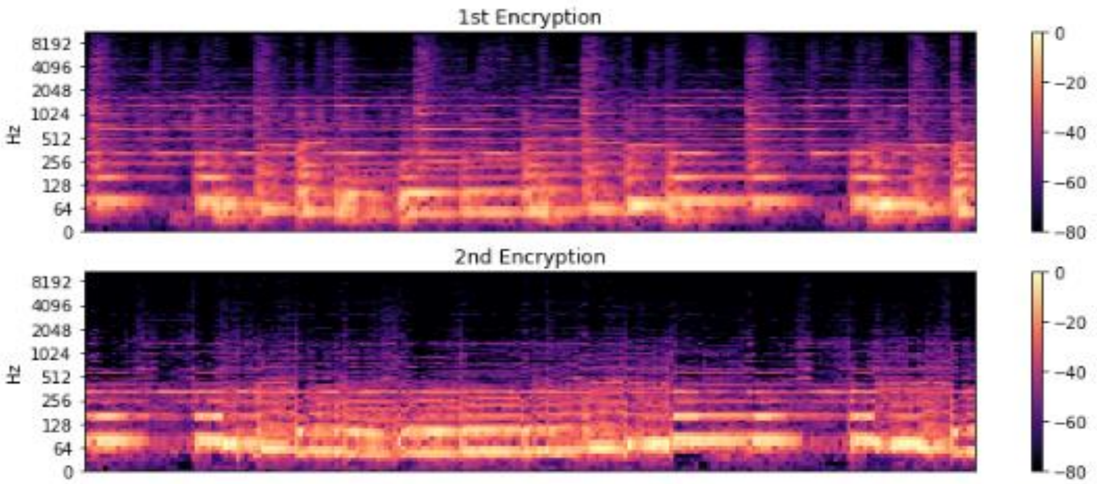

Fig.10. Encrypted Waves

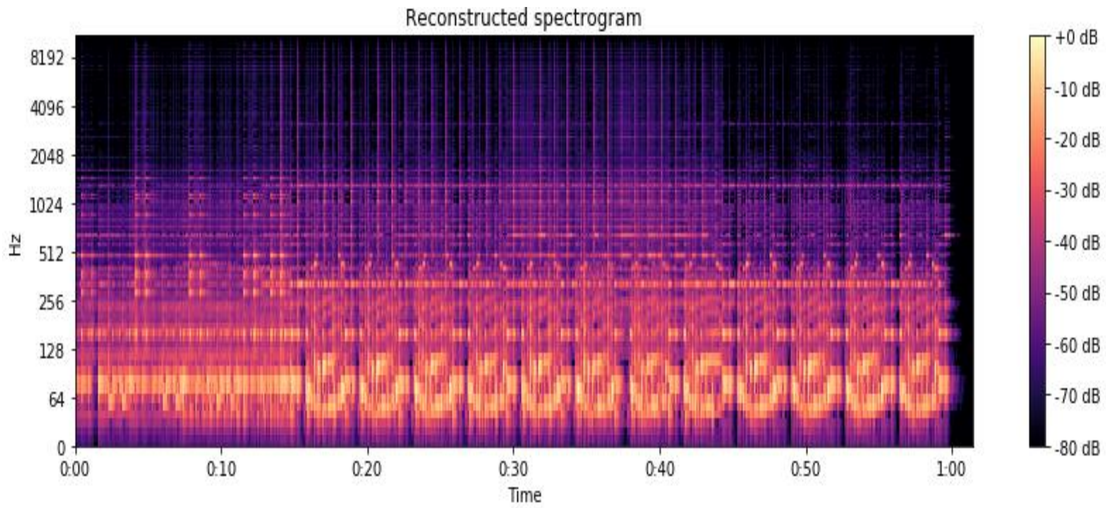

Fig.11. Reconstructed Spectrogram

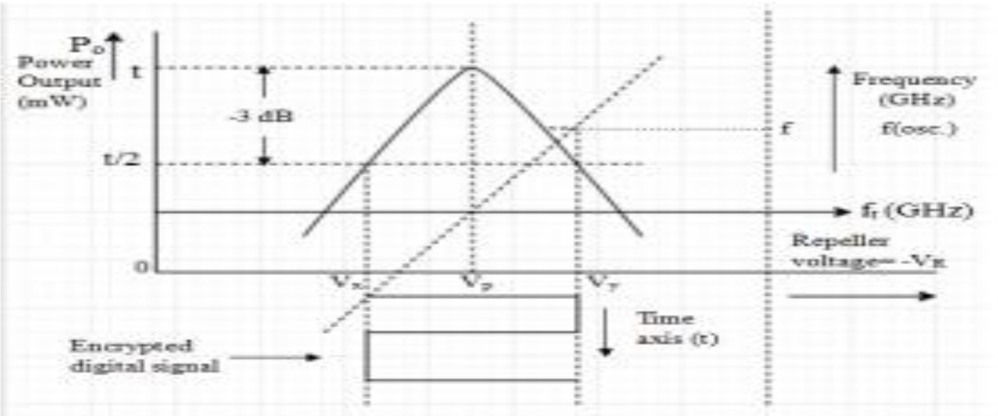

Fig.12. Graph of klystron characteristics using external modulation [3]

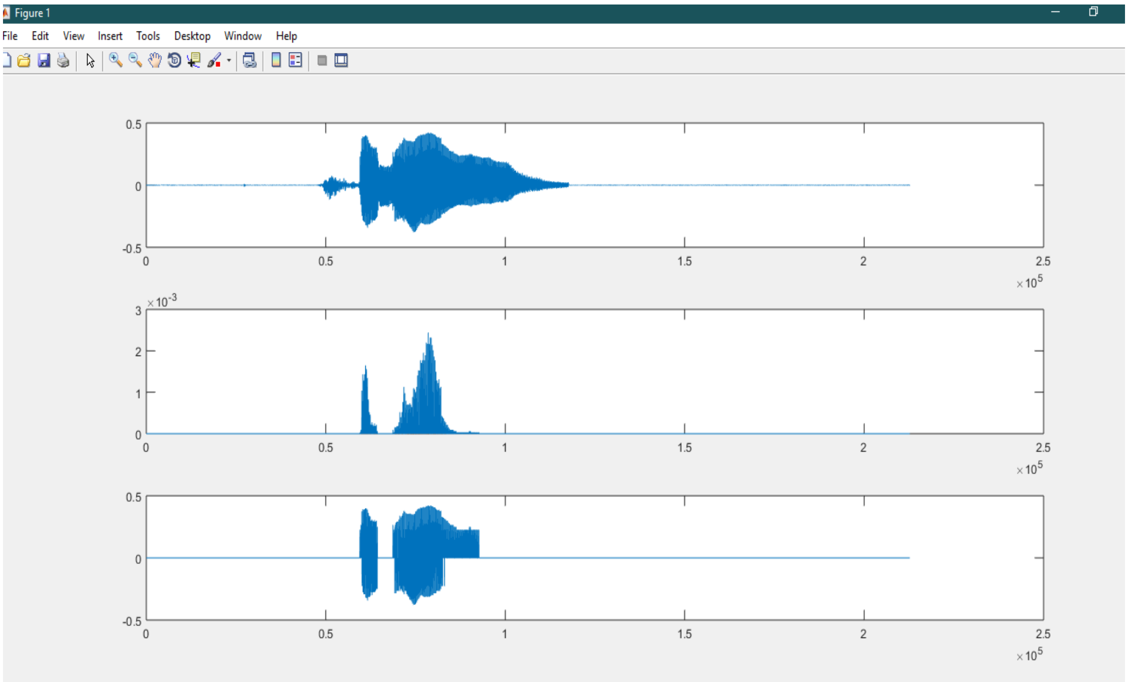

Fig.13. Time-domain representation of i) encrypted signal, ii) chipper signal, iii) decrypted signal [18]

CONCLUSIONS

In addition to ElGamal or other encryption (like) techniques, voice-based authentication of authentic persons gives additional security. Therefore, only authentic person data would be taken for encryption and finally for transmission. The methodology can be used in a highly secured environment like in defense applications. High-power microwave devices like reflex klystron are used for the generation of FSK modulated signals. The signals are reconstructed after the FSK demodulation and decryption process. The similarity index of the reconstructed signal is very high concerning the transmitted signal. Here ElGamal algorithm is preferred as it is more time-efficient and faster concerning the other like as RSA algorithm. [18]

REFERENCES

Journal Articles

1. T. Elgamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms," in IEEE Transactions on Information Theory, vol. 31, no. 4, pp. 469-472, July 1985.

2. Richard W. Middlestead, "Frequency shift keying (FSK) modulation, demodulation, and performance," in Digital Communications with Emphasis on Data Modems: Theory, Analysis, Design, Simulation, Testing, and Applications, Wiley, 2017, pp.207-225

3. Mohuya Chakraborty, & Amiya Kumar Mallick, (2010). "AES Encrypted FSK Generation at X-Band Frequency using a Single Reflex Klystron." Wireless Communication over ZigBee for Automotive Inclination Measurement. China Communications. 7. pp 1-9. 2010

4. Yiu-Kei Lau and Chok-Ki Chan, "Speech recognition based on zero-crossing rate and energy," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, no. 1, pp. 320-323, February 1985.
5. T. Tan, "The effect of voice disguise on Automatic Speaker Recognition," 2010 3rd International Congress on Image and Signal Processing, Yantai, 2010, pp. 3538-3541.

6. N. Obin and A. Roebel, "Similarity Search of Acted Voices for Automatic Voice Casting," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 9, pp. 1642-1651, Sept. 2016.

7. Tahir, Ari. . "Design and Implementation of the RSA Algorithm using FPGA". International Journal of Computers & Technology. Vol 14. 6361-6367. (2015)

Conference Papers

8. H. Bae, H. Lee and S. Lee, "Voice recognition based on adaptive MFCC and deep learning," 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 1542-1546.

9. N. C. Bui, J. J. Monbaron and J. Michel, "An Integrated Voice Recognition System," ESSCIRC '82: Eighth European Solid-State Circuits Conference, Brussels, 1982, pp. 158-161.

10. S. J. Wenndt and R. L. Mitchell, "Machine recognition vs. human recognition of voices," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, 2012, pp. 4245-4248.

11. Y. Yamazaki, M. Tamaki, C. Premachandra, C. J. Perera, S. Sumathipala and B. H. Sudantha, "Victim Detection Using UAV with On-board Voice Recognition System," 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 2019, pp. 555-559.

12. J. Pak and M. Kim, "Convolutional Neural Network Approach for Aircraft Noise Detection," 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 2019, pp. 430-434.

13. H. AlShu'eili, G. S. Gupta and S. Mukhopadhyay, "Voice recognition based wireless home automation system," 2011 4th International Conference on Mechatronics (ICOM), Kuala Lumpur, 2011, pp. 1-6.

14. M. S. I. Sharifuddin, S. Nordin and A. M. Ali, "Voice Control Intelligent Wheelchair Movement Using CNNs," 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), Ipoh, Perak, Malaysia, 2019, pp. 40-43.

15. S. Pleshkova, Z. Zahariev, and A. Bekiarski, "Development of Speech Recognition Algorithm and LabView Model for Voice Command Control of Mobille Robot Motio," 2018 International Conference on High n (HiTech), Sofia, 2018, pp. 1-4.

16. A. Prodeus and K. Kukharicheva, "Automatic speech recognition performance for training on noised speech," 2017 2nd International Conference on Advanced Information and Communication Technologies (AICT), Lviv, 2017, pp. 71-74.

17. Che Yong Yeo, S. A. R. Al-Haddad and C. K. Ng, "Animal voice recognition for identification (ID) detection system," 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, 2011, pp. 198-201.

18. Prashnatita Pal, Bikash Chandra Sahana, Jayanta Poray, and Amiya Kumar Mallick "Generation of Encrypted FSK RF Signals for Secured Communication Inspired with High-Frequency Technique" International Conference on Recent Trends in Artificial Intelligence, IoT, Smart Cities & Applications 2020 (ICAISC-2020).