Towards Security of GSM Voice Communication

Fauzia I. Abro*, Farzana Rauf†, Mobeen-ur-Rehman*, B. S. Chowdhry†, Muttukrishnan Rajarajan‡
*Air University, Islamabad, Pakistan
†Mehran University of Engineering and Tech. Jamshoro, Pakistan
‡City University of London, United Kingdom

***Abstract*—Global System for Mobile Communication (GSM) is widely used digital mobile service around the world. Although GSM was designed as a secure wireless system, it is now vulnerable to different targeted attacks. There is a need to address security domains especially the confidentiality of communication. This paper presents a novel framework for end to end secure voice communication over the GSM networks using encryption algorithm AES-256. A special Modem and speech coding technique are designed to enable the transmission of encrypted speech using GSM voice channel. To the best of our knowledge, this is first solution that uses single codebook for transmission of secure voice. An efficient low bit-rate (1.9 kbps) speech coder is also designed for use with the proposed modulation scheme for optimal results. Different speech characteristics such as pitch, energy and Line Spectral Frequencies (LSF) are extracted and preserved before compression and encryption of speech. Previously, the best achieved data rate was 1.6 kbps with three codebooks, whilst the proposed approach achieves 2 kbps with 0% bit error rate. The empirical results show that the methodology can be used for real time applications to transmit encrypted voice using GSM network.**

***Index Terms*—Global System for Mobile Communication (GSM), Modulator, De-modulator, Codebook, Synthesizer, Analyzer, Encryption.**

I. INTRODUCTION

GSM is the most widely used technology for mobile communication. It is the only communication facility accessible in most rural areas and under developed countries. After unveiling of 3G, 4G and LTE, higher data rates are still available with GSM which makes it viable for data transmission [1]. However, the weak security architecture of GSM makes it vulnerable to eavesdropping and other malicious attacks [2].

Currently, GSM data is encrypted between Mobile Station (MS) and Base station (BS). At BS, signals are decrypted and sent to Mobile Switching Station Centre (MSSC) in clear. The plain speech can be compromised by unintended and malicious parties. Even the encryption between MS and BTS is provided at the discretion of the service providers. GSM uses A3 and A5 encryption algorithms for encrypting data, which are already compromised and breakable in less than a second [3] [4].

Security of GSM channel is imperative as it carries personal data of mobile users. Besides, the Military and other government organizations also use GSM network for sensitive communication. There is a need to have some mechanism to protect the confidentiality and integrity of the communication. In order to achieve end to end security of voice communication, the data needs to be encrypted before transmission over GSM and decrypted at the receiver [5] [6].

Encryption of voice is a challenging task due to design complexity of GSM system and overheads associated with the encryption process. The GSM voice channel is designed only for human speech, and it is not possible to transfer data that does not have speech like characteristics such as an encrypted data or image file [7]. It uses Discontinuous Transmission, Voice Activity Detection and Comfort Noise Generation to detect the voice and cancels the part that is not voice during transmission due to its limited bandwidth. Therefore, it is necessary that encrypted data should be converted to speech like signals before channelling over GSM network. However, the encryption can only be applied on digital data and encryption randomizes the bit stream to a level where recovery of original signal becomes challenging [5] [6] [8] .

Another challenge is the recovery of compressed and randomized speech signals at the receiver end. The inherent characteristics of human speech such as pitch, energy and Linear Spectral Frequencies (LSF) are required to be preserved to accurately recover speech from the randomized data [9].

Limited available channel bandwidth is another challenge for secure communication as encryption adds overheads [5]. The commercially available off the shelf modems are unsuitable for this application as they are not designed to overcome these challenges. Apropos, there is a need to design a customized Modem for secure GSM communication.

To address the aforesaid challenges, this paper proposes an efficient approach for end to end security of GSM voice communication. We present a method for transmission of encrypted data over the GSM voice channel with the help of a specially designed Modem. The voice produced by synthesizer is an artificial voice whose reflection coefficients are provided by the codebooks. The proposed system uses speech like symbols which are selected from the codebook trained on human speech. Our approach uses only one codebook as opposed to the existing approaches which use multiple codebooks. A specially designed low bit rate speech codec [10] is used to provide good quality speech. To encrypt the speech we use AES 256 block cipher.

The proposed approach achieves the data rate of 2 kbps with a reduced processing time. The achieved data rate is comparable with the existing state-of-the-art solutions. The paper provides the first single codebook based method for transmitting secure voice, as the major novelty of our work. Our approach does not need separate noise suppression cir-

Corresponding author: Fauzia I. Abro
Email: Fauzia.Idrees.1@city.ac.uk

cuitry for background noise cancellation as used by other related approaches. The selected design parameters ensure the production of high quality speech. The voicing decision is not required separately since it is deduced from the pitch processing. This results in reduction of system processing time, which makes the proposed approach more efficient. Empirical results suggest that the proposed approach can be effectively used for secure voice communication over GSM voice channel.

To summarize, this paper makes the following contributions.

- A special Modem is designed and developed with a single codebook for efficient transmission of secure voice.
- The most secure digital encryption algorithm AES-256 is used to enhance the security of the system.
- This work uses sensitive speech parameters such as *pitch*, *energy* and *LSF coefficients* for re-producing a high quality speech with nearly no background noise.
- Our approach does not need separate noise suppression circuitry for background noise cancellation. The selected design parameters ensure the production of high quality speech.
- The voicing decision is not required separately since it is deduced from the pitch processing. This results in reduction of system processing time, which makes the proposed approach more efficient.
- Performance of proposed system is validated not only on Computer Environment but also on real-time GSM network.

The rest of this paper is organized as follows. Section-II gives an overview of relate work. Section-III briefly discusses about the background related to the paper; Section-IV describes the proposed approach; Section-V highlights on the experiments and results and Section-VI concludes the paper with possible future work.

II. RELATED WORK

The research on the secure GSM voice communication can be classified into three categories: Voice coding, Encryption and Modulation. Mostly the researchers have worked on one or two out of these three aspects. To the best of our knowledge, so far there is not a single work which has implemented all the three stages.

Aruna and Sukriti [3] presented an encryption algorithm: Tiny Encryption Algorithm (TEA) for encrypting the GSM traffic. Their work is on development of TEA rather than the complete system. The developed encryption algorithm was not tested on GSM network, thereby raising the questions on the efficacy of algorithm itself.

In [6] [5], N. Katugampala et al proposed secure transmission method for security over GSM. In this research codebook, technique is discussed. They also proposed to add encryption before the modulator and after the encoder stage. However, they did not implement any encryption or other security features on their system. The reported results are without any encryption, hence the performance of their system cannot be fairly validated.

In [11], H.F. Qi et al present an end to end Encryption framework over GSM data channel. They presented a general overview and limitations of transmitting voice over the GSM data channel. There are no specific design and implementation details.

Rekha A.B et al [12] gave the basic concept of ciphering model that can be implemented for GSM security. They compared different GSM based secure handsets and proposed to use standard speech codec for both coding and encryption of speech. However, in reality speech codec cannot be used as an encryptor.

Claudi et al [13] proposed Frequency Modulation technique to generate symbols for the modulator stage. Their presented solution is at a preliminary stage and simulated using built-in functionalities of GNU radio simulation environment. Due to poor BER and bit rate, the proposed prototype is not suitable for real-time communication.

Similarly, a QAM based modulated signal is proposed in [14] for using as an encrypted data for real time communication. Although, QAM based coding can provide weak security through scrambling the speech signal, however, it cannot be categorized as a digital encryption algorithm. They did not report any performance parameters (bit rate and BER), which can ascertain the suitability of their proposed approach for real time GSM voice communications.

In another relevant work [15], three encryption algorithms: AES, RSA and NTRU are implemented on two different DSP platforms: fixed-point processor (Blackfin ADSP-BF537) and floating-point processor (TMS320C6711) to identify the best encryption algorithm among three on DSP platfporm for real time secure communication. Their work recommends AES as the most suitable encryption algorithm for real time secure communication application.

Comparing with the existing approaches, our proposed system is the first one which integrated all the sub-modules necessary for secure communication and the resultant solution was successfully tested on real time GSM network with the best throughput 2 kbps with 0% BER.

III. BACKGROUND

This section briefly describes the cryptographic algorithm, modulation and speech coding used in the paper.

A. Speech Coding

Instead of straight conversion of speech to an analogue signal, digitally-encoded speech offers many advantages such as security, signalling and easy regeneration [16]. However, it requires more bandwidth; hence signal compression is necessary for efficiently utilizing the available channel bandwidth. The digitization of speech signal consists of sampling and quantization processes in which speech signal is first sampled and then digitized continuous-amplitude signal is converted to a discrete-amplitude signal (binary).

The signal compression of digitized speech is not straightforward; it involves various digital signal processing techniques to produce low bit rate speech coders. Once the speech is digitized, it is compressed with speech coding algorithms by applying quantization techniques.

The very low bit-rate Vocoders (1.2 or 2.4 kb/s) involve high design complexity. They analyse speech samples to extract the distinct speech parameters which are used at the receiver end to synthesize the reconstructed speech. Linear Prediction (LP) based vocoders are mainly used for this purpose which is especially designed to emulate the human speech production mechanism for high quality speech reconstruction. CELP, MELP and LPC-10 are few examples of LP based Vocoders.

The LPC based vocoder models the vocal tract with a linear prediction filter, the glottal pulses with a periodic pulses and turbulent air flow at the glottis with Gaussian noise. Different speech parameters such as LP filter coefficients signal power, binary voicing decision and pitch are estimated for transmission.

B. Modulation Schemes for Speech Synthesis

The GSM uses a band-limited voice channel designed to transmit only voice-like signals. To transmit the encrypted speech which is actually digital data, the modem needs to be capable of converting digital data into speech like signal at the transmitting side for transmission over low bit rate GSM voice channel and to convert back the speech like signals to digital data (encrypted speech) at the receiver side so that it can be decrypted.

Data communication using GSM voice channel has its own peculiarities that merit designing a new system for transmitting data over voice channel. There can be different ways to represent the data as speech like symbols for transmitting over speech sensitive voice channel. Multiple compression stages and the error corrections techniques used in GSM network cause a significant different between the waveforms produced at the decoder of receiver end and one generated by the data modulator at the transmit end, however, the data modulator has to be capable of recovering the transmitted parameters.

There are two broad approaches to reproduce the speech waveform closest to the original speech: Aanalysis-and-synthesis (AaS) and analysis-by-synthesis (AbS). AaS is not proved good for a low rate applications despite of its capability to produce high quality speech [17].

C. Advanced Encryption Standard

Advanced Encryption Standards (AES) is a symmetric cryptographic algorithm mostly used by the government organizations due to its strong security and fast processing. It is a symmetric block cipher which means it encrypts data on block basis and uses same key for encryption and decryption [18]. Blocks are measured in bits and they represent the length of plaintext and ciphertext. AES has three variants: AES-128, AES-192, and AES-256. Each of these uses data in chunks of 128 bits by using keys of 128, 192 or 256 bits. AES 256 is the strongest cipher which can protect data up to the top Secret level. It is an iterative cipher with 14 rounds as shown in Fig. 1 of substitution and permutation operations to randomize the data. In substitution, inputs are replaced with specific outputs and in permutation, bits are shuffled. AES handles 128 bits of plaintext as 16 bytes (1 byte equals to 8 bits). These 16 bytes are arranged in a matrix of four columns and four rows. Each round of AES comprises of four sub-processes as shown in Fig. 2: Byte substitution, shift rows, mix columns and add round key. In byte substitution, the 16 bytes are substituted with the entries of a pre-defined table called S-box. The second transmutation shifts the four rows of the matrix to the left in a manner that first row remains same, second row is shifted one byte to the left, third row by two positions to the left and fourth row, three positions to the left; producing a new matrix of the same but shifted 16 bytes. Third operation transforms the columns by mixing them to replace the original column. The last process of adding round key is done by performing exclusive XOR on each column using a different part of 256 bits key [18] [19].

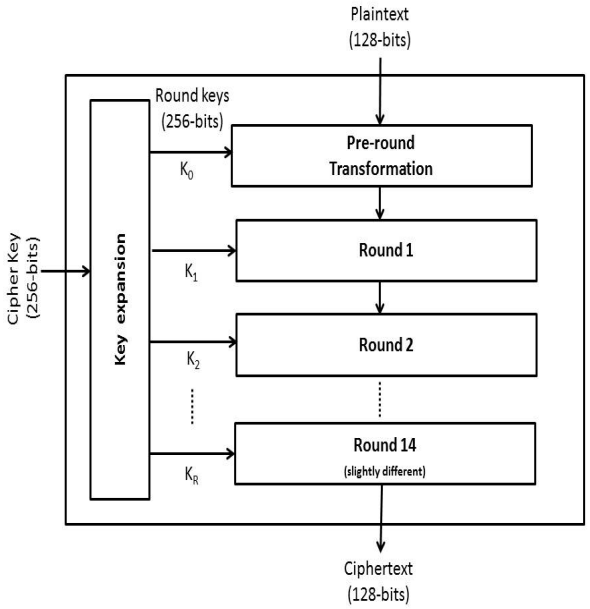

Fig. 1. AES Algorithm operations

D. Elliptic Curve Diffie-Hellman Key Exchange Protocol

Elliptic Curve Diffie-Hellman (ECDH) is a secret key agreement algorithm which defines how to generate and exchange asymmetric keys (public and private key pair) between two parties using elliptic curve to share a secure secret/key over an insecure channel. Once the keys are generated and exchanged, encryption algorithm is selected to encrypt the data. Fig. 3 depicts the working of protocol, when Bob and Alice want to communicate securely; they first generate their key pairs (public and private keys) with the agreed domain parameters. They exchange their public keys and multiply them with their own private keys, these results in an equal shared secret for

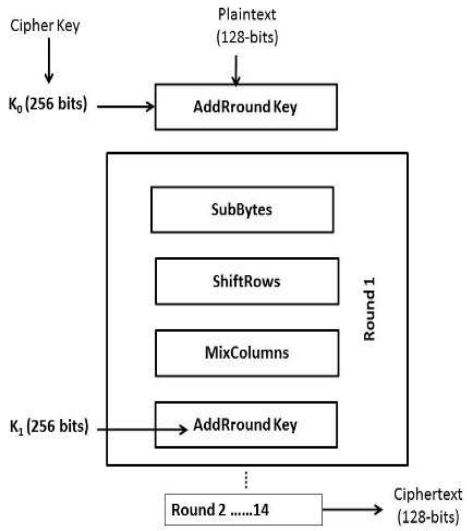

Fig. 2. AES round

both the parties [20]. The shared key can either be used directly as an encryption key or hashed to compute a derived key for encrypting the communication with some symmetric encryption algorithm such as AES. With ECDH secure key generation and exchange mechanism, nobody but the authorized parties knows the secret.

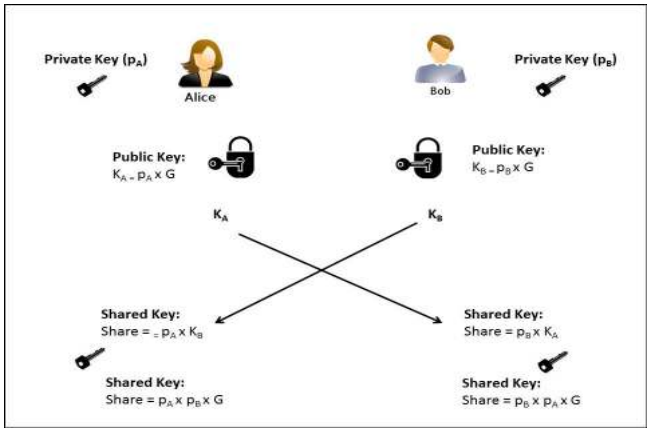

Fig. 3. Elliptic Curve Diffie-Hellman key exchange

IV. Proposed Methodology

The proposed methodology is designed for GSM Full Rate (GSM-FR) which is standardised with a sampling frequency of 8 kHz and data rate of 13 kbps [21]. Fig. 4 shows the basic block diagram of the proposed system. The speech signal is first compressed using a very low bit-rate speech encoder and digitized. The digital bit stream is encrypted and the randomized bits are modulated back onto speech-like waveform, suitable to transmit over GSM voice channel. At the receiver side, the speech like waveform is demodulated and converted into the digital bit stream. After decryption, the bit stream is passed through the decoder for decompression and recovery of speech signal.

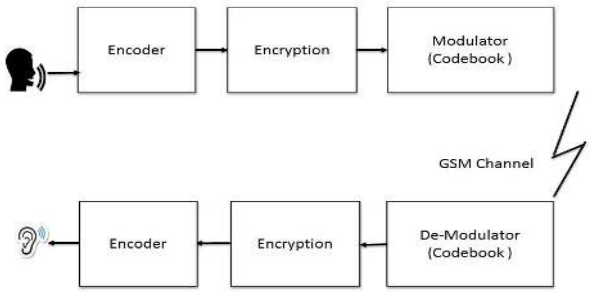

Fig. 4. Diagram of the complete system

In first stage, a low bit lossy speech encoder [10] is used, which extracts the speech characteristics and arrange them into a bit stream. It uses 10th order Linear Predictive Coding (LPC) analysis for compression of input speech to preserve required speech characteristics while compressing the signal. It yields low data rate with a high quality speech.

The digital data stream from the encoder is sent to the Encryptor module. Encryption randomises the bit stream so that it cannot be deciphered by unauthorized parties. The encrypted data can only be decrypted with the right cipher key. There are various Stream and Block ciphers available for use with different applications. The most common encryption algorithms are DES, AES, RSA and ECC [19].

We use Advanced Encryption Standard (AES) 256 bits with 128 bit block size for encrypting the digitized voice stream and Diffie-Hellmann (DH) key exchange protocol with Elliptic Curve 25519 (EC25519) for exchanging the secret key. AES-256 bits is a symmetric block cipher. It works on a fixed length blocks of data and widely used in Military and Government applications due to its strong security architecture. We use the libcrypto library of OpenSSL to perform AES encryption, decryption and secret key generation and exchange operations.

The data stream is split into 128 bits packets and each packet is encrypted by applying rounds of permutation and substitution.The encrypted data is sent to the Modulator stage for conversion to speech like waveform i.e., synthetic speech.

The last stage at the transmitter end is the Modulator (Figure 2), which is the most important module of the system. Without specialised Modulator, the encrypted Speech which is a randomized digital bit stream, cannot pass through speech sensitive GSM channel. The modulator converts the digitized bit stream into speech-like waveform to trick the GSM channel .

A. Speech Coding Technique

To transmit the encrypted voice stream over GSM voice channel, it is necessary to provide an efficient methodology

that overcomes the challenges of limited available bandwidth and randomization of speech signals due to encryption. The coding scheme extracts the vital characteristics of vocal tract that are essentially required for recovery of speech. It also uses new quantization technique for efficiently digitizing the voice stream. Complete design details of Voice coder are presented in [10].

1) Encoder: The encoder works on a 13 bit speech signal sampled at 8 kHz and processed on a frame-by-frame basis. The input signal is divided into frames of 20 ms each (160 samples). Each Frame is processed through the Encoder to extract the required vocal features from it. Firstly, the speech signal is converted into digital signal using a sampling frequency of 8000 Hz .It is then divided into short chunks of 20 ms frames (160 samples), with an overlap of 10 ms (80 samples). Each frame passes through Pre-Emphasizer filter.

Pre-Emphasize block increases the amplitude of high frequency band and decrease the amplitude of low frequency band. High frequencies are more important for signal disambiguation and improve the quality of sound. The output of Pre-Emphasizer block is passed through LPC, Pitch and Energy blocks for extraction of required vocal tract features from the signal [10]. Fig. 5 depicts the block diagram of the Encoder. The input speech signal is processed in real time and the audio stream loop iteratively reads the frames.

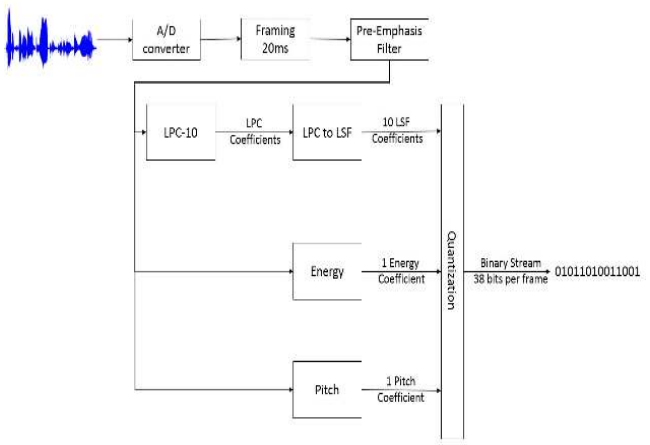

Fig. 5. Encoder diagram

Tenth order auto-correlation coefficients LPC-10 are calculated from the autocorrelation using Levinson-Durbin algorithm, which is a reursive procedure for finding an all-pole IIR filter with a prescribed deterministic autocorrelation sequence. It produces minimum phase filter. LPC is a model for speech signal, based on the assumption that speech is produced by a very specific model consisting of excitation signal and a filter. Excitation is an impulse train or white noise. The LPC analysis filter is an all pole zero filter that extracts the reflection coefficients from the signal. LPC calculates the coefficients by acting as a predictor. The output of this block is the residual signal. The speech signal recovered from residual signal can be expressed as Eq.1.

$$Y(n) = X(n) + \sum_{p}^{k=1} a^k y_{-k} \tag{1}$$

LPC predicts the current output as a linear combination of previous sample. The parameter estimation is repeated for each frame, with the results representing information on the frame. In LPC predictor, the output is feedback to LPC to predict the next value of the sample. The output of LPC-10 is 10 coefficients for each frame. The plot of LPC coefficients is shown in Fig. 6. Parameter estimation process is repeated for each frame.

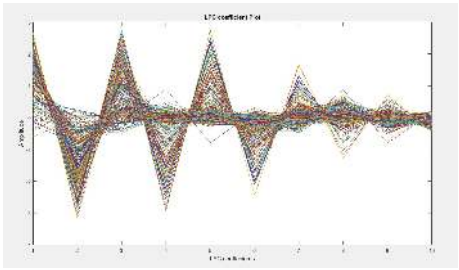

Fig. 6. LPC Plot

LPC coefficients are then transformed into corresponding Line Spectral Frequencies (LSF). LSF normalizes the LPC coefficients lying in the range [-1 to 1] to [0 to pi] by using lpc2lsf function.The transformation of LPC to LSF is shown in Fig. 7. The main objective of conversion of LPC to LSF is to encode LPC parameters with as few possible bits as possible without introducing any other distortion in the signal. The conversion to LSF reconstructs acceptable results at the decoder end.

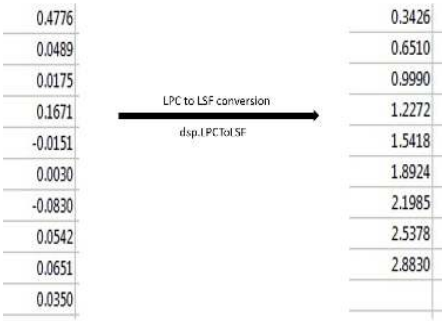

Fig. 7. LPC to LSF conversion

The residual signal is passed through an inverse filter Eq. 2 and output is given to energy block which calculates the energy of signal within the frame. The plot is shown in Fig. 8. This energy information is then passed on to decoder end.

$$A(z) = 1 - \sum_{p}^{k=1} a^k z_{-k} \tag{2}$$

The pitch is calculated using sub-harmonic to harmonic ratio algorithm. It is based on handling the alternative cycles in speech, in which spectrum is compressed along the frequency

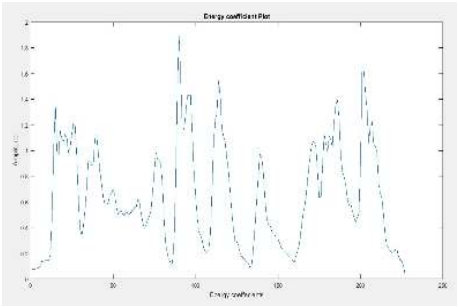

Fig. 8. Energy Plot

axis at different ratios and compressed spectra are added together to make one value of the pitch for each frame. The complete pitch algorithm is based on calculation of certain maxima and minima from the frame. Harmonics and sub-harmonic is calculated on the frame. The Subharmonic-to-Harmonic Ratio (SHR), which is the amplitude ratio between subharmonics and harmonics, is calculated using the difference function as shown in Eq. 3. The plot of pitch is shown in Fig. 9.

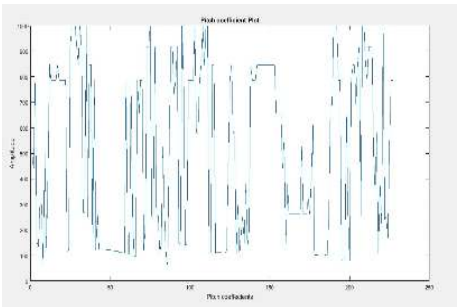

Fig. 9. Pitch Plot

$$SHR = \frac{DA(log(f1)) - DA(log(f2))}{DA(log(f1)) + DA(log(f2))} \tag{3}$$

If value of SHR is less than threshold, it indicates that sub-harmonic are weak and harmonics are strong. So, harmonic value is selected for pitch otherwise sub-harmonic value is selected as a pitch.

The output of frame consists of 10 coefficients of LSF, 1 coefficient of pitch and 1 coefficient of Energy per frame. Fig. 10 shows the energy, pitch and LSF coefficients.

The pitch and energy is converted to binary format using de2bi function in Matlab, while LSF coefficients are quantised using the Linked Split Vector quantization technique [22].

The complete pseudocode is shown in Algorithm 1.

2) Decoder: At decoder end all the above calculated parameters are gathered to produce the speech (Fig. 11). The output of synthesis filter is the original signal. This is played through Audio Device Writer block. The synthesis filter is based on the assumption that speech is either voiced (having pitch) or unvoiced (no pitch). Voiced sound is generated through vowels by the vibration of vocal cord. These vibration is periodic in time, thus are approximated by an impulse train. Pitch is distributed as the spacing between impulses. Voiced sound

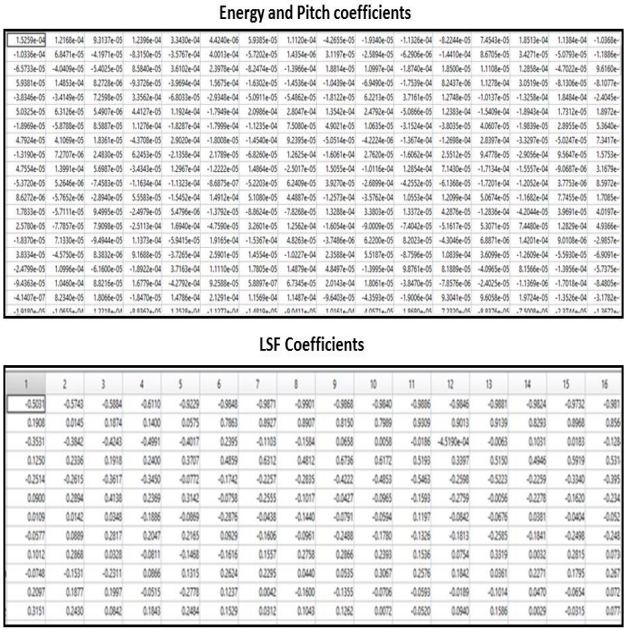

Fig. 10. Energy, Pitch and LSF coefficients

Algorithm Split Band LPC ()

```
[input_speech, freq]←audiodata(speech)
[data,frame] ← data to frame(input_speech)
for frame in Frames
    [lpc,energy]←LPC_func(data,frame)
    [pitch]←SHR_ratio (data,frame)
end
for frame in Frames
    win ← hamming_window (frame);
    energy(:, frame) ← frame(1,frame)*energy(:,frame);
    Pulse←excitation(size(data),Pitch);
    excit←LPC_Filter(Pulse,lpc(1,frame),energy);
    syn←synthesizer(excit,fft_size,win,11000);
end
```

(having pitch) is modelled by an impulse train of the signal while unvoiced sounds have no pitch since they are excited as white Gaussian Noise. Energy is added together to pulse excitation. Output is finally passed through LPC filter that restore the original signal power distribution. The frequency response of LP filter is found from the equation Eq. 4.

$$He^{-j2pikf/fs} = \frac{Energy}{1 - \sum_{p}^{k=1} a_k e^{-j2pikf/fs}} \tag{4}$$

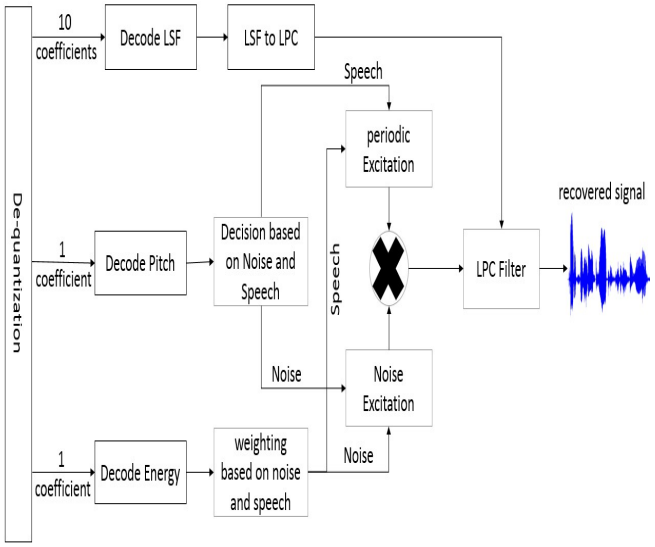

Fig. 11. Decoder diagram

B. Modulation Scheme

The modulator shown in Fig. 12 uses a codebook trained on human voice. The codebook is trained using Linde-Buzo Gray (LBG) algorithm, which needs less number of inputs and operates efficiently on the real network [23] [24]

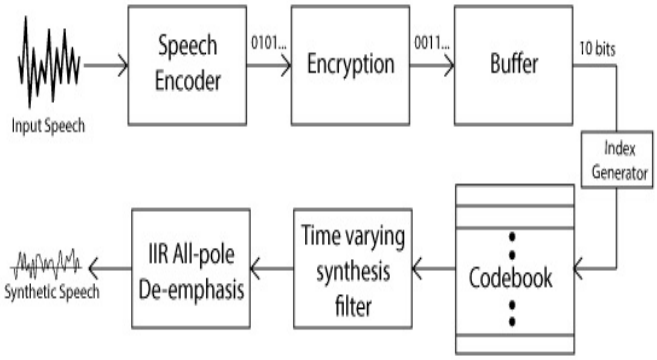

Fig. 12. Diagram of the Modulator

Fig. 13 shows the codebook generation process. The Reflection coefficients are extracted from the training speech dataset and used to train the codebooks.

The incoming data is analysed with the codebook. It uses K-means method (shown in Algorithm 1) to set codewords of the codebook for the incoming data. Both transmitting and receiving sides use same Codebook based on Multiple-Input-Multiple-Output (MIMO) technology. The transmitter encodes the bits through codebook while receiver selects the best possible index to recover encoded signal.

The modulator carries a fixed trained codebook where each codeword carries 20 reflection coefficients to make synthetic speech. A codebook contains a total of 1024 codewords . At

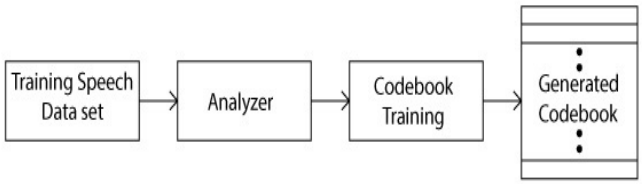

Fig. 13. Codebook Training Flow

Algorithm 1: K-Means Algorithm

Input: $E = \{e_1, e_2, \ldots, e_n\}$ (set of entities to be clustered)
k (number of clusters)
$MaxIters$ (limit of iterations)
Output: $C = \{c_1, c_2, \ldots, c_k\}$ (set of cluster centroids)
$L = \{l(e) \mid e = 1, 2, \ldots, n\}$ (set of cluster labels of E)

foreach $c_i \in C$ **do**
  $c_i \leftarrow e_j \in E$ (e.g. random selection)
end
foreach $e_i \in E$ **do**
  $l(e_i) \leftarrow argminDistance(e_i, c_j) j \in \{1 \ldots k\}$
end

$changed \leftarrow false;$
$iter \leftarrow 0;$
repeat
  foreach $c_i \in C$ **do**
    $UpdateCluster(c_i);$
  end
  foreach $e_i \in E$ **do**
    $minDist \leftarrow argminDistance(e_i, c_j)\ j \in \{1 \ldots k\};$
    if $minDist \neq l(e_i)$ **then**
      $l(e_i) \leftarrow minDist;$
      $changed \leftarrow true;$
    end
  end
  $iter++;$
until $changed = true$ *and* $iter \leq MaxIters$;

single time, modem takes 10 bits as input and these bits are treated as index of codebook and the particular codeword of the input index is selected.

The analyzer module uses different functions to extract the reflection coefficients as shown in Fig. 14. The major components of analyzer and synthesizer are discussed in subsequent sub-paragraphs.

Pre-Emphasis Filter: Pre-Emphasis Filter inputs the audio signal and increases the magnitude of the higher frequencies with respect to other low frequency components. It enhances the overall signal to noise ratio by reducing the adverse effect of attenuation or saturation given in eqs 5 and 6 respectively.

$$y_n = x_n - ax_{n-1}, \text{where a}\epsilon[0, 1] \tag{5}$$

$$Y(z) = (1 - az^{-1})X(z) \tag{6}$$

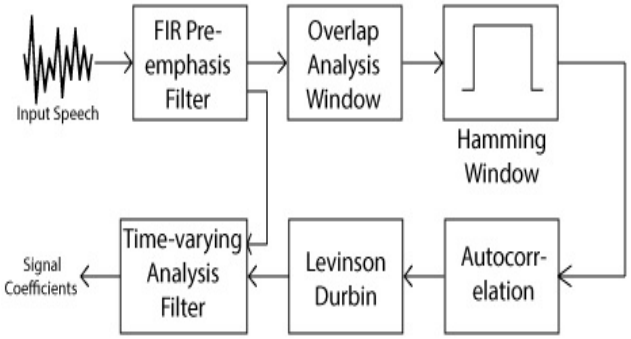

Fig. 14. Diagram of the Analyzer module

De-Emphasis Filter: De-Emphasis Filter inputs the speech waveform and reduces the magnitude of the higher frequency components with respect to the lower frequency components. It is the mirror operation of Pre-Emphasis Filter. This operation is also performed in order to reduce the signal to noise ratio.

Overlap Analysis Window: Overlap Analysis Window performs analysis of similarity between different frames. It subdivides the input speech and than performs overlap analysis and windows them separately as given in eq. 7.

$$X(n, w_0) = \sum_{m=-\infty}^{\infty} (x[m]e^{-jw_0 m} w[n-m]) \tag{7}$$

Hamming Window: Hamming Window is utilized here as a low pass filter which allows low frequency components to pass and blocks the high frequency components. As the higher frequency components contain noise and other unwanted information. Ripples are removed and input speech is smoothed (eq. 8).

$$w(n) = 0.54 - 0.46cos(\frac{2\Pi n}{M}) \tag{8}$$

Auto-correlation: Auto correlation coefficients are found using autocorrelation operation of one frame with another frame.

Analysis Filter: The original speech signal is passed through an analysis filter, which is an all-zero filter with coefficients as the reflection coefficients obtained above. Residual signal is the output of analysis filter. The transfer function used for this purpose is given in eq. 9.

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k * z^k} \tag{9}$$

Where G is kept as 1 for all the frames. Gain is kept constant all over.

Levinson-Durbin: The Levinson-Durbin algorithm [25] is used to compute the linear prediction filter coefficients in speech encoders.It uses autocorrelation to estimate the linear prediction parameters of speech sample. Fig. 15 shows the reflection coefficients computed with Levinson-Durbin algorithm.

Function input: $r_{j=[0,P]}$ is the array of autocorrelation of the array of real numbers $x_{n=[0,N-1]}$

$$r_{j=[0,P]} = \frac{1}{N-1} \sum_{n=1}^{N-1} x_n x_{n-1}$$

and P is the number of past samples of x_n which we wish to examine.

Function output: $a_{i=[0,P]}$ array of LP coefficients.

function $lpc(r_{i=[0,P]})$

$E_0 \leftarrow r_0$
for $i = 1$ to P
  $k_i \leftarrow \frac{1}{E_{i-1}}(r_i - \sum_{j=1}^{i-1} b_j \cdot r_{|i-j|})$
  for $j = 1$ to $i = 1$
    $a_j \leftarrow b_j - k_i b_{i-j}$
  end for
  $a_i \leftarrow k_i$
  $E_i \leftarrow (1 - k^2) E_{i-1}$
  $b_m = [1, P] \leftarrow a_m = [1, P]$
end for
$a_0 \leftarrow 1$
$a_{m=[1,P]} \leftarrow -(a_{m=[1,P]})$
return a_i

end function

Fig. 15. Levinson Durbin algorithm [25]

Synthesis Filter: Synthesis filter inputs coefficients as well as the error signal calculated. So in proposed methodology the error signal is kept constant and coefficients are varied.

The variation in coefficients is decided by the input bits of modulator. Synthesis filter gets residual signal and coefficients to generate a speech like waveform. That speech like waveform is transmitted over GSM channel after passing through the de-emphasis filter. the output of the synthesizer is shown in Fig. 16.

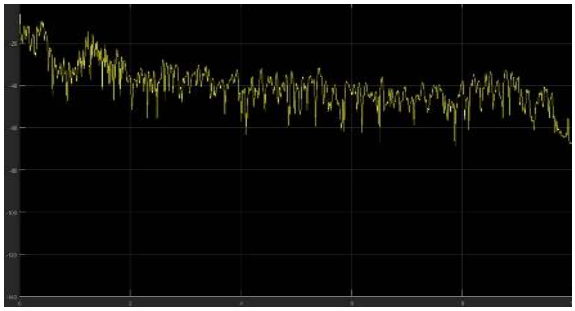

Fig. 16. Synthesizer output

The demodulator is shown in Fig. 17. It receives speech like symbols and applies same analysis as at the transmitter side (discussed above) to calculate the coefficients of received symbol. These coefficients are auto-correlated with the codebook entries. Index of the highly correlated entry is extracted and converted back to bits. We received the transmitted bit

with high accuracy due to extensive training of codebooks.

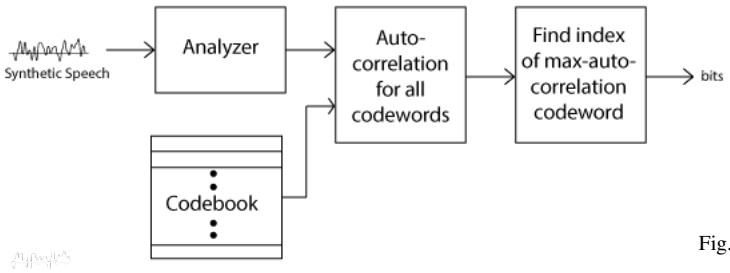

Fig. 17. Demodulator Block Diagram

For the purpose of codebook training 128 different audio files were used. By applying efficient LBG algorithm for codebook training, a highly varied codebook was generated. Fig. 18 shows the signal values compared with the estimated ones. These estimated values are used for auto-correlation with the codebook centroids. Fig. shows the reconstructed speech at the receiver end.

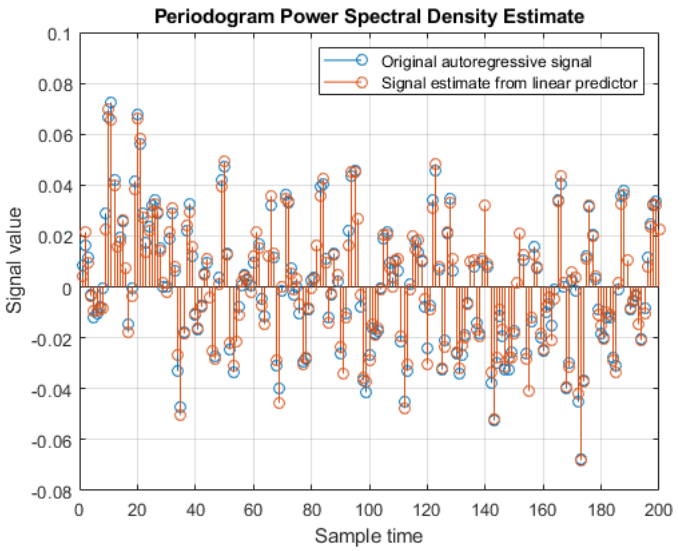

Fig. 18. Power spectral density estimate

V. Experiment and Results

The proposed system was implemented on two Personal Computers (PCs), each with an Intel Core i3-8100 processor and 8 GB RAM and Microsoft Windows 7 Operating system. Two GSM mobile phones were connected to the PCs with the hands free cable. All the functionalities (encode, decoder, modulator, demodulator, encryptor and decryptor) were implemented on both the PCs. Real-time GSM-to-GSM call was made with phones passing through the PCs. The quality of speech without encryption was better and the throughput of 2 kbps was achieved with 0% BER, which is the best data rate achieved so far. The reconstructed speech waveform is shown in Fig. 19. The prototype of system is presented in Fig. 20.

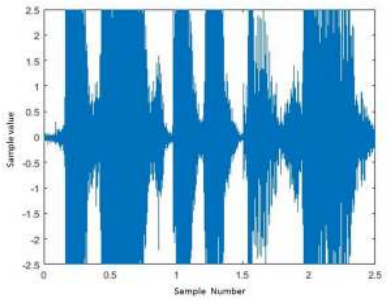

Fig. 19. Reconstructed speech

Table 1 compares the results of the proposed technique with the existing state of the art. Our approach outperforms the state of the art in throughput and BER. The compared approaches were tested in computer environment unlike the proposed approach which was tested both in computer and real time environment.

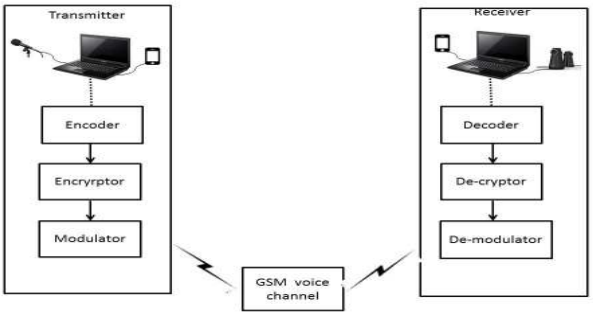

Fig. 20. Proposed prototype

TABLE I
Comparison with other algorithms

Algorithm	Vocoder	Data Rate	BER
Proposed	FR	2(kbps)	0.00
[21]	FR	1.6(kbps)	0.00
[26]	FR	0.6(kbps)	0.00
[6]	EFR	3(kbps), 4(kbps)	0.5,0.4
[27]	FR	1.15(kbps)	0.02

VI. Conclusion

GSM security framework relies on weak encryption algorithms (A3 and A5), which are optionally activated between mobile phone and base station. We have presented a secure voice solution for GSM. Speech is compressed with an indigenously developed low bit rate Voice coder then encrypted with a military grade encryption algorithm AES-256 and finally modulated with a specially designed modem to SL symbols for relaying over the GSM network. The proposed system achieved a data rate of 2 kbps with 0% BER when no encryption is applied. To the best of our knowledge, this is the optimum result reported in this particular research.

The proposed system is efficient due to reduction of number of codebooks used for bit retrieval. In future, we plan to implement the complete system on two Raspberry Pi devices. We also plan to optimize the approach to further improve the data rate without increasing the BER.

REFERENCES

[1] Xu, Zhan, *Data Transmission Method based on Single Carrier over GSM Voice Channel*, *Revista de la Facultad de Ingeniera*, vol. 32, no. 9, 2017.

[2] Cattaneo Giuseppe, De M Giancarlo and Petrillo U Ferraro, *Security Issues and Attacks on the GSM Standard: a Review*, *J. UCS*, Vol. 19, no. 16, pp 2437-2452, 2013.

[3] Aruna Chouhan and Sukriti Singh, *Real Time Secure End to End Communication over GSM Network*, *Int. Conf. on Energy Systems and Applications*, 2015.

[4] Biancucci G, Claudi A, Dragoni AF, *Secure data and voice transmission over GSM voice channel: applications for secure communications*, *4th Int. Conf. on Intelligent Systems Modelling & Simulation*, pp. 230-233, 2013.

[5] N. Katugampala, K.T. Al-Naimi, S. Villete and A.M. Kondoz, *Real Time End to End Secure Voice Communication over GSM Voice Channel*, *13th European Signal Processing Conference*, 2005.

[6] N. Katugampala, S. Villeile and A.M. Kondoz, *Secure Voice over GSM and Other Low Bit Rate Systems*, *IET*, 2003.

[7] LaDue Christoph K, Sapozhnykov Vitaliy V and Fienberg Kurt S, *A data modem for GSM voice channel*, *IEEE Transactions on Vehicular Technology*, Vol. 57, no. 4, pp 2205-2218, 2008.

[8] A.Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, *J. Wiley* New York, 1994.

[9] Yang Yucun, Feng Suili, Ye Wu and Ji Xinsheng, *A transmission scheme for encrypted speech over GSM network*, *Computer Science and Computational Technology, Int. Symp.on*, pp 805-808, 2008.

[10] Fauzia I. Abro, Farzana Rauf, Mouazma Batool, B. S. Chowdhry and Saleem Aslam, *An Efficient Speech Coding Technique for Secure Mobile Communications*, *In 8th IEEE Info. Tech., Electronics and Mobile Comm.Conf. (IEMCON)*, 2018.

[11] H.F. Qi, X.H. Yang et al, *Novel End-to-End Voice Encryption Method in GSM System*, *IEEE Int. Conf. on Networking, Sensing and Control*, 2008.

[12] Rekha A B, Umadevi B, Yogesh Solanke and Snnivasa Rao Kolli, *End to End Security for GSM Users*, *IEEE International Conference on Personal Wireless Communications*, 2005.

[13] Biancucci, Gianluigi and Claudi, Andrea and Dragoni, Aldo Franco, *Secure data and voice transmission over GSM voice channel: applications for secure communications*, *Intelligent Systems Modelling & Simulation (ISMS), 4th Int. IEEE Conf.*, pp. 230–233, 2013.

[14] Islam, Saad and Ajmal, Fatima and Ali, Salman and Zahid, Jawad and Rashdi, Adnan, *Secure end-to-end communication over GSM and PSTN networks*, *IEEE Int. Conf. on Electro/Information Technology*, pp. 323–326, 2009.

[15] Duta, Cristina-Loredana and Gheorghe, Laura and Tapus, Nicolae, *Real-time DSP Implementations of Voice Encryption Algorithms*, *ICISSP*, pp.439–446, 2017.

[16] Zhan Xu, *Data Transmission Method based on Single Carrier over GSM Voice Channel*, *Revista de la Facultad de Ingenier, vol. 32, no. 9*, 2017.

[17] Lo, Chi-Chun and Chen, Yu-Jen, *Secure communication mechanisms for GSM networks*, *IEEE Transactions on Consumer Electronics*, vol. 45, no. 4, pp. 1074–1080, 1999.

[18] Daemen, Joan and Rijmen, Vincent, *The design of Rijndael: AES-the advanced encryption standard*, *Springer Science & Business Media*, 2013.

[19] Singh G, *A study of encryption algorithms (RSA, DES, 3DES and AES) for information security*, *Computer Applications*, vol. 67, no. 19, 2013.

[20] Hankerson, Darrel and Menezes, Alfred J and Vanstone, Scott, *Guide to elliptic curve cryptography*, *Book: Springer Science & Business Media*, 2006.

[21] Ozkan Mehmet Akif and Ors Berna, *Data transmission via GSM voice channel for end to end security*, *IEEE 5th Int. Conf. on Consumer Electronics-Berlin (ICCE-Berlin)*, pp. 378–382, 2015.

[22] M. Y. Kim, N. K. Ha and S. R. Kim, *Linked Split-Vector Quantizer of LPC Parameters*, *Proc. IEEE ICASSP*, Vol. 2, pp. 741-744, 1996.

[23] Chang Chin-Chen and Hu Yu-Chen, *A fast LBG codebook training algorithm for vector quantization*, *IEEE Transactions on Consumer Electronics*, vol. 44, no. 4, pp. 1201–1208, 1998.

[24] Pal Arup Kumar and Sar Anup, *An efficient codebook initialization approach for LBG algorithm*, *arXiv preprint arXiv:1109.0090*, pp. 54–58, 2011.

[25] Lee LM, Wang H, *An extended Levinson-Durbin algorithm for the analysis of noisy autoregressive process*, *EEE Signal Processing Letters*, 3(1):13-5. 1996.

[26] Ozkan Mehmet Akif, Ors Berna and Saldamli Gokay, *Secure voice communication via GSM network*, *Electrical and Electronics Engineering (ELECO), 2011 7th Int. Conf. on*, pp 282-288, 2011.

[27] Rashidi Mahsa, Sayadiyan Abolghasem and Mowlaee Pejman, *Data mapping onto speech-like signal to transmission over the GSM voice channel*, *40th IEEE Symp on System Theory*, pp. 54–58, 2008.