Secure Voice Communication via GSM Network

Mehmet Akif Ozkan[1], Berna Ors[1] and Gokay Saldamli[2]
[1]Istanbul Technical University, Turkey
akif.ozkan@itu.edu.tr , siddika.ors@itu.edu.tr
[2]Bogazici University, Turkey
gokay.saldamli@boun.edu.tr

***Abstract*—In this study, a system is developed which communicates through GSM mobile phones and provides protection for interviews against third parties including with service providers developed. GSM line is sensitive to human speech to be more efficient and provide more quality for transmission. In addition, a tool should be used to compress speech to transmit speech over GSM. For these reasons, speech cannot be transmitted to the GSM line directly after encrypted. In this study, the encrypted speech which is a digital data stream, formed speech like waveform by the designed coder to transmit through the GSM line.**

FPGA implementation of AES is used for encryption of digital data stream. Desired speech characteristics are obtained by scanning the database of NTIMIT, and then LBG algorithm is used to design codebooks which include speech parameters. A coder is designed to synthesize speech like waveforms from the encrypted digital data stream.

I. Introduction

With developments in Global System for Mobile Communications (GSM) technology, communication has become possible to almost every point in the world without the need of a wired connection. Since telephone speeches can have content in each area of life, including personal information to be kept confidential until the state information to be kept secret. For this reason, security of GSM communication is very important.

Public Switched Telephone Network (PSTN), is the network of the world's public circuit switched telephone networks which allows any telephone in the world to communicate with any other. Interconnection of speech signals in radio frequency is provided by core circuit switched networks as GSM-GSM or GSM-PSTN. In GSM communication, speech signal is transmitted as encrypted to the radio access channel but across the core circuit switched networks voice is transmitted as a form of PCM or ADPCM [1]. Moreover encryption is optional and in control of network operators [1]. To provide end to end security, speech signal have to be encrypted before the GSM communication.

In this paper, a system which is based on paper [1], [3] is developed for protection of GSM communication against third party including network operator. Firstly, speech signal is digitized and the bit streams are encrypted. Secondly, encrypted bit streams are coded as synthetic speech signals. Finally, speech signals are transmitted to the GSM coder and GSM line. After the receiver gets the speech signal, synthetic speeches are decoded to bit streams and decrypted to get original bit stream.

For the security of digital data 128 bit Advanced Encryption Standard (AES) is used. To decrease delay AES-128 is designed as hardware for Field Programmable Gate Array (FPGA) board. It is clear that a codec module should be designed to produce synthetic speech from digital bit stream. The codec module produces synthetic speech using codebooks, which include speech parameters, by linear predictive coding method. Using 13 kbps GSM Full Rate (GSM-FR) low bit rate vocoder 06.10 source code, system is tested in computer environment without the line effects.

In second chapter the scheme of designed system, in third chapter module structure, in fourth chapter designed AES hardware is described.

II. Secure Voice Communication over GSM

For secure communication speech could be transmitted by using GSM voice channel, GSM data channel or 3G channels. GSM data channel which is used for message transmission is restricted and has bigger delay process so that getting more capacity and speed could be achieved with sending data over voice channel [3]. Although 3G channel is more convenient to transmit data it is a new technology, yet uncertain and not spread as GSM [1].

To use GSM channel more efficient speech is compressed before entering the access channel by GSM codec. Speech is coded secondly at the voice channel and this is called tandeming.

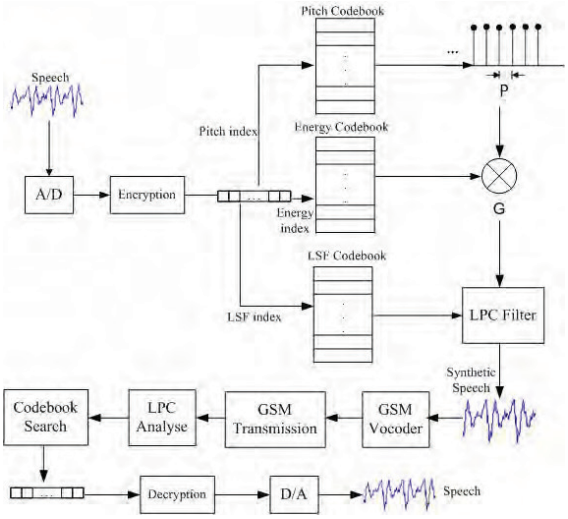

Fig. 1. Overall Scheme for End to End Secure GSM Communication

GSM channel is speech sensitive and suppresses other forms of signals. To decrypt data correctly, transmitted and received bit streams must be same. Data transmission can be achieved if transmitted symbols are understood correctly when received.

Codec module uses LPC model to synthesize speech like signals, which don't change more than acceptable level when coded by GSM-FR codec. While communication bit streams indicates index numbers of codebooks including energy, vocal filter coefficients and pitch parameters of speech. Codec module produces speech and transmitted to the GSM codec. After

received speech signals are analyzed and parameters are searched in codebooks, index numbers are extracted and bit stream is produced.

After desired properties of speech parameters are found, they are searched in NTIMIT database. Then codebooks are designed by Linde-Buzo-Gray (LBG) algorithm so as to find the most distributed codewords and provide error correction.

III. Design of the Codec Module

GSM-FR is 13 kbps low bit rate vocoder design algorithm with 8 khz sampling frequency and standardized by European Telecommunications Standards Institute (ETSI) for mobile phones. It is based on Regular Pulse Excitation-Long Term Prediction (RPE-LTP) compression algorithm which uses Code Excited Linear Prediction (CELP) method. CELP is a model based lossy compression method and compressed speech signal is not the same sample by sample with the original signal but perceptually resembled. For these reasons codec module should synthesize speech by model based low bit rate algorithm which is based on LPC method.

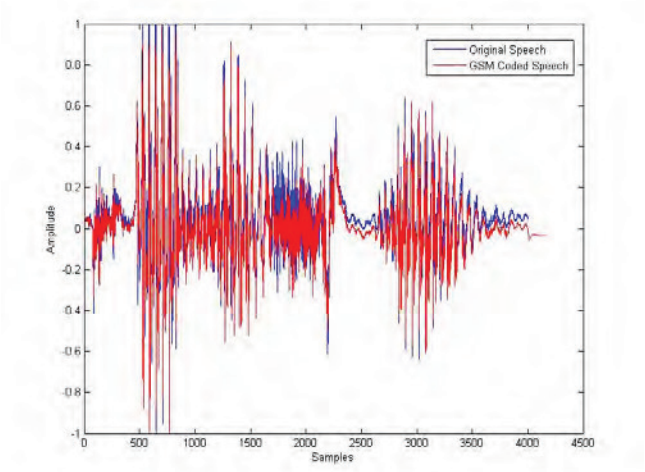

Fig. 2. Original Speech and GSM Coded Speech

While speech is coded by GSM codec, unvoiced speech is modeled with white noise and voiced speech is coded with impulse train. After experimental results with this knowledge, it was decided that the symbols which don't change after coded by GSM should be voiced speech signals.

20 ms hamming windowed voiced speech is coded with GSM codec. Vocal filter frequency responses and their waveforms are compared in Fig. 5-6. LPC coefficients, logarithms of energies and pitch frequencies are given in Table I.

As observed in Table I, Line Spectrum frequencies (LSF) of LPC coefficients and nominal energy give better results for coding. Energy is observed logarithmic for better sensitivity.

A. Normalized Energy

At the LPC algorithm, energy of excitation signal and error signal are the same and gain is found with this equation. Nominal mean squared error is the ratio of energy of error signal and analyzed signal. The formula of the nominal energy in autocorrelation coefficients is given in the equation below [6].

V_n: Nominal Energy;

TABLE I
Changes in model parameters for an example voiced speech signal after coded with GSM

	Analyzed Speech		Speech coded with GSM	
Pitch Period	59 sample		59 sample	
LPC Coefficients	-2.1645	1.3734	-2.2109	1.3781
	-0.5692	0.7886	-0.3141	0.5654
	0.1465	-0.6921	-0.0512	-0.4917
	-0.2131	0.1596	-0.2163	0.4587
	0.4616	-0.2602	-0.0218	-0.0639
LSF Parameters (rad)	0.1888	0.2785	0.1883	0.2816
	0.4215	0.5062	0.4209	0.5049
	0.8091	1.6516	0.8124	1.5144
	1.7357	1.9526	1.7734	1.9756
	2.3579	2.7669	2.4102	2.7515
Log. of Energy	-1.6544		-3.6290	
Log. of Nom. Energy	-110.3271		-112.8483	

e_n: Error signal of LPC;
s_n: Analyzed signal of LPC
α_i: LPC coefficient;
$r_n(i)$: Autocorrelation coefficient
$V_n = \frac{\sum_{m=0}^{N+p-1} e_n(m)^2}{\sum_{m=0}^{N+p-1} s_n(m)^2} = -\sum_{i=0}^{p} \alpha_i \frac{r_n(i)}{r_n(0)}$;

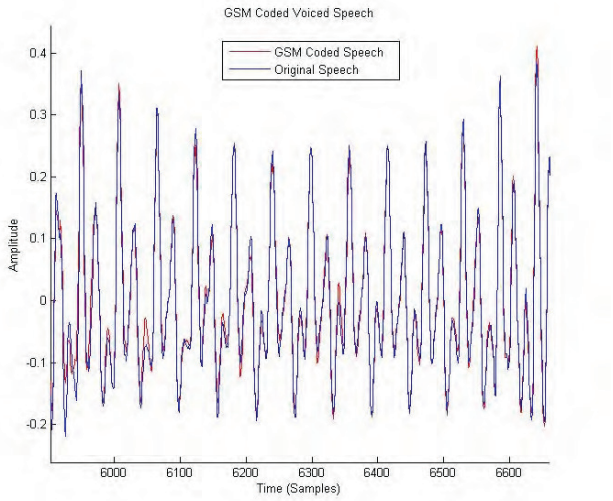

Fig. 3. GSM Coded Voiced Speech Frame

B. Line Spectrum Frequencies

Line Spectrum Frequencies (LSF) are a different representation of LPC filter coefficients. Two polynomials of LPC coefficients can be found with the below equations bt assuming p is the order of LPC coefficients.

$P(z) = A(z) + z^{-(p+1)}A(z)^{-1}$
$Q(z) = A(z) - z^{-(p+1)}A(z)^{-1}$

The roots of $P(z)$ and $Q(z)$ polynomials are interlaced in the unit circle and lie in complex conjugate pairs. LSFs can be computed by taking the $p/2$ different roots.

LSFs are located in ascending order thus provide a sufficient method for checking stability. LSFs are plotted as vertical lines on spectrum of the spectrum of LPC filter and related closely with the formant frequencies.

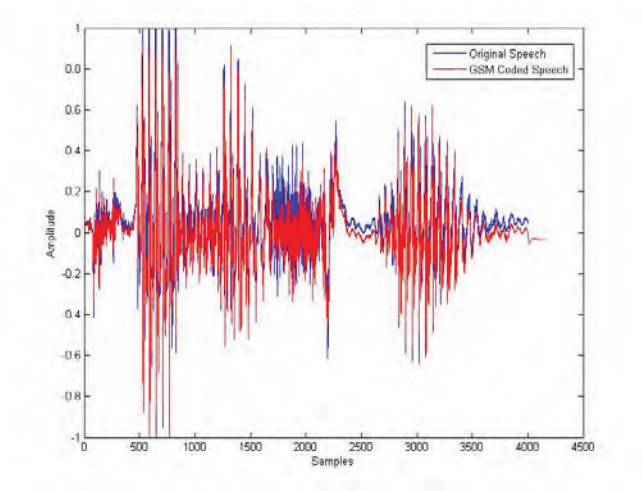

Fig. 4. GSM Coded Unvoiced Speech Frame

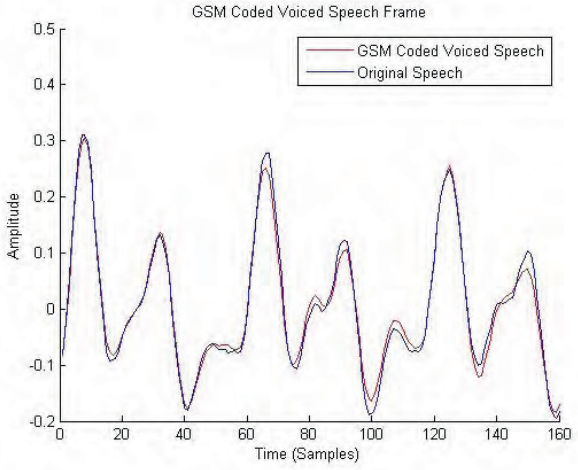

Fig. 5. Original and GSM Coded Voiced Speech Frame

LPC coefficients can easily be found from LSFs with the equation as follows, $A(z) = \frac{P(z)+Q(z)}{2}$.

C. Design of Codebooks

Codebooks of energy and pitch period are produced with scalar vector quantization. Logarithm of nominal energy is coded with 2 bit; pitch period is coded with 4 bit.

LSFs for voiced speech signals are searched from NTIMIT voice database after designing a voiced activity detector based on below voiced speech properties. LSF codebooks are designed with the splitting vector quantization method using LBG algorithm. Splitting vector quantization is a method which uses multiple codebooks instead of one so producing less codeword and having wider region for codeword provide better quantization performance. Two LSF codebooks are designed for ten order LPC algorithm. One of the codebooks is designed for first four LSFs; another is designed for six LSFs and they are coded with 5 bits.

1) Design of Voiced Activity Detector: Voiced speech signals have bigger amplitudes and energy levels than unvoiced speech signals. Although voiced speech is periodic, unvoiced speech is

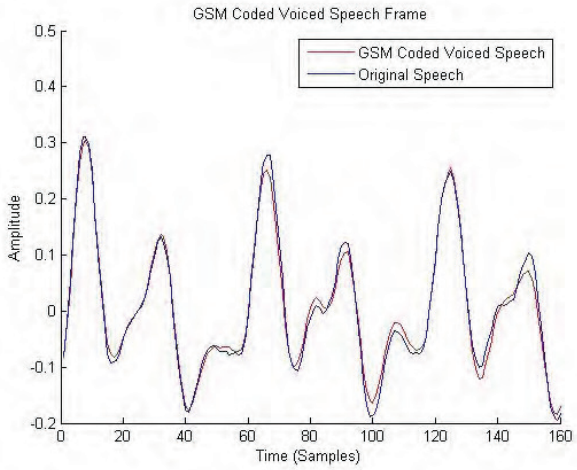

Fig. 6. Spectrum of LPC Filters

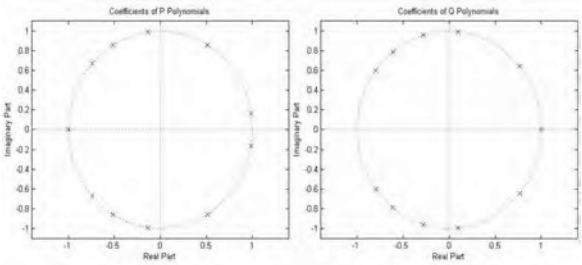

Fig. 7. Roots of $P(z)$ and $Q(z)$ polynomials

random noise structured. A voiced activity detector is designed to distinguish voiced speech frames from unvoiced ones with looking into periodicity, low band to full band ratio, autocorrelation coefficient at simple unit and zero crossing rate [6], [2].

Voiced speech has bigger density of energy at low frequencies. Energy ratio of voiced speech signal being smaller frequency from 1 kHz to full band energy is almost 1 where unvoiced speech has much lower ratio.

Low band to Full band Ratio= $\frac{\sum_{i=1}^{N} s_{lpf}(i)^2}{\sum_{i=1}^{N} s(i)^2}$

Resemblance of signal at a simple unit is very high at voiced speech unlike unvoiced speech which has random values. So normalized first autocorrelation coefficient of voiced speech is almost 1 where that of unvoiced speech is generally negative.

Normalized first Autocorrelation Coeff= $\frac{\sum_{i=1}^{N} s(i)s(i-1)}{\sum_{i=1}^{N} s(i)^2}$

Human voiced speech has a periodicity with a frequency between 50 Hz and 200 Hz and if pitch frequency of the signal is out of this border it is decided as unvoiced speech.

Since voiced speech has lower frequencies, it crosses from zero much lower than unvoiced speech. Zero crossing rate is lower than 3 kHz unlike unvoiced speech.

2) Linde-Buzo-Gray Algorithm: Linde-Buzo-Gray (LBG) algorithm provides an efficient way for vector quantization. Vector quantization is a lossy data compression algorithm based on the principle of block coding. It produces codebooks and each block is considered as a vector which is called codeword. Algorithm changes original data with index numbers referring codewords for compression. LBG algorithm searches for the most separated codewords for an efficient vector quantization.

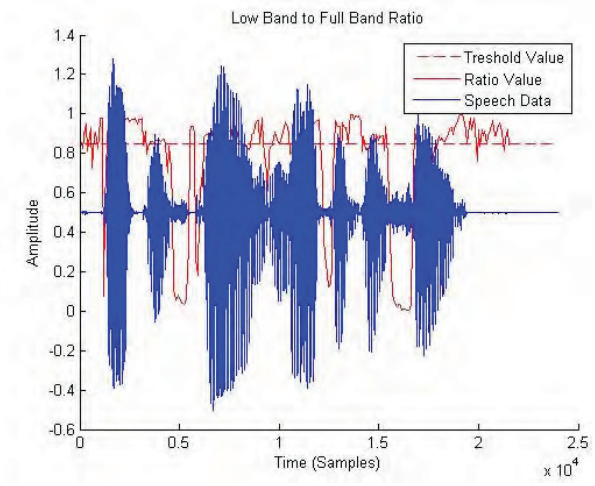

Fig. 8. Voiced Classification of Speech According to Low Band to Full Band Ratio

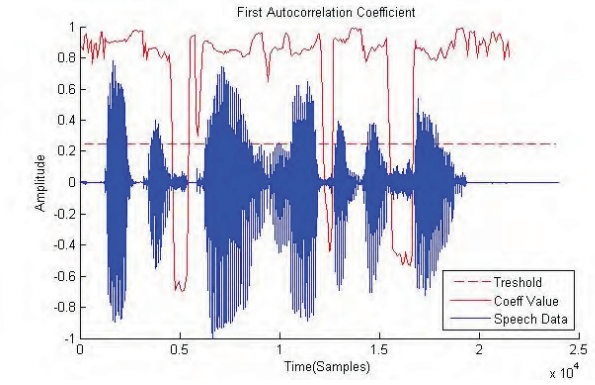

Fig. 9. Voiced Classification of Speech According to Normalized First Autocorrelation Coefficient with threshold of 0.25

A thousand of two dimensional random vectors are quantized to sixteen clusters with splitting LBG algorithm in Fig. 10. Centers of each cluster are the codewords. Sixteen codewords are coded with 4 bit.

Splitting LBG algorithm gives better results because no starting cluster information is needed. LSF codebooks are produced by following below steps using Euclidean formula as distance measure [4].

Common measures for algorithm:

Training Vectors: $B = \{B_1, B_2, \ldots, B_m\}$

Codewords: $C = \{C_1, C_2, \ldots, C_m\}$

$B_i = \{b_1, b_2, \ldots, b_l\}$, $C_i = \{c_1, c_2, \ldots, c_l\}$; b_i and c_i are real numbers.

Euclidean Distance: $ED\left(B_i, C_j\right) = \sum_{k=1}^{l} (B_k - C_k)^2$.

There are two main processes at the algorithm which are splitting and iteration [4].

Step 1: At the start of algorithm there is one partition including all training vectors.

Number of codeword: $N = 1$

Distortion: $D^{(0)} = 0$

Set a very small value: $\epsilon > 0$

Initial codeword for the cluster: $C_1 = \frac{1}{m} \sum_{k=1}^{m} B_k$

Step 2: This is the splitting process and numbers of partitions are multiplied by two.

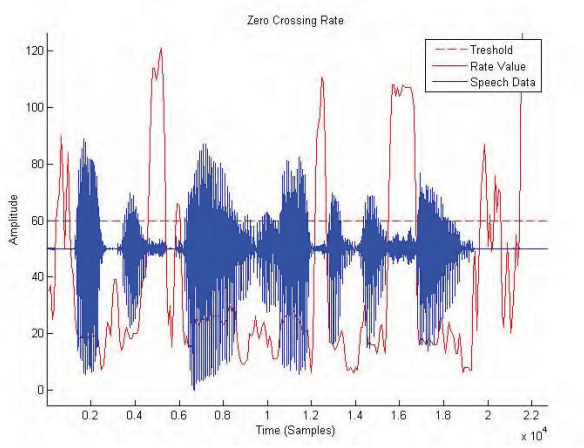

Fig. 10. Voiced Classification of Speech According to Zero Crossing Rate with threshold of 3kHz

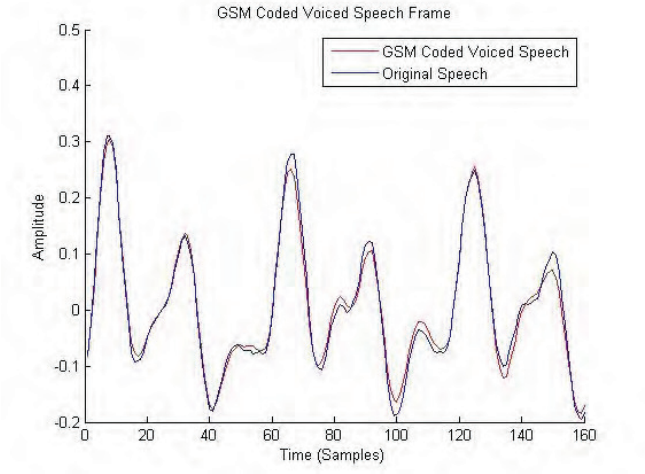

Fig. 11. Codebook Design with LBG Algorithm

For $i = 1, 2, \ldots, N$:

$C_i^{(k)} = 1 + \epsilon * C_i^{(k)}$, $C_{N+i}^{(}k) = 1 - \epsilon * C_i^{(k)}$, $N = 2N$

Step 3: This is the start of iteration process in which codewords are found for each partition. Algorithm searches for each training vector for the closest codeword and sets these training vectors as members of partitions of the found codewords.

Assuming elements of P as partitions of codewords and k is the iteration number:

$P_i^{(k)} = B_n$, $d\left(B_n, C_i^{(k)}\right) < d\left(B_n, C_i^{(k)}\right)$, $j \neq i$, $i = 1, 2, \ldots, M$

Step 4: New codewords are found by calculating center of mass for each partition.

For $i = 1, 2, \ldots, N$

$C_i = \frac{1}{length\left(P_i^{(k)}\right)} \sum_{k=1}^{length\left(P_i^{(k)}\right)} B_k$

Step 5: Distortion is calculated for new codebook.

$D^{(k)} = \frac{1}{m} \sum_{i=1}^{N} D_i$, $D_i = \sum_{j=1}^{length\left(P_i^{(k)}\right)} \left(C_j - B_j\right)^2$

Step 6: If distortion did not get a smaller value or equation below is true for a small value of δ algorithm goes to Step 7.

$\frac{D^{(k-1)} - D^{(k)}}{D^{(k)}} \leq \delta$

Otherwise, after iteration number is increased go to Step 3.

$k = k + 1$

Step 7: Iteration is finished. If partition number is enough

algorithm ends after codewords are updated with the below formula, else it goes to Step 2.

$C_i = C_i^{(k)}$.

3) Improving the Codebooks for A Better Performance: GSM codec synthesize speech with its quantization codebooks at quantization values, after speech is analyzed and modeled. So if speech compressed after analyzed recursively, after sufficient iteration its LSFs are expected not to change. To improve symbols, codec module values of LSF codebooks are decoded with LPC method then encoded and decoded by GSM codec. After sufficient iteration codebooks are modified. Figure 12 shows the frequency response of LPC filter and Table II shows LSF parameters after coding speech signal multiple times. After LSF parameters updated for the GSM CODEC, transmission change decreased to very small numbers. So symbols which are used for transmission become more trustable for data communication.

IV. Design of Encryption Hardware Using Advanced Encryption Standard

To secure communication, digital data encrypt with AES-128 algorithm which is designed as hardware for FPGA board. A recursive process in an algorithm is called round. The algorithm has ten rounds consisting two parts which are encryption and decryption. These parts have four basic modules which are Sub Bytes, Shift Rows, Mix Columns, Add Round Key and their inverses. Key is updated in every round by key producer [5]. AES is used in Electronic Code Book (ECB) mode.

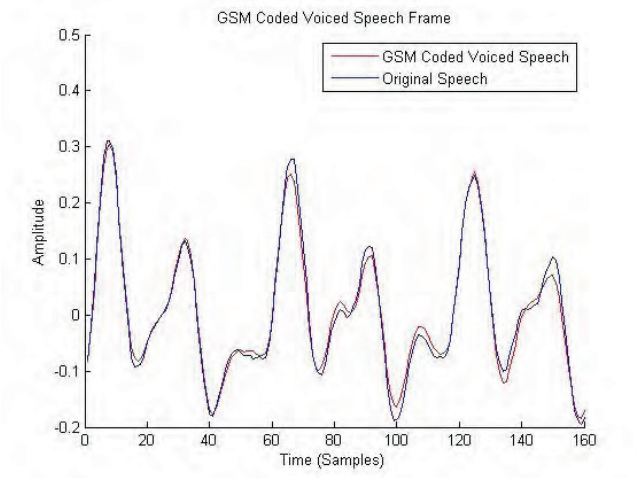

Fig. 12. Signal Coded with GSM Multiple Times

TABLE II
LSFs of speech signals after coded with GSM codec multiple times

	LSFs			
Speech Coded with GSM 20 times	0.1888	0.2859	0.4506	0.6994
	1.4886	1.8211	1.8826	2.2529
		2.4056	2.6704	
Speech Coded with GSM 19 times	0.1886	0.2859	0.4561	0.7060
	1.4823	1.8205	1.8866	2.2493
		2.4087	2.6715	

Hardware synthesis reports which are designed in Xilinx Spartan 3E is given in Table III. It shows that AES hardware is fast enough for speech communication.

TABLE III
Synthesis Reports of AES-128 Hardware

	Encryption	Decryption
# of Slices	4162	2722
# of Slice FFs	2083	475
# of LUTs	7931	5031
# of BRAMs	20	0
# of GCLKs	1	1
Min. Period	10.515 ns	10.389 ns
Max. Clock Freq.	95.098 MHz	96.256 MHz

V. Conclusions

In this paper a system for secure end to end voice communication via GSM network is designed and simulated. In our system the speech is converted to digital bit stream, encrypted and then encrypted digital data is converted to speech like waveform before given to 13 kpbs GSM FR codec at the sender side. At the receiver side the digital bit stream is recovered back by analyzing the received speech like waveform.

Because the digital voice data is sent encrypted, none of the bits should be transmitted incorrectly. If that is the case the original message can not be recovered back. Any signal whose characteristic is different from speech is filtered out in low bit rate GSM network. Hence the challenge is finding the best method to convert the digital data to speech like waveform and vise versa without any error.

AES algorithm is implemented on an FPGA for the encryption of the speech after converted to digital bit stream. The characteristics that the speech like waveform should satisfy in order to be transmitted correctly are decided. Then the corresponding LSF parameters are found by scanning the NTIMIT voice database. After having the parameters, the codec designed by using the LBG algorithm is made usable. The codec implemented by MATLAB and the FPGA implementation of the encryption algorithm is communicated via serial port. The GSM network is also modeled in MATLAB. If we ignore the wireless communication errors in GSM network, we have shown that the system is working correctly.

Acknowledgments

Note that, Gokay Saldamli is partially funded by TUBITAK research project No: 109E180 and Berna Ors is partially funded by TUBITAK research project No: 110E172.

References

[1] N. N. Katugampala, K.T. Al-Naimi, S. Villette, and A. Kondoz. Real time data transmission over GSM voice channel for secure voice and data applications. In *Proceedings of the 2nd IEE Secure Mobile Communications Forum: Exploring the Technical Challenges in Secure GSM and WLAN*, London, UK, September 2004.

[2] A. M. Kondoz. *Digital Speech:Coding for Low Bit Rate Communication Systems*. John Wiley & Sons, 2nd edition, 2004.

[3] C. K. LaDue, V. V. Sapozhnykov, and K. S. Fienberg. A data modem for GSM voice channel. *IEEE Transactions on Vehicular Technology*, 57(4):2205–2218, July 2008.

[4] Y. Linde, A. Buzo, and R. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1):84 – 95, January 1980.

[5] National Institute of Standards and Technology. FIPS 197: Advanced Encryption Standard, November 2001.

[6] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Signal Processing Series. Prentice-Hall, 1978.