Data Transmission via GSM Voice Channel for End to End Security

Mehmet Akif Özkan
Istanbul Medeniyet University
akif.ozkan@itu.edu.tr

S. Berna Örs
Istanbul Technical University
siddika.ors@itu.edu.tr

***Abstract*—Global System for Mobile Communications (GSM) technology still plays a key role because of its availability, reliability and robustness. Recently, additional consumer applications are proposed in which GSM is used as a backup or data transmission service. Unfortunately sending data via GSM channel is a challenging task since it is speech sensitive and suppresses other forms of signals. In this paper, a systematic method is proposed to develop a modem that transmits data over GSM voice channel (DoGSMV) using speech like (SL) symbols. Unlike the previous approaches an artificial search space is produced to find best SL symbols and analyses by synthesis (AbyS) method is introduced for parameter decoding. As a result 1.6 kbps simulation data rate is achieved when wireless communication errors are ignored.**

***Keywords*—Data over GSM voice, Secure GSM communication, GSM Full Rate, Speech like signal**

I. Introduction

GSM still protects its popularity with recent consumer applications because of its wide service availability in all over the world. Even now it is the only communication service in many developing countries and rural areas. In addition to that GSM voice dedicated channel has high priority, low latency and better quality of service (QoS) comparing to data channels. These advantages make GSM voice channel attractive as a data transmission service or a backup service for mission critical applications even when higher data rates are offered with recent technologies such as 3G, 4G LTE.

Security of GSM communication is vital for many life scenerios since it can have a content in each area of life, including personal information to be kept confidential until the governmental information to be kept secret. However GSM network still has serious vulnerabilities [1]. The content of a GSM communication is only encrypted between the mobile phone and mobile station (MS), while the security of transmission through the network only depends on service providers. Furthermore lightweight A5 algorithms, to which a lot of vulnerabilities have been reported [1], are used for the encryption. A solution is to encrypt and decrypt speech signals before and after the GSM communication to provide end-to-end security [2]–[4] against third party including the network operator. However a special modem is necessary to transmit data over GSM voice channel as shown in Figure 1.

Data transmission over GSM voice channel (DoGSMV) is suggested for many other applications recently. It is proposed to be used in Point of Sale (POS) payment systems [5] and wireless automatic teller machines (ATM) [6]. ETSI's (European Telecommunications Standards Institute) emergency call

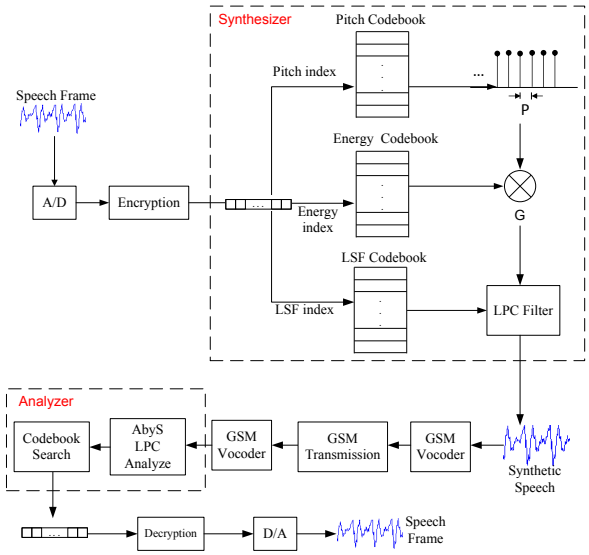

Fig. 1. The structure of DoGSMV modem

(eCall) system for road safety monitoring is a mission critical example. It consists of an in-car system that calls the closest emergency center and sends the details in case of an accident. Another important example is the pedestrian guidance of WikiWalk project [7], in which geographical information is sent and guidance messages are received through GSM voice channel. A different usage is the network address translation (NAT) [8] that establishes a direct peer to peer communication between mobile users without a middle server. Covert data communication [9], in which data is hidden into speech signals, is an interesting application. Telemetry, automated meter reading, alarm systems, vehicle tracking/fleet management, security systems are foreseen as other applications [6].

DoGSMV is a challenging task since GSM channel is speech sensitive and suppresses other forms of signals. In addition several operations are applied to use the bandwidth efficiently. The voice channel has a narrow band with maximum bandwidth of 4 kHz and speech is compressed with GSM codec before it is transmitted. Then voice activity detection (VAD), discontinuous transmission (DTX) and comfort noise generation (CNG) are applied in the transmission line so that only voiced parts are carried. Furthermore compression rate is adjusted adaptively according to cellular network traffic.

Previous approaches for DoGSMV needs a special modem before and after GSM network as shown in Figure 1. The

DoGSMV modem converts digital bit stream to special signals that can be translated back after they are compressed, transmitted and decompressed via GSM. Related work can be investigated in three group [6], [10]. First group includes methods that use digital modulation techniques [5], [7], [11], [12] to produce the special signals. Second group uses predefined symbols [6], [10], [13]. Third group maps digital bit stream to speech parameters and synthesize speech-like (SL) signals [14], [15]. Although producing real speech signals have significant advantages over modulation methods to pass voice activity detectors (VADs), previous SL symbol based methods are either very slow or not fully explained.

In this paper we proposed a systematic method to transmit digital bit streams using speech like (SL) symbols. We introduced analyses by synthesis (AbyS) for parameter decoding and designed the modem by selecting best performing SL symbols after a convenient artificial database is produced. As a result performance increased from 0.6 kbps to 1.6 kbps data rate for GSM FR.

II. Structure of DoGSMV Modem

We focused on GSM Full Rate (GSM-FR) [16] in this work since it is one of the widely used standards for mobile phones. Its sampling frequency is 8 kHz and data rate is 13 kbps. Its compression algorithm is Regular Pulse Excitation-Long Term Prediction (RPE-LTP), which is a type of Code Excited Linear Prediction (CELP).

CELP is a lossy compression based on Linear Predictive Coding (LPC) and analyses by synthesis (AbyS) methods. Input signal is divided to short frames and vocal properties of speech are analyzed with LPC. The error in AbyS is the difference between the input signal and the synthetic speech that is produced from analyzed parameters. Speech parameters of input are extracted with LPC and error is calculated with AbyS in a CELP based compression. An iteration is run until a small error is reached. A detailed explanation can be found in [17].

Considering the structure of CELP, we chose LPC method to produce SL symbols from the basic parameters, which are Line Spectrum Frequencies (LSF), pitch and energy, as shown in Figure 2. The DoGSMV modem consists of look up tables, which are called codebooks, to store the parameters of selected SL signals. Input bit stream is used as a set of index numbers to the codebooks. The biggest challenge is that the same exact parameters cannot be found even when the synthesized signal is analyzed with no change. Furthermore additional distortions to produced signal come from the (de)compression of GSM-FR. Therefore we used a clever parameter extraction method based on AbyS for the decoder part.

SL symbols are selected from voiced frames because of two reasons: 1) Not to trigger VADs of GSM. 2) Distortion coming from GSM (de)compression is much bigger for unvoiced speech frames [4]. On the other hand silence intervals should be placed between voiced frames to pass VADs.

Design of the DoGSMV modem has three stages. First, an objective function is determined to measure the performance. Second, quantization tables for the parameters of the SL symbols are determined from the properties and requirements

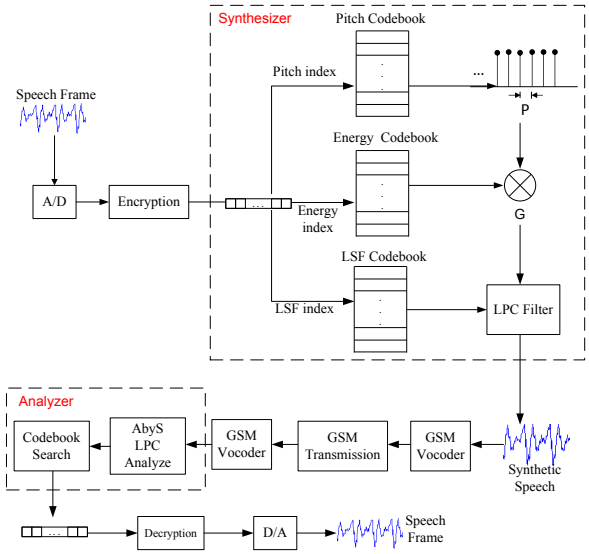

Fig. 2. The structure of DoGSMV modem

of GSM channel. Then a parameter pool is created from the determined combinations and a synthetic SL symbol database is produced as a search space. Finally, best SL symbols are selected according to the objective function and the codebooks of DoGSMV modem are designed.

A. Objective Function

Symbol Error Rate (SER) is chosen for the cost function. DoGSMV modem is designed in two steps to have most fault tolerant codebook and decoding scheme. In the first step, euclidean distances of codewords are maximized in order to produce most distinguished SL symbols for different inputs. In the second step, error correction blocks are designed for the decoding part. Conditional probability of $P(IN(Y_i) = i|X_i)$ is maximized where $IN(Y_i)$ is the codeword index of GSM channel output for the input of X_i signal.

B. Search space

Optimizing the cost function in a speech database has two advantages: 1) Finding SL symbols. 2) Reducing the search space to a subset most probably provides a global minimum for the cost function. Previous works targets TIMIT speech database as the search space [4], [10], [13] for the SL symbol selection. Then they applied several pre-processing operations are applied such as voiced symbol selection, power normalization and similar symbol election before the optimization of cost function.

In this paper, a novel approach is proposed. An artificial database is created using LPC from the constraints of GSM instead of filtering a speech database. Since LPC uses IIR filter for speech production, producing a database consisting stable sets of LPC parameters is a challenge. We overcome this problem by using Line Spectrum Frequencies (LSF) instead of LPC parameters. LSF is a transformation of LPC that provides better performance in quantization [17]. A monotonically increasing LSF parameters, obeying to (1), assures the stability of LPC

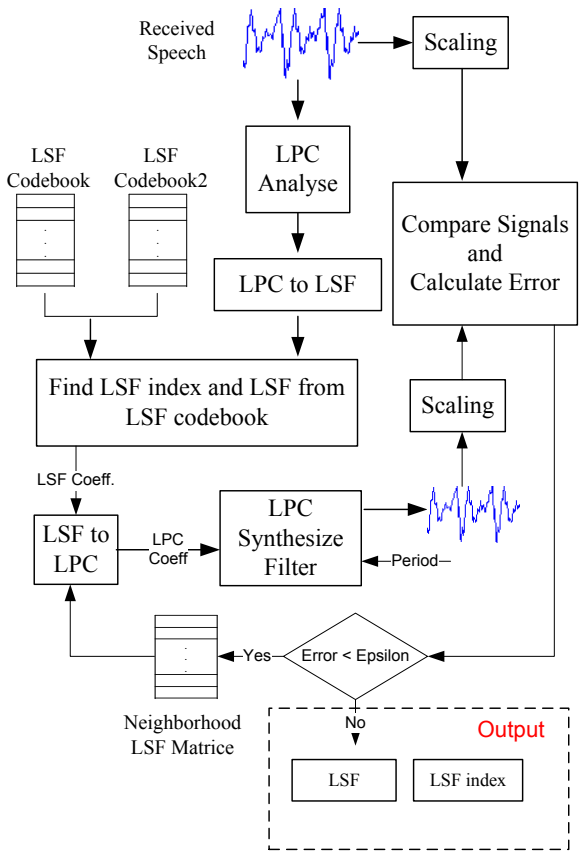

Fig. 3. LSF parameters decoding

synthesis filter.

$$LSF[i] < LSF[j] \quad \text{for } i < j \text{ and } i, j \in [1, 10] \qquad (1)$$

The main advantage of the proposed method is that we only introduced the SL signals showing good performances in GSM (de)coding. We made sure that speech signals produced from the parameters of the quantization levels of GSM are in the database. In addition, we included outputs of GSM after a signal is repeatedly coded as explained in paper [4].

C. LSF Parameter (De)Coding

LSF codebook of DoGSMV modem is designed to map 10 bits of digital data. It is experimented that 10 LSF parameters provide a good performance for 10/20 ms frames. Splitted vector quantization method is used with euclidean distance. (4, 6) splitting of LSF parameters is reported to give best performance [17]. Therefore two codebooks having $2^5 + 2^5$ codewords are designed for the sets of 4 LSFs and 6 LSFs instead of a one big 2^{10} sized codebook.

All of the LSF combinations that are produced from splitted codebooks should be monotonically increasing to preserve the stability. This means that the maximum codeword of 4 LSF codebook is smaller than the minimum codeword of 6 LSF codebook. Mathematical explanation is given in (2).

$$K| \begin{cases} K > LSF[i] & \text{for } i \leq 4 \\ K < LSF[i] & \text{for } i \geq 4 \end{cases} \qquad (2)$$

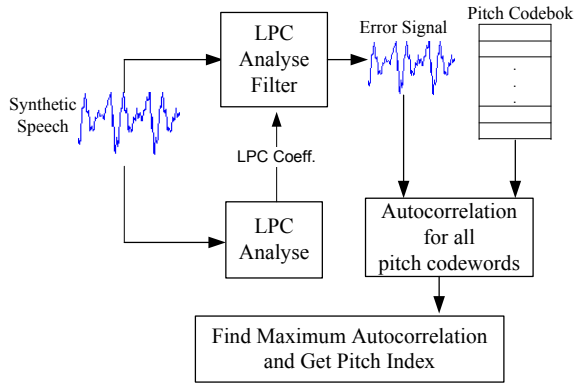

Fig. 4. Pitch parameter decoding

Cost of LSF codebooks are minimized with an iteration of two steps:

1) K in (2) is selected and all the combinations of 4/6 LSF codewords are created from the quantization table of CELP [17].
2) Most distinct codewords are found using LBG (Linde-Buzo-Gray) algorithm and Codebooks are updated when the minimum cost is reached.

As a result optimal K is selected and best LSF database is created. In the second step, LBG is replicated with many different set of initial vectors to avoid local minimum. Mean shift algorithm can also be used in here. These clustering algorithms are modified to force selected codeword to be an element of its search database.

LSF decoder structure of DoGSMV modem is given in Figure 3. AbyS method is used with a concept called Neighborhood matrix (NM). NM consists of a set of indexes in its ith row in order to address most probable symbols that can be selected instead of codeword i after GSM coding. The term "nearest neighbor" is used for the symbol having the biggest probability of confusion. Affordable calculation cost of AbyS can be achieved thanks to NM concept. In addition, NM can be updated according to system noise.

LSF decoder extracts the indexes in six steps:

1) Calculate LSF parameters of received signal.
2) Extract LSF indexes from codebooks.
3) Produce SL symbol from the selected LSF codewords.
4) Calculate a scaled error of produced SL symbol
5) Return the index if the error is acceptable
6) If there is a nearest neighbor, select its LSF codewords and skip to step 3. Else return the index of the codeword with minimum error.

D. Pitch Parameter De(Coding)

Frequency of the SL symbols should be in the range of human voice to trick VADs of GSM. It is observed that less distortion come for the frequencies between 50 Hz and 400 Hz in GSM coding.

Experimental results show that pitch parameter can be extracted correctly even for the GSM coded signals. On the

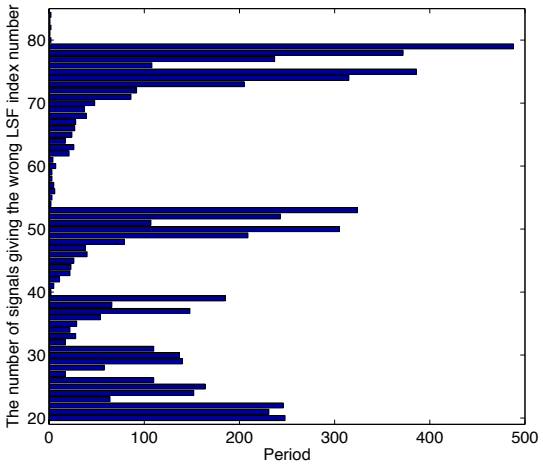

Fig. 5. Symbol error rate (SER) / pitch histogram

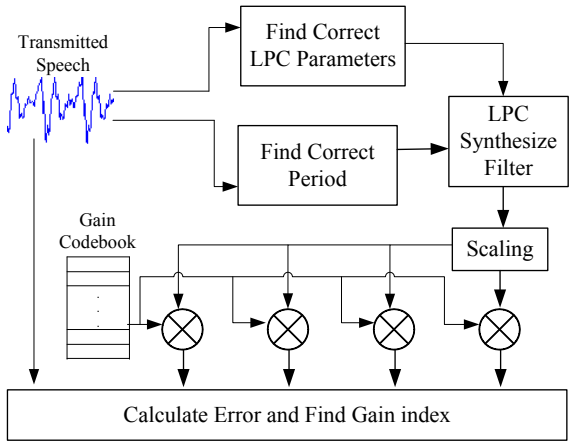

Fig. 6. Energy parameter decoding

other hand, changing the frequency affects fault tolerance to GSM coding of SL symbols differently. Therefore designing the pitch codeword with a linear division of the specified frequency range gives a poor performance. Concerning these observations, a nonlinear approach is proposed to design 4-bit pitch codebook. Proposed algorithm has three steps:

1) Search space is extended by producing SL symbols using all of the combinations of LSF codebooks and selected frequency range of pitch.
2) Symbol error rate (SER) / pitch histogram is calculated over the search space as given in Figure 5.
3) Codewords are selected from pitch values providing minimal SER and their distances are maximized.

Autocorrelation, which is a general approach, is used in the decoding part as shown in Figure 4. It has three steps:

1) Filter the received speech signal by LPC analyze using the parameters calculated at the 3rd step of LSF error correction algorithm. This is the result of first LPC analyze.
2) Calculate autocorrelation values of the error signal for all of the codewords of the pitch codebook.
3) Extract the index number of the codeword that has maximum auto-correlation result.

E. Energy Parameter De(Coding)

Energy is calculated using the error of LPC analyze since it provides a more reliable result [17]. 2 bits of energy codebook is produced with a linear approach. AbyS is used in the decoding as shown in Figure 6. Decoding has two steps:

1) Synthesized speech frame is multiplied with codewords of energy codebook.
2) Energy index is extracted by selecting the codeword providing minimum distance with the received signal.

III. Results & Discussions

GSM-FR 06.10 is selected to show the boundaries of the proposed method since it is the hardest scenario [6]. GSM-GSM communication is simulated in computer environment using SOX tool [18]. Additional VAD programs, which check low band to full band ratio, zero crossing rate and normalized first autocorrelation coefficient features to estimate the silence, are designed. 1.6 kbps data rate is reached by coding 16 bit streams on 10 ms speech frames with the proposed method.

Coherency of simulation platforms is very important for a fair comparison of the related work. While some work simulates the voice channel only with speech compression source code [2], [4], [15], results on real time implementations exist [3], [5]. Although there are certain challenges in real time implementation of the system, Katugampala et. al. proves that closed results of a simulation [2] can be achieved in real time [3]. This work proposes a more efficient algorithm to use auto-regressive speech production methods. It can also be applied in other works. Related work giving the simulation results on GSM-FR channel or using the speech production methods are given in Table I. A fair comparison can be made with [4], [15]. Bit Error Rate (BER) in a simulation system points constant design faults when only speech compression codec is implemented as in [15]. Moreover, Sapozhnykov [6] shows that vocoder selection and acceptable BER highly effects maximum achievable data rate. Even accepting a small BER rises the data rate of the same technique significantly. It can be seen that the proposed algorithm increases the data rate of [4] by 33 % when 20 ms frames are processed in both algorithms.

AbyS approach increases the performance of speech production methods significantly. A competitor in this class is Katugampala's design. A raw data rate of 3 kbps has been reported with a simulation BER of 2.9 %. Adding error correcting codes (rate 1/2 convolutional codes) yielded a throughput of 1.2 kb/s with a 0.03 % BER. However their results are on GSM EFR. Moreover, they gave no detail about their codebook design and error handling algorithms. Concerning the advantages of using speech signals instead of modulation methods to pass VADs, the results of the proposed system is promising.

IV. Conclusion

In this work a system is given to provide end to end data communication over GSM voice channel. A systematic

TABLE I. COMPARISON WITH PREVIOUS WORKS USING GSM FR.

	Vocoder	Data Rate	BER %
^{*,CE}Proposed Method	FR	1.6 (kbps)	0.00
^{*,RT}Kotnik et al. [5]	FR	0.533 (kbps)	30
	FR	0.107 (kbps)	10
^{*,CE}Ozkan et al. [4]	FR	0.600 (kbps)	0.00
^{*,CE}Katugampala et al. [2]	EFR	3 (kbps)	0.5
		4 (kbps)	4
^{*,RT}Katugampala et al. [3]	EFR	3 (kbps)	2.9
		1.2 (kbps)	0.03
^{*,RT}Sapozhnykov et al. [6] (same technique)	FR	1.4 (kbps)	0.0004
	EFR	2.1 (kbps)	0.0001
	EFR	3 (kbps)	0.3466
	AMR 12.2	3 (kbps)	3.1
	AMR 12.2	0.76 (kbps)	0.00
^{*,CE}Rashidi et al. [15]	FR	1.15 (kbps)	0.02
$^{*,-}$Peyvandi et al. [19]	FR	0.8 (kbps)	0.4

(*: Methods using autoregressive speech production)
(Simulation in ^{CE}: Computer Environment, ^{RT}: Real Time)

method is proposed to design a special data modem that transmits digital bit streams using speech like (SL) symbols. The modem converts SL symbols to digital data correctly after it is compressed and uncompressed with the vocoders of GSM. The system is simulated using GSM-FR 06.10 vocoder and VAD programs.

SL symbols are produced from LSF, pitch and energy properties of speech using splitted vector quantization with two main contributions. First, analyses by synthesis (AbyS) method of CELP compression is used in the decoding part. Second, cost function of the codebook design is optimized in artificial databases that are created from the quantization levels of the selected GSM vocoder. A significant increase in performance is observed with a 1.6 kbps data rate.

Furthermore design of the modem is defined as a speech identification problem. As a future work decoding part will be designed with the identification methods such as hidden markov models and it will be applied to other GSM voice channels, such as EFR, AMR, HR. In addition, synchronization algorithms will be investigated and the proposed method will be tested in real time.

REFERENCES

[1] G. Cattaneo, G. De Maio, and U. F. Petrillo, "Security issues and attacks on the gsm standard: a review," *Journal of Universal Computer Science*, vol. 19, no. 16, pp. 2437–2452, 2013.

[2] N. Katugampala, S. Villette, and A. M. Kondoz, "Secure voice over gsm and other low bit rate systems," 2003.

[3] N. N. Katugampala, K. T. Al-Naimi, S. Villette, and A. M. Kondoz, "Real time data transmission over gsm voice channel for secure voice & data applications," *IET Conference Proc.*, 2004.

[4] M. A. Ozkan, B. Ors, and G. Saldamli, "Secure voice communication via gsm network," in *IEEE International Conference on Electrical and Electronics Engineering (ELECO)*, 2011.

[5] B. Kotnik, Z. Mezgec, J. Svečko, and A. Chowdhury, "Data transmission over gsm voice channel using digital modulation technique based on autoregressive modeling of speech production," *Digital Signal Processing*, vol. 19, no. 4, 2009.

[6] V. V. Sapozhnykov and K. S. Fienberg, "A low-rate data transfer technique for compressed voice channels," *Journal of Signal Processing Systems*, vol. 68, no. 2, 2012.

[7] B. T. Ali *et al.*, "Data transmission over mobile voice channel based on m-fsk modulation," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.

[8] A. Patro, G. Baudoin, and O. Venard, "A system for audio signalling based nat traversal," in *IEEE International Conference on Communication Systems and Networks (COMSNETS)*, 2011.

[9] M. Boloursaz, R. Kazemi, F. Behnia, and M. A. Akhaee, "Performance improvement of spread spectrum additive data hiding over codec-distorted voice channels," in *IEEE European Signal Processing Conf. (EUSIPCO)*, 2014, pp. 2510–2514.

[10] M. Boloursaz, A. H. Hadavi, R. Kazemi, and F. Behnia, "A data modem for gsm adaptive multi rate voice channel," in *IEEE East West Design and Test Symposium*, 2013.

[11] T. Chmayssani and G. Baudoin, "Data transmission over voice dedicated channels using digital modulations," in *IEEE International Conference Radioelektronika*, 2008.

[12] A. Dhananjay, A. Sharma, M. Paik, J. Chen, T. K. Kuppusamy, J. Li, and L. Subramanian, "Hermes: data transmission over unknown voice channels," in *ACM International Conference on Mobile Computing and Networking*, 2010.

[13] A. Shahbazi, A. H. Rezaei, A. Sayadiyan, and S. Mosayyebpour, "Data transmission over gsm adaptive multi rate voice channel using speech-like symbols," in *IEEE International Conference on Signal Acquisition and Processing (ICSAP'10)*, 2010, pp. 63–67.

[14] N. N. Katugampala, K. T. Al-Naimi, S. Villette, and A. M. Kondoz, "Real time end to end secure voice communications over gsm voice channel," in *IEEE European Signal Processing Conference (EUSIPCO)*, 2005, pp. 1–4.

[15] M. Rashidi, A. Sayadiyan, and P. Mowla, "Data mapping onto speech-like signal to transmission over the gsm voice channel," in *IEEE Southeastern Symposium on System Theory*, 2008.

[16] E. S. T. Series, "Etsi standard gsm 06.10," 1997, eTSI SMG 2 Group.

[17] A. M. Kondoz, *Digital speech: coding for low bit rate communication systems*. John Wiley & Sons, 2005.

[18] "Sox - sound exchange," *http://sourceforge.net/projects/sox/*.

[19] H. Peyvandi and A. R. Ebrahimi, "A neural approach for compensation of nonlinear distortion effect in up to 1600bps data communication over voice-dedicated channels," in *IEEE International Conference on Telecommunications (ICT)*, 2010.