A New Approach for Digital Data Transmission over GSM Voice Channel

MAHSA RASHIDI
Electrical Engineering Department
Amirkabir University of Technology
15914, Tehran
IRAN

ABOLGHASEM SAYADIYAN
Electrical Engineering Department
Amirkabir University of Technology
15914, Tehran
IRAN

Abstract:- One of the important objectives in mobile communication systems is secure voice and data communication (including text, picture, video and voice) especially in high bit rate. For this reason, in this paper, a new procedure is proposed in which the intended data or voice is encrypted and modulated onto speech-like waveforms; Then the modulated waveforms are transmitted over the global system for mobile communications (GSM) voice channel. Finally, these waveforms are demodulated and decrypted at the receiver. This paper focuses on designing an appropriate model for the GSM Full Rate (FR) speech codec by describing the data mapping scheme onto the basic parameters of speech-like waveforms i.e. formants' phase, frequency and pitch frequency. The proposed model has been evaluated for a GSM-to-GSM connection. It is observed that, using the proposed method, it's possible to transmit data at 1.15kbps with 0.2% bit error rate (BER) and after error control codes we achieve a 600 bps channel with 0.02% BER.

Key-Words: - Speech-like waveform, GSM, Formant, speech codec, BER.

1 Introduction

Hardware and protocol deficiencies are two drawbacks for the second-generation mobile communications systems which makes them only capable for data transmission in low bit rates (e.g. 1120 bits per page) through Short Message Service (SMS) in G.7 signaling channel. However, in this generation using limited number of data channels available for subscribers, data transmissions will be possible to a maximum rate of 9.6 kbps.

The novelty in this paper is to consider the data transmission over the GSM voice channel in high bit rates. The major advantage in this method is the lack of data channel deficiencies like interoperability problems especially in international networks and frequent channel delays [1].

The most important problem in data transmission over the GSM voice channel is to make sure whether the transmitted data is highly secure. To this end, the resulting bit stream from a low bit rate speech coder [2] implemented for voice channel adaptation, enters into data encrypting block. Next, this data will be modulated on the speech-like waveform prior entering the GSM network. The resulting waveform then enters the first GSM handset, communications channel and then reaches the second GSM handset. Finally, the received waveform from second handset is demodulated, decrypted and decoded [3]. This completes the whole process for data communication. Due to low bit rate of speech channels in such communications, we require modems having data transfer capability with low bit rate. Based on the proposed approach in this paper, an appropriate modem is designed for such communications. Finally, using a Punctured ½-rate convolutional code with a constraint length of 7, we reach a system with a throughput of R=1.15 kbps rate with 0.2 % BER.

In some recent works Katugampala, in modulator side a codebook including the values of speech-like waveform parameters is defined including pitch frequency, LSF coefficients and energy of frame. Next these parameters are used for waveforms synthesis. Finally, the encrypted data are mapped onto these waveforms. In demodulator, these parameters will be derived from the received speech-like waveforms and compared to codebook and finally the best one is chosen [3]. Meanwhile, this approach has been adopted for GSM Enhanced Full Rate (EFR) speech codes 12.2 kbps whereas the proposed approach in this paper is considered for 13 kbps GSM FR speech codes corresponding to ETSI GSM 06.10. In the following sections, we will present the best way for producing appropriate speech-like signal for digital data transmission.

2 Speech-like signal production pro-Cedure

2.1 Fundamentals of speech-like signal production

We require mapping data bits stream on speech-like waveforms with 20ms length (equal to what is available in GSM speech coder). Therefore in this paper, we produce speech-like waveforms with Auto-regressive (AR) modeling; waveforms should be produced with four formants so that they can be adapted to GSM coder. As formants are sensitive to changes, the corresponding transfer functions should be parallel [4]. Finally, by applying excitation signal to resulting transfer function, appropriate speech-like signal will be produced. Transfer function of each formant is a second-order difference equation as follows:

$$H(k) = \frac{A}{1 - BZ^{-1} - CZ^{-2}}$$

$$\begin{aligned} A &= 1 - B - C \\ B &= 2\exp(-\pi\,\Delta f\,/\,fs).\mathrm{Cos}(2\pi\,f\,/\,fs) \\ C &= -\exp(-2\pi\,\Delta f\,/\,fs) \end{aligned} \tag{1}$$

Where in equation (1) Δf, f are formants' frequency and bandwidth and f_s is sampling frequency. Note that paralleling of these functions is done under conditions as follows: Firstly, we normalized the transfer function of each formant to its central frequency:

$$\left| H_i(e^{j\omega}) \right|_{\omega=\frac{2\pi k_i}{N}} = 1 \qquad i = 1,2,3,4 \tag{2}$$

Secondly, the normalization condition should be presented in a parallel format as below:

$$\boxed{\sum_{n=1}^{4}\left[\begin{array}{l}\alpha_n\left|H_1(e^{j\omega_n})\right| + \beta_n\left|H_2(e^{j\omega_n})\right| + \\ \rho_n\left|H_3(e^{j\omega_n})\right| + \lambda_n\left|H_4(e^{j\omega_n})\right| = 1\end{array}\right]} \tag{3}$$

Where in equation (3) α_n, β_n, ρ_n and λ_n are the so-called normalized equation coefficients . Finally, the speech-like waveforms will be resulted from spectral envelope using the harmonic synthesis method. Equation (4) shows the complete process for waveforms production; where $H_{Total}(k)$ is the same Paralleling transfer function and, N_{anly} is the analysis window length; N_{fr} is the frame shift length.

$$\begin{aligned} a(i) &= H_{Total}\,(f(i)) \\ \phi(i) &= \frac{2\pi a(i)}{N}\left(\frac{N_{anly}}{2} - N_{fr}\right) \\ s(n) &= \sum_{i=1}^{M} a(i)\cdot\cos\left(\frac{2\pi f(i)}{N}n + \phi(i)\right), n = 0,1,..,(160-1) \end{aligned} \tag{4}$$

In (4), three parameters $a(i)$ ، $\phi(i)$ ، $f(i)$ are available from M picked peaks among N/2 samples that has been selected by peak picking approach. Fig.1 shows a prototype of speech-like waveforms with 20ms length produced by harmonic modeling approach.

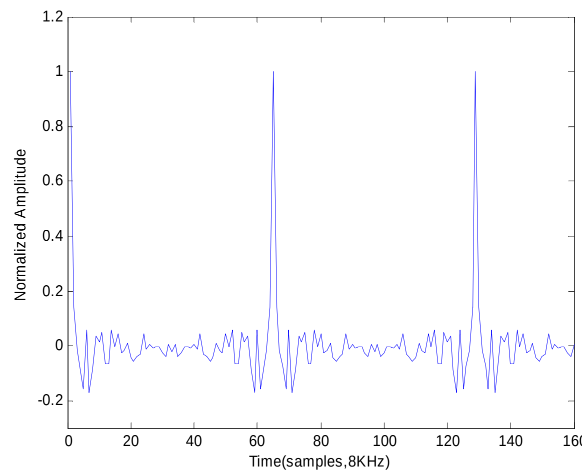

Fig.1: Synthesized speech-like signal

2.2 Data mapping scheme on produced speech-like signal

One of the important things is the correct selection of formants' frequency within telephone voice band (400-3500 Hz). During experiments and investigations we concluded that among the mentioned parameters in formants, their frequency and phase can be only detected as the passing signals speech-coded voice channel. As a result, we will explain in detail how to select parameters and to allocate data bits to frequency and phase parameters. We also concluded that we should select the frequency of first and second formants among the frequency range of 400 to 1000 Hz and coded this frequency range by eight states equal to 3 bits. Note that, the frequency range of the third formant is 1400 to 2500 Hz which is coded by 3 bits and fourth format range is 2900 to 3500 which is coded by 2 bits. The selection criteria are as follows:

1- Received signal envelope range should be more than 70 percent of transmitted signal envelope.

2- Frequency displacement of received formants should not be more than defined frequency steps for each formant. Otherwise, it causes incorrect extraction of saved information in the resulting frequencies.

3- Another important point is the lack of proximity in two adjacent formants. As a result, there are unusable band regions in boundary between formants. According to the experiment results, minimum distance for two adjacent formants is twice the bandwidth considered.

As there are phase fidelity in frequencies under 1000 Hz, we concluded that some information should be stored in phase related to first and second formants and this phase should be coded with 3 bits. Thus the extracted phase from received signal envelope is equal to the mapped phase in that particular frequency plus the primary phase that is acquired from transmitted signal envelope. As a result, selecting appropriate bandwidth is very important. On the other hand there is a direct relation between bandwidth, pitch frequency and intera-frame interpolation, thus we tried to select these two parameters in such a way to have minimum amount of unusable frequency band between formants so that mapped data on waveform signal in maintained in interpolation stage due to low pitch frequency as close as possible. In addition, we observed that, pitch frequency ranging from f_p=125 Hz to f_p=100 Hz gives acceptable results; therefore we coded the mapped data on pitch frequency with 1 bit. Due to the lake of fidelity in GSM coder/decoder to formants bandwidth, we only consider constant and similar bandwidth of $\Delta f = 180Hz$. Finally, speech-like waveform is modulated by 12 bits digital data in a 20ms transmission. Fig.2 illustrates one frame of synthesized and received signals with the same length of 400ms which produced by the proposed method. Note that signals are not selected from Interpolated samples. In addition, Fig.3 depicts spectral envelopes of these signals which have been shown in Fig.2.

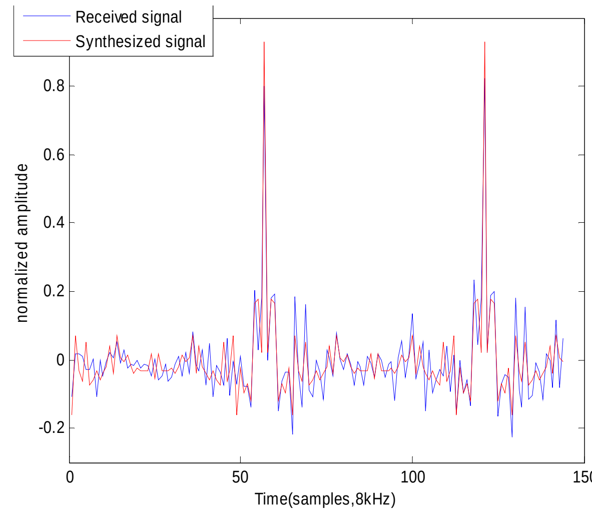

Fig.2: Synthesized and received signals

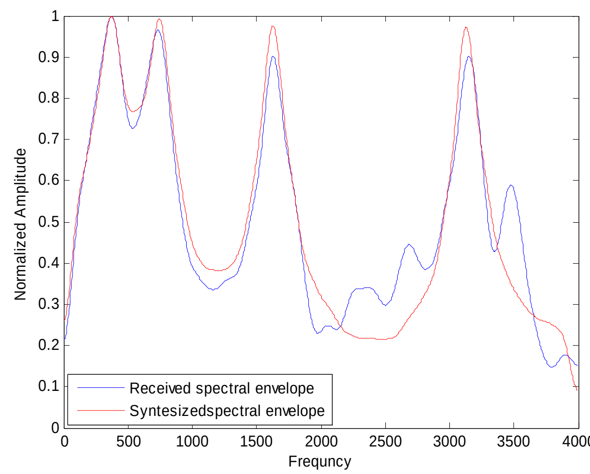

Fig.3: Synchronized and received envelopes

2.3 Intera-frame Interpolation

In order to achieve phase continuity that is an important characteristic in speech signals, it is necessary to overlap the produced speech-like waveforms with above approach. Also it should be considered that data bit streams on speech-like signal remain undamaged. To this end, it is so important to select pitch period i.e. $1/f_p$ that has direct relation to digital data mapped on each frame. GSM codec does a linear interpolation between Log Area Ratio (LAR) coefficients to avoid spurious transients, as well as interpolating LAR coefficients of the last frame's the primary 40 samples with LAR coefficients of the current frame's the primary 40 samples [5]. This motivates us to idea that adjacent frames should have the minimum overlapping so that when a PCM waveform signal starts GSM tandem connection; high overlapping of intera-

ISSN: 1790-5117

ISBN: 978-960-6766-34-3

frames does not cause tremendous changes in reflection coefficients of each frame and as a result does not result in incorrect detection of transmitted data. The important thing in designing the demodulator is that one should not select overlapping samples in each frame. In order to prevent from inter-modulation effect between formants. Next, appropriate PCM waveform signal has been prepared to enter into the speech coded voice channel. Equation 5 presents the linear interpolation for proposed modulator:

$$\begin{aligned}
&\begin{cases} y_1 = \dfrac{(n-153)}{16} \\ y_2 = \dfrac{(169-n)}{16} \qquad n = 153,154,..,160 \end{cases} \\
&\begin{cases} y_3 = \dfrac{(8+m)}{16} \\ y_4 = \dfrac{(8-m)}{16} \qquad m = 1,2,..,8 \end{cases} \\
&L_1 = [y_1, y_3] \\
&L2 = [y_2, y_4] \\
&Inter0 = L_1 \cdot [s_{(i-1)}(153:160), s_i(1:8)] \\
&Inter1 = L_2 \cdot [s_{(i-1)}(153:160), s_i(1:8)] \\
&Inter = Inter0 + Inter1
\end{aligned} \tag{5}$$

$s_{(i-1)}, s_i$ are the last frame and current frame.

3 Synchronization

Because of modulator and demodulator symbols are asynchronous, at the start of any communication a predefined synchronization sequence is sent from the modulator to the demodulator. This sequence of samples is known to both. Since in the simulation it is known that there will be a synchronization sequence in the input signal, the synchronization module cross-correlate a fixed predefined number of input samples in the beginning of the transmission with the predefined synchronization sequence. The sample sequence that best matched is used for synchronization in the demodulator. In addition, the speech codec frames in two base-stations are unsynchronized so that in this paper we considered the effects introduced by unsynchronized codecs. To consider this, we inserted a random number of samples before the signal passed to the second codec.

4 Conclusions

An appropriate method is proposed for digital data transmission over a GSM voice channel. The method was based on transmitting intended data on the produced speech-like signal frequency and phase of which result in transferring 12 bits data on a speech-like waveform with a frame size of 20ms. It was demonstrated that the proposed method could result in acceptable results achieving at a 1.15kps channel with 0.2% BER and 600 bps with 0.02% BER after error correcting codes were imposed.

5 References

[1] M. Street, “Interoperability and international operation: An introduction to end to end mobile security”, IEE Secure GSM and Beyond: End to End Security for Mobile Communications, London ,February 2003.

[2] M. Stefanovic, Y. D. Cho, S. Villette, and A. M. Kondoz, “A 2.4/1.2 kb/s speech coder with noise pre-processor”, proceedings EUSIPCO 2000, Tampere, Finland, pp. 4-8, September 2000.

[3] N. Katugampala, S. Villette, and A. Kondoz, “Secure voice over GSM and other low bit rate systems,” IEE Secure GSM and Beyond: End to End Security for Mobile Communications, London, February 2003.

[4] D. H. Klatt, ”Software of cascade/parallel formant synthesizer”, J. Acoustic. Soc. Am. 67 (3), Massachusett,pp 971-996, Mar.1980.

[5] J. Degener and C. Bormann.” Gsm 06.10 lossy speech compression”.ftp://ftp.cs.tu-rlin.de/ pub/local/ kbs / tubmik/gsm/gsm-1.0.10.tar.gz