Performance Evaluation of Variable Bitrate Data Hiding Techniques on GSM AMR coder

Dr.G.Indumathi ,R.Anusuya

Professor
Mepco Schlenk Engineering College

Abstract Wireless systems involved in transmission of speech envisage that efficient and effective methods be developed to transmit and receive the same while maintaining quality-of-speech, especially at the receiving end. The goal of the project is to implement variable bitrate steganographic data transmission over GSM Adaptive Multi Rate (AMR) coder. In order to facilitate the same bitrate steganographic data transmission a few pulses are identified and utilized for embedding and hiding the data bits into them. Joint speech coding and data hiding is accomplished. This is implemented on the standard AMR coder and the performance evaluation using subjective measures that include mean opinion score and the objective measures like Perceptual Evaluation of Speech Quality, Segmental SNR are done. The stego signal, which is small data information, is used for embedding over the encoded wave file.

Keywords: GSM adaptive multi rate coder, Joint speech coding, Mean Opinion Score, Perceptual Evaluation of Speech Quality, Segmental Signal to Noise Ratio.

I. INTRODUCTION

In Wireless communication networks, in order to ember the steganography data within the encoded data and to transmit the signal, various methods have been used. Initially, the signal was encoded and then the data was embedded using source coding with the reduction in audibility (Geiser and Vary 2008). Recently, Joint Source Coding and Data Hiding for data encoding and hiding is used. In this popular technique, the quality of the speech signal provided by the host signal is considered to be important. The steganographic data is embedded after the speech signal is compressed and encoded. Authentication and digital rights management are provided by watermark data hiding. For higher variable data transmission rate, the performance of stego signals for robustness is more relevant (Shahbazi et al 2010). The GSM voice channel, in general, uses one of the standard speech codecs like Half Rate (HR), Full Rate (FR), Enhanced Full Rate (EFR) and Adaptive Multi Rate (AMR).

This work aims at implementing and performance evaluation of variable bitrate data transmission over GSM Adaptive Multi Rate coder. The method used for embedding the data stream is Joint source coding and data hiding approach. The performance of the GSM coder is evaluated using the subjective measure like Mean Opinion Score (MOS) and the objective measures that include Perceptual Evaluation of Speech Quality (PESQ), Segmental SNR (segSNR), Log-Likelihood Ratio (LLR) and Weighted slope-spectral distance (WSS).

II. ETSI GSM ADAPTIVE MULTI RATE CODER

In preprocessing, the input signal is high-pass filtered and scaled. The input for all the subsequent analysis is the preprocessed signal. The LP filter coefficients are computed using LP analysis which is done once per 10ms frame which are then converted to line spectral pairs (LSPs) and quantized. The error signal is filtered using a perceptual weighting filter which improves the performance for input signals with a flat frequency-response. The quantized and un-quantized LP filter coefficients are used for the second sub-frame, while in the first sub-frame interpolated LP filter coefficients are used (both quantized and un-2 quantized). Based on the perceptually weighted speech signal, an openloop pitch delay is estimated. The target signal is computed by filtering the LP residual through the weighted synthesis filter. Using the target signal and the impulse response, closed-loop pitch analysis is done. The fixed-codebook excitation is done using an algebraic codebook with 17 bits. The filter memories are then updated using the determined excitation signal.

The received bit stream is used for extracting the parameter's indices which are then decoded to obtain the coder parameters that include LSP coefficients, the two fractional pitch delays, the two fixed-codebook vectors and the two sets of adaptive and fixed-codebook gains. The LP filter coefficients for each sub-frame are formed by interpolating and converting the LSP coefficients. For each 5ms sub-frame, the adaptive and fixed-codebook vectors are added and scaled by their respective gains to construct the excitation. This excitation is filtered using the LP synthesis filter to reconstruct the speech signal. The reconstructed speech signal is then passed through an adaptive post-filter followed by a high pass filter and up-scaling.

III .JOINT SOURCE CODING AND DATA HIDING

For steganographic data transmission over the encoded speech stream, several researches have been carried out over a few years. The most popular methods are Least Significant Bit (LSB) insertion, Spread Spectrum, Echo and Phase Coding, Auditory Masking and Quantization Index Modulation (QIM). This work is mainly based on Joint Source Coding and Data Hiding techniques. (Fig 1)
The three prominent data embedding approaches are hiding in temporal domain, in frequency/wavelet domains and in coded domain. Low-bit encoding technique is employed in majority of temporal domain methods.

3.1. Low-bit encoding

Low-bit encoding also known as Least Significant Bit (LSB) is one of the earliest methods used for information hiding. Traditionally, it is based on embedding each bit from the message in the least significant bit of the cover audio in a deterministic way. Thus, 16 kbps of data are hidden for a 16 kHz sampled audio. The LSB method allows high embedding capacity for data. It is relatively easy to combine with other hiding techniques. However it becomes vulnerable to simple attacks thereby reducing its security performance being characterized by low robustness to noise addition. The stego data is very likely to be destroyed by filtration, amplification, noise addition and lossy compression.

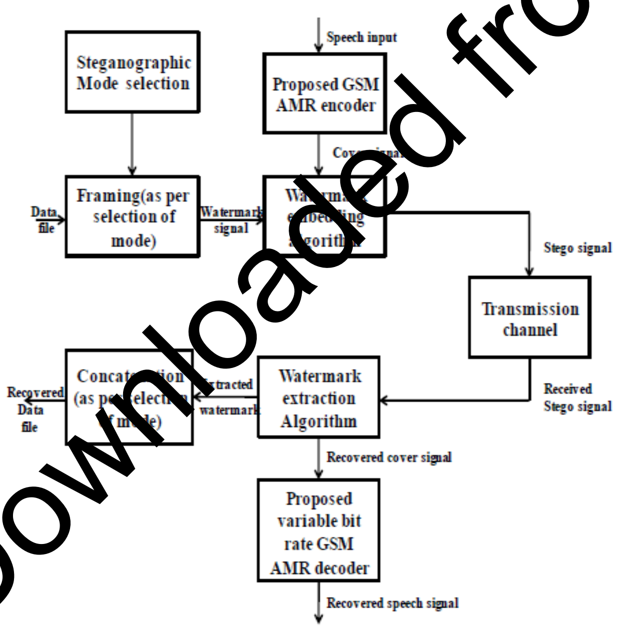

Figure1. Joint variable bitrate GSM AMR coding and data hiding system

IV OBJECTIVE MEASURES

Several objective speech quality measures can be evaluated: segmental SNR (segSNR), weighted-slope spectral distance (WSS), PESQ, LPC-based objective measures including the log-likelihood ratio (LLR), Itakura-Saito distance measure (IS), and cepstrum distance measures (CEP) and frequency-weighted segmental SNR (fwsegSNR).

4.1. Perceptual evaluation of speech quality

Among all objective measures considered, the PESQ measure is the most complex to compute and is the one recommended by ITU-T for speech quality assessment of 3.2 kHz (narrow-band) handset telephony and narrow-band speech codecs. The PESQ score is computed as a linear combination of the average disturbance value and the average asymmetrical disturbance values. The parameters were optimized for speech processed through networks and not for speech enhanced by noise suppression algorithms. The PESQ score is computed as a linear combination of the average disturbance value and the average asymmetrical disturbance values as follows:

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \qquad (1)$$

$a_0 = 4.5, a_1 = -0.1$ and $a_2 = -0.0309$. The parameters a_0, a_1 and a_2 in the above equation were optimized for speech processed through networks and not for speech enhanced by noise suppression algorithms.

4.2. LPC-based objective measures

Three different LPC-based objective measures can be considered: the LLR, the IS, and the cepstrum distance measures. The LLR measure is defined as

$$d_{LLR}(\overrightarrow{a_p}, \overrightarrow{a_c}) = \log\left[\frac{\overrightarrow{a_p} R_c \overrightarrow{a_p}^T}{\overrightarrow{a_c} R_c \overrightarrow{a_c}^T}\right] \qquad (2)$$

where is the LPC vector of the original speech signal frame, is the LPC vector of the enhanced speech frame, and is the autocorrelation matrix of the original speech signal. The segmental LLR values were limited in the range of [0, 2] to further reduce the number of outliers. The IS measure is defined as

$$d_{IS}(\overrightarrow{a_p}, \overrightarrow{a_c}) = \frac{\sigma_c^{\,2}}{\sigma_p^{\,2}}\left(\frac{\overrightarrow{a_p} R_c \overrightarrow{a_p}^T}{\overrightarrow{a_c} R_c \overrightarrow{a_c}^T}\right) + \log\left(\frac{\sigma_c^{\,2}}{\sigma_p^{\,2}}\right) - 1 \qquad (3)$$

where and are the LPC gains of the clean and enhanced signals, respectively. The IS values were limited in the range of [0,100]. This is necessary in order to minimize the number of outliers. The cepstrum distance provides an estimate of the log spectral distance between two spectra. The cepstrum distance was limited in the range of [0, 10].

4.3. Time domain and frequency-weighted SNR measures

The time-domain segmental SNR (segSNR) measure can be computed. Only frames with segmental SNR in the range of 10 to 35 dB can be considered in the average. The frequency-weighted segmental SNR (fwSNRseg) was computed using the following equation:

$$fwSNRseg = \frac{10}{M} * \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W(j,m) \log_{10} \frac{|X(j,m)|^2}{(|X(j,m)| - |\overline{X}(j,m)|)^2}}{\sum_{j=1}^{M} W(j,m)} \quad (4)$$

where W is the weight placed on the j_{th} frequency band, K is the number of bands, M is the total number of frames in the signal, is the weighted (by a Gaussian-shaped window) clean signal spectrum in the j_{th} frequency band at the m_{th} frame, and is the weighted enhanced signal spectrum in the same band. For the weighting function, we considered the magnitude spectrum of the clean signal raised to a power, i.e.,) ,

$$W(j,m) = |X(j,m)|^{\gamma} \quad (5)$$

where is the weighted magnitude spectrum of the clean signal obtained in the j_{th} band at frame m and is the power exponent, which can be varied for maximum correlation.

The WSS distance measure computes the weighted difference between the spectral slopes in each frequency band. The WSS measure evaluated is defined as

$$d_{WSS} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W(j,m)(S_c(j,m) - S_p(j,m))^2}{\sum_{j=1}^{M} W(j,m)} \quad (6)$$

where are the weights computed, and are the spectral slopes for j_{th} frequency band at frame m of the clean and processed speech signals, respectively.

V RESULTS AND DISCUSSIONS

The speech waveform is encoded using GSM AMR encoder and then the steganographic data is embedded into the encoded signal using the watermark embedding algorithm. The stenographic bit-stream is then extracted using the watermark extraction algorithm in the decoder section. The objective and the subjective measures like Mean opinion score, PESQ, segmental SNR, LLR and WSS are evaluated.

In subjective analysis, the mean opinion score (MOS) of the signals reduces with the increase in the bitrate of the steganographic data transmission. (Table 1)

The distortion, background distortion and the overall quality of different speech signals were compared and tabulated as shown in Table 2. The objective measures like SegSNR, PESQ, WSS and LLR were calculated and compared for different signals as shown in Table 3. Figure 2 shows the comparison chart of different objective measures.

Table 1 Mean opinion score

Mean opinion score (MOS)		
MOS	***Quality***	***Impairment***
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Table 2 Comparison of distortion, background distortion and overall quality for different speech signals

Signal	**Distortion**	**Background Distortion**	**Overall Quality**
Sp01	3.9896	2.7315	3.3126
Sp02	4.5529	3.1[illegible]90	4.0207
Sp03	4.1303	2.7852	3.5039
Sp04	3.3327	2.6736	3.0499
Sp05	3.4847	2.4426	2.8606
Sp06	3.3050	2.6160	2.7133

Table 3 Comparison of objective measures

Signal	**PESQ**	**LLR**	**SNRseg**	**WSS**
Sp01	2.654	0.448	0.274	26.938
Sp02	3.488	0.413	0.912	24.286
Sp03	2.896	0.451	0.677	27.196
Sp04	2.905	1.017	0.2054	51.705
Sp05	2.339	0.609	0.0277	43.473
Sp06	2.256	0.681	0.991	49.671

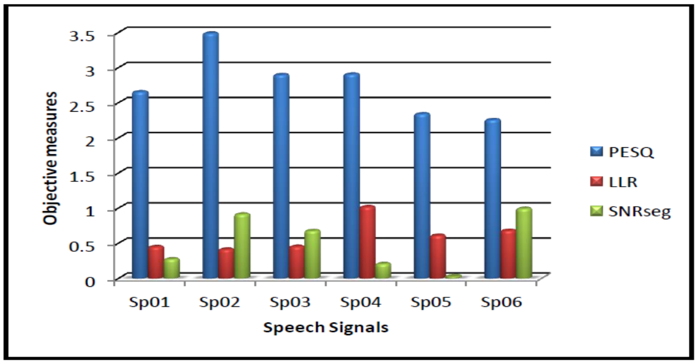

Figure2. Comparison chart of different objective measures

CONCLUSION

In this work, the original speech signal is encoded using GSM AMR encoder. The embedding and extraction of small text files into the encoded signal using LSB steganography have been conducted for steganography modes for GSM AMR coders. The implementation and analysis shows the trade-off between the speech quality and the embedding capacity. The subjective measure that includes MOS and the composite measures like LLR, WSS, PESQ and SegSNR reveal a gradual reduction in speech quality with reference to the steganography bitrate modes. However if both the GSM FR coder and the AMR coder are compared in performance the latter coder has improved speech quality compared with the former coder.

REFERENCES

[1]. A. Kataoka, T. Moriya, and S. Hayashi, “An 8 kbit/s speech coder based on conjugate structure CELP,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 1993, pp. II-592–II-595.

[2]. “Implementation and performance of an 8 kb/s conjugate structure CELP speech coder,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 1994, pp. II-93–II-96.

[3]. “An 8-kb/s conjugate structure CELP (CS-CELP) speech coder,” IEEE Trans. Speech Audio Processing, no. 6, pp. 401–411, Nov. 1996.

[4]. ITU-T, “Recommendation G.729—Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CSACELP),” Mar. 1996.

[5]. CCITT/ITU-T, “Rec. G.729: Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)” in General Aspects of Digital Transmission Systems; Terminal Equipments, Series G Recommendations, International Telecommunication Union, 1996.

[6]. A.Kataoka et al, “An 8-kbit/s conjugate structure CELP speech coder”. IEEE Transaction on speech and audio processing.Vol.4, No.6, November 1996.

[7]. D. Malkovic, “Speech Coding Methods in Mobile Radio Communication Systems”, 17th International Conference on Applied Electromagnetics and Communications, Croatia, 2003, pp.103-108.

[8]. N. Bhatt, Y. Kosta. “Overall Performance Evaluation of Adaptive Multi Rate 06.90 Speech Codec based on Code Excited Linear Prediction Algorithm using MATLAB”, International Journal of Speech Technology, springer, published online Jan. 2012

 ISBN : 978-81-929742-1-7