Proposed Modifications in ETSI GSM Full Rate Speech Codec in line with bitrates of GSM EFR Speech Codec and its Objective Evaluation of Performance using MATLAB

Ninad Bhatt[1], Yogeshwar Kosta[2]

[1]*Research Scholar, Veer Narmad South Gujarat University, Surat, Gujarat, India*
bhattninad@gmail.com

[2]*Principal, Marwadi Education Foundation, Rajkot, Gujarat, India*
ypkosta@yahoo.com

Abstract

Today, the primary constrain in wireless communication system is limited bandwidth and power. Wireless systems involved in transmission of speech envisage that efficient and effective methods be developed (bandwidth usage & power) to transmit and receive the same while maintaining quality-of-speech, especially at the receiving end. Speech coding is a technique, since the era of digitization (digital) and computerization (computational and processing horsepower - DSP) that has been a material- of- research for quite sometime amongst the scientific and academic community. Amongst all elements of the communication system (transmitter, channel and receiver), transmission channel (carrier of information/data, also called the medium) is the most critical and plays a key role in the transmission and reception of information/data.

This paper proposes some modifications in the Long Term Gain and selection of Grid position sections of 13kbps GSM Full Rate coder so that it can be mapped to bitrates of GSM Enhanced Full Rate at12.2kbps. Here, Objective analysis is carried out on a proposed system using MATLAB to evaluate its performance and the results obtained are then compared to the results of GSM Full Rate coder. Proposed coder, at bitrates of 12.2 kbps, provides moderately low computational complexity [1] as its implementation is with reference to GSM full rate coder rather than using Algebraic Code Excited Linear Prediction (ACELP) algorithm in case of standard GSM EFR coder but it offers small degradation in speech quality compared to standard GSM EFR coder. In comparison with standard GSM Full Rate coder, here the proposed coder offers other benefit of saving of 0.8kbps which can be useful for better error protection during channel coding while also provides satisfactory results for the parameters of Objective analysis.

Keywords: *Speech Coding, GSM, ETSI, RPE-LTP coder, GSM EFR coder, MATLAB*

1. Introduction

Full Rate GSM 06.10 Speech Coder basically belongs to Hybrid coder (Analysis by Synthesis coder) which provides attractive trade off between waveform coders and vocoders, both in terms of speech quality and transmission bit rate, although generally at the price of higher complexity. The speech encoder takes its input as a 13 bit uniform PCM signal either from the audio part of the mobile station or on the network side, from the PSTN via an 8bit / A-law to 13 (13bit* 8KHz=104Kbps) bit uniform PCM as specified in GSM 06.01 [8]. The encoded speech at the output of the speech encoder is delivered to a channel encoder unit which is specified in GSM 05.03 [10]. In the receive direction, inverse operations take place. GSM 06.10 describes the detailed mapping between input blocks of 160 speech samples in 13 bit uniform PCM format to encoded blocks of 260 bits and from encoded blocks of 260 bits to output blocks of 160 reconstructed speech samples. The rate of sampling is 8000 samples/s leading to an average bit rate for the encoded bit stream of 13kbps. The coding scheme is so called Regular Pulse Excitation-Long Term Prediction-Linear Predictive Coder.

2. GSM Full Rate Encoder

The detailed block diagram of GSM 06.10 Speech Encoder is shown in Figure 1. The input speech frame, consisting of 160 signal samples is first pre-processed to produce an offset free signal, which is then subjected to a first order pre-emphasis filter. The 160 samples obtained are then analyzed to determine the coefficients for the short term analysis (LPC Analysis). These parameters are then used for the filtering of the same 160 samples. The result is 160 samples of the short term residual signal. The filter parameters, termed reflection coefficients, are transformed to log area ratios, LARs, before transmission. The speech frames are divided into 4 sub- frames with 40 samples of the short term residual signal in each. Each sub-frame is processed block wise by the subsequent functional elements. Before the processing of each sub block of 40 short term residual samples, the parameters of the long term analysis filter, the LTP lag and the LTP gain, are estimated and updated in the LTP analysis block, on the basis of the current sub-block of the present and a stored sequence of the 120 previous reconstructed short term residuals. A block of 40 long term residual signal samples is obtained by subtracting 40 estimates of the short term residual signal from the short term residual signal itself. The resulting block of 40 long term residual samples is fed to the Regular Pulse Excitation analysis which performs the basic compression function of the algorithm. As a result of the RPE analysis, the block of 40 input long term residual samples are represented by one of 4 candidate sub-sequences of 13 pulses each. The subsequence selected is identified by RPE grid position (M). The 13 RPE pulses are encoded using Adaptive Pulse Code Modulation (APCM) with estimation of the sub-block amplitude which is transmitted to the decoder as side information. The RPE parameters are also fed to a local RPE decoding and reconstruction module which produces a block of 40 samples of the quantized version of the long term residual signal. By adding these 40 quantized samples of the long term residual to the previous block of short term residual signal estimates, a reconstructed version of the current short term residual signal is obtained. The block of reconstructed short term residual signal samples is then fed to the long term analysis filter which produces the new block of 40 short term residual signal estimates to be used for the next sub-block thereby completing feedback loop [8].

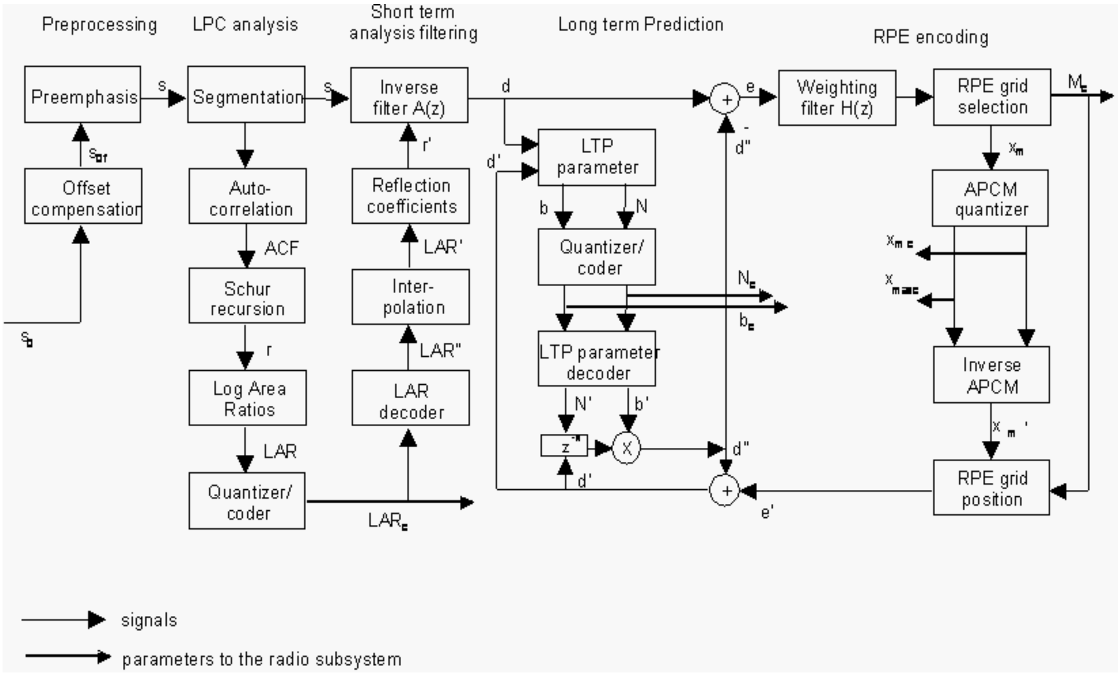

Figure 1. Detailed block diagram of Full Rate GSM 06.10 Speech Encoder [8]

3. GSM Full Rate Decoder

The detailed block diagram of GSM 06.10 Speech Decoder is shown in Figure 2. The decoder includes the same structure as the feedback loop of the encoder. In error free transmission, the output of this stage will be the reconstructed short term residual samples. These samples are then applied to the short term synthesis filter followed by the de-emphasis filter resulting in the reconstructed speech signal samples. GSM 06.10 describes the detailed mapping between input blocks of 160 speech samples in 13 bit uniform PCM format to encoded blocks of 260 bits and from encoded blocks of 260 bits to output blocks of 160 reconstructed speech samples. The sampling rate is 8000 samples/sec leading to an average bit rate for the encoded bit stream of 13 kbit/s. The bit allocation for the ETSI GSM 06.10 Full Rate Speech coder is as shown in Table 1 [8].

Table 1. Bit allocation for GSM Full Rate Speech Coder [8]

Parameter	No. per frame	Resolution	Total bits / frame
LPC	8	6,6,5,5,4,4,3,3	36
Pitch Period	4	7	28
Long Term Gain	4	2	8
Grid Position	4	2	8
Peak Magnitude	4	6	24
Sample Amplitude	4*13	3	156
Total			260

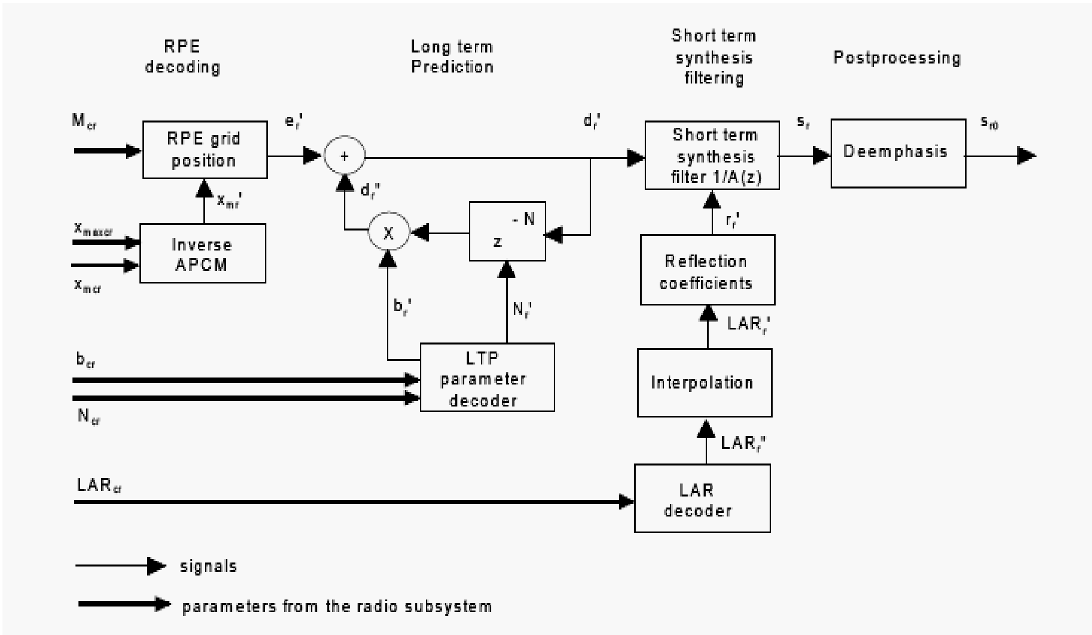

Figure 2. Detailed block diagram of Full Rate GSM 06.10 Speech Decoder [8]

4. Proposed Modification in GSM Full Rate Speech Coder

GSM Full Rate Coder Consists of three major blocks Linear Predictive Coding Section, Long Term Predictive Section and Regular Pulse Excitation Section. The proposed modifications are suggested in LTP and RPE Section in the selection of Long Term gain and grid positions respectively. To provide better quantization, one bit per sub segment is added in Long Term Gain Section. In RPE section, selection of grid position and samples is modified such a way that no samples repeat in multiple grids which is the case of GSM Full Rate coder in first and forth grid where except sample number 0 and sample number 39 don't repeat where as all other samples in both grids repeat. A new proposed grid selection strategy is as shown in Figure 3.

As can be seen in Figure 1, if the weighting filtered Prediction-error sequence is down-sampled by a ratio of 4 instead of 3, it results into four interleaved sequences with regularly spaced pulses. These are defined with

$$X_m[k]=X[m+4k]\ ; \qquad m=0,1,2,3; \qquad k=0,1,2\ldots\ldots\ldots10 \qquad (1)$$

Where m = no. of grids per sub segment and k = no. of samples per grid

The benefit in this sampling grid position selection is, there is no repetition of any sample in multiple grids where as now the total number of samples per grid reduces from 13 to 10 so ultimately there is a reduction in overall bit-rate which can be useful for mapping GSM EFR having 244 bits / frame of 20 ms (resulting 12.2 kbps) from GSM FR having 260 bits / frame of 20 ms (resulting 13 kbps). The proposed modification in GSM FR offers a new bit allocation table for mapping GSM EFR as shown in Table 2.

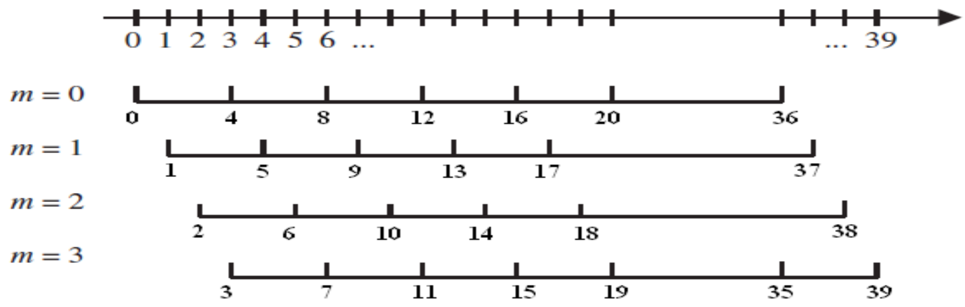

Figure 3. Sampling Grids used in Position Selection for proposed GSM EFR 12.2 Kbps coder

Table 2. Bit allocation for Proposed GSM Enhanced Full Rate Speech Coder

Parameter	No. per frame	Resolution	Total bits / frame
LPC	8	8,8,7,7,6,6,5,5	52
Pitch Period	4	7	28
Long Term Gain	4	3	12
Grid Position	4	2	8
Peak Magnitude	4	6	24
Sample Amplitude	4*10	3	120
Total			244

5. Objective Analysis

To evaluate the performance of proposed GSM EFR, the different types of Objective Analysis have been carried out in this paper. Objective Analysis has been categorized into waveform, spectral, perceptual and composite measures.

5.1. Waveform based Objective Analysis

The following parameters are evaluated in this category of Objective Analysis.

(1) Signal to Noise Ratio is mathematically defined as

$$SNR = 10\log_{10}\frac{\sum|Si|^2}{\sum|Si - So|^2} \tag{2}$$

Where S_i= input signal, S_o= decoded signal and N = total no. of frames

(2) Segmental SNR is mathematically given as

$$SNR_{SEG} = \frac{1}{M}\sum_{j=0}^{M-1} 10\log_{10}\left[\frac{\sum_{n=m_j-N+1}^{m_j} s^2(n)}{\sum_{n=m_j-N+1}^{m_j}\left[s(n)-\hat{s}(n)\right]^2}\right] \tag{3}$$

Where, S(n)= input signal, $S^{\wedge}(n)$= decoded signal, N=segment length, M= no. of segments and m_j = end of the current segment[10,20].

5.2. Perceptual and based Objective Analysis and Composite measure

The following is the important parameter for performing perceptual based analysis.

5.2.1. Perceptual Evaluation of Speech Quality (PESQ)

In comparison with other objective measures, the PESQ measure is the most complex to compute and is the one recommended by ITU-T P.862 for speech quality assessment of 3.2 kHz (narrow-band) handset telephony and narrow-band speech codecs [17,18]. PESQ score is computed as a linear combination of the average disturbance value D_{ind} and the average asymmetrical disturbance values A_{ind} as follows:

$$\mathrm{PESQ} = a_0 + a_1 D_{\mathrm{ind}} + a_2 A_{\mathrm{ind}} \tag{4}$$

Where $a_0 = 4.5, a_1 = -0.1, a_2 = -0.0309$

5.2.2. Composite measures

As conventional objective measures are not sufficinet to provide high correlations in terms of speech/noise distortion and overall speech quality, it is hence necessary to combine different objective measures in order to produce Composite measure [10]. Here, Multiple Linear Regression Analysis and Mulitvariate Adaptive Regression Splines (MARS) techniques are used to produce different parameters. With reference to ITU P.835, the following parameters are used for evaluation of Composite measure: Predicted rating of Overall Speech Quality (Covl), Rating of speech distortion (Csig) and Rating of background distortion (Cbak)[12,15,23,37].

5.3. Spectral based Objective Analysis

The following parameters are evaluated in this category of Objective Analysis.
(1) Frequency Weighted Segmental SNR (fwSNRseg) is expressed as follows

$$\mathrm{fwSNRseg} = \frac{10}{M} \times \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W(j,m) \log_{10} \frac{|X(j,m)|^2}{(|X(j,m)| - |\hat{X}(j,m)|)^2}}{\sum_{j=1}^{K} W(j,m)} \tag{5}$$

where W(j, m) 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, $|X(j,m)|$ is the weighted (by a Gaussian-shaped window) clean signal spectrum in the j^{th} frequency band at the m^{th} frame, and $|x^{\wedge}(j,m)|$ in the weighted decoded signal spectrum in the same band [10].

6. Performance Evaluation of Proposed GSM EFR Coder

Here, both standard GSM FR and proposed GSM EFR coders are implemented in MATLAB and performance of both coders is evaluated using different Objective measures.

First standard GSM FR coder is implemented in MATLAB and then proposed modifications are carried out in GSM FR in line with GSM EFR bit-rates of 12.2 kbps. For the purpose of Objective analysis, four different wave files have been chosen. Each Wave file is sampled at 8 kHz and coded by 16 bits mono [38]. Table 3 demonstrates the result analysis and comparison between both coders. As can be seen from Table 3, all parameters offer satisfactory results and it can be observed that with reduction in bit-rates from 13 kbps of standard GSM FR to 12.2 kbps of proposed GSM EFR offers small degradation in almost all parameters. Still the obtained results for proposed GSM EFR coder for all parameters are

Table 3. Performance comparison between standard GSM FR and proposed GSM EFR Coder

Algorithm	Wave files (.wav)	Perceptual Analysis	Composite measures			Waveform Analysis		Spectral Analysis
		PESQ	Csig	Cbak	Covl	SNR	SNRseg	fwSNRseg
GSM FR 13kbps	Five	2.8038	2.7017	2.4217	2.1832	3.5587	3.6560	8.4468
	Sp21	2.8598	2.6280	2.4405	2.3222	2.0206	2.4059	7.3820
	Ninad	3.0812	1.6643	2.1895	1.6737	3.9566	1.7324	1.1441
	Doormono	2.6190	2.9858	2.5744	2.4626	3.3471	3.5598	8.3558
Proposed GSM EFR 12.2kbps	Five	2.9818	2.8776	2.4756	2.4091	3.1805	2.9332	8.1343
	Sp21	2.6034	2.4903	2.2664	2.1055	1.6998	2.0749	6.2471
	Ninad	2.9213	1.5675	2.1926	1.6206	3.6330	1.9379	0.7384
	Doormono	2.3397	2.8603	2.4257	2.2762	2.8488	3.1982	7.9584

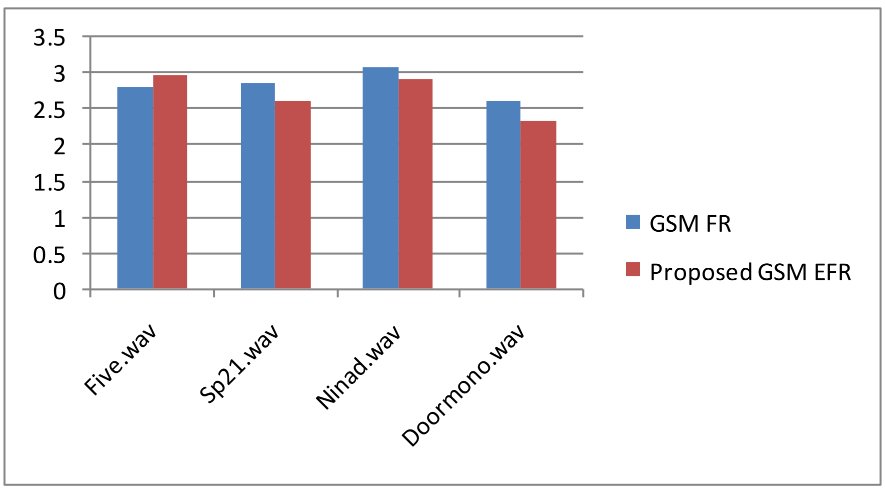

Figure 4. PESQ score comparison between standard GSM FR and proposed GSM EFR coders

quite comparable with respect to standard GSM FR coder. Figure 4 demonstrates the comparison of PESQ score between both coders. Except one wave file (i.e. Five.wav), for all other wave files PESQ score degrades with small amount for proposed GSM EFR when compared with standard GSM FR.

7. Concluding Remarks

In order to conserve the channel bandwidth, the role of a speech coder is to provide toll quality recovered speech signal even with comparatively lower bit rate and also with less delay and complexity. There is a trade off between Quality of Speech and Bit Rate. Full Rate GSM Speech codec offers moderate delay and less complexity in comparison with other coders but at the cost of comparatively moderately high bit rate.

The idea behind implementation of proposed GSM EFR coder is to reduce the bit-rate of standard GSM FR coder in line with bit-rates of GSM EFR coder. Standard ETSI based GSM EFR coder makes use of Algebraic Code Excited Linear Prediction (ACELP) algorithm which provides toll quality speech in comparison with standard GSM FR coder but at the cost of high complexity and higher coding delay [1,4]. The proposed GSM EFR coder, as is implemented in line with GSM FR coder, reduces overall complexity and delay (which is an inherent benefit of GSM FR coder) but provides small degradation in speech quality as observed in Figure 4 for PESQ scores. As can be observed from Table 3, results of different parameters of Objective evaluations are obtained and compared for both GSM FR and proposed GSM EFR coders and they are found satisfactory.

8. References

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

[2] M. Budagavi, “Speech Coding in Mobile Radio Communication”, proceedings of IEEE, Vol. 86, No.7, July-1998

[3] N. Bhatt, Palav Patel. “Real Time Implementation of Full Rate GSM 06.10 Speech Codec on TMS 320C6713 DSP kit and its Analytical Evaluation of Performance in Simulink”, IJERIA, Vol.2 No. 4, july 2009

[4] Jerry Gibson, “Speech Coding Methods, Standards and Applications”, IEEE Circuits and Systems Magazine, IEEE 4th Quarter 2005

[5] L. Basacier, “GSM Speech Coding and Parameter Recognition”, University of Neuchatel, A.L. Breguet, Switzerland

[6] R. Meston, “GSM Vocoders Improve Speech Transmission”, Recal Instrumentation

[7] Jie Yang, Sheng sheng Yu, Mian Zhao “ The Implementation And Optimization Of Amr Speech Codec On Dsp” , 2007 International Symposium on Intelligent Signal Processing and Communication Systems

[8] ETSI, Digital Cellular Telecommunications System (Phase 2+), Full Rate Speech, Transcoding, (GSM 06.10 version 8.2.0 Release 2005-06), pp.10-59

[9] K.Jarvinen “ Standardisation Of the Adaptive Multirate Codec”, IEEE Conference 2000.

[10] Yi Hu, Philipos C. Loizou “Evaluation of Objective Quality Measures for Speech Enhancement”, IEEE Transactions on Audio, Speech and Language processing, Vol. 16, No. 1, Jan. 2008

[11] S. Quackenbush, T. Barnwell, and M. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[12] “Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm,” ITU-T, ITU-T Rec.P. 835, 2003.

[13] P. Kroon, , W. Kleijn and K. Paliwal, Eds., “Evaluation of speech coders,” in *Speech Coding and Synthesis*. New York: Elsevier, 1995,pp. 467–494.

[14] L. Thorpe and W. Yang, “Performance of current perceptual objective speech quality measures,” in *Proc.* IEEE Speech Coding Workshop,1999, pp. 144–146.

[15] T. H. Falk and W. Chan, “Single-ended speech quality measurement using machine learning methods,” IEEE Trans. Audio, Speech, Lang.Process., vol. 14, no. 6, pp. 1935–1947, Nov. 2006.

[16] L. Malfait, J. Berger, and M. Kastner, “P.563-the ITU-T standard for single-ended speech quality assessment,” IEEE Trans. Audio,Speech,Lang. Process., vol. 14, no. 6, pp. 1924–1934, Nov. 2006.

[17] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, “Perceptual evaluation of speech quality (PESQ)-A new method for speech quality assessment of telephone networks and codecs,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2001, vol. 2, pp. 749–752.

[18] “Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” ITU, ITU-T Rec. P. 862, 2000.

[19] R. Kubichek, D. Atkinson, and A. Webster, "Advances in objective voice quality assessment," in Proc. Global Telecomm. Conf., 1991, vol.3, pp. 1765–1770.

[20] Y. Hu and P. Loizou, "Subjective comparison of speech enhancement algorithms," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,2006, vol. 1, pp. 153–156.

[21] Y. Hu and P. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun., vol. 49, pp. 588–601,2007.

[22] T. Rohdenburg, V. Hohmann, and B. Kollmeir, "Objective perceptual quality measures for the evaluation of noise reduction schemes,"in Proc. 9th Int. Workshop Acoust. Echo Noise Control, 2005, pp.169–172.

[23] J. Salmela and V. Mattila, "New intrusive method for the objective quality evaluation of acoustic noise suppression in mobile communications," in Proc. 116th Audio Eng. Soc. Conv., 2004, preprint 6145.

[24] V. Turbin and N. Faucheur, "A perceptual objective measure for noise reduction systems," in Proc. Online Workshop Meas. Speech Audio Quality Netw., 2005, pp. 81–84.

[25] E. Paajanen, B. Ayad, and V. Mattila, "New objective measures for characterization of noise suppression algorithms," in IEEE Speech Coding Workshop, 2000, pp. 23–25.

[26] V. Mattila, "Objective measures for the characterization of the basic functioning of noise suppression algorithms," in Proc. Online Workshop Meas. Speech Audio Quality Netw., 2003 [Online]. Available:http://wireless.feld.cvut.cz/mesaqin2003/contributions.html

[27] A. Bayya and M. Vis, "Objective measures for speech quality assessment in wireless communications," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1996, vol. 1, pp. 495–498.

[28] V. Turbin and N. Faucheur, "Estimation of speech quality of noise reduced signals," in *Proc. Online* Workshop Meas. Speech Audio Quality Netw., 2007 [Online]. Available: http://wireless.feld.cvut.cz/mesaqin2007/contributions.html

[29] Y. Hu and P. Loizou, "Evaluation of objective measures for speech enhancement," in Proc. Inter speech, 2006, pp. 1447–1450.

[30] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in Proc. Int. Conf. Spoken Lang. Process., 1998, vol. 7, pp. 2819–2822.

[31] D. Klatt, "Prediction of perceived phonetic distance from critical band spectra," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,1982, vol. 7, pp. 1278–1281.

[32] "Application guide for objective quality measurement based on recommendationsP.862, P.862.1 and P. 862.2," ITU-T Rec. P. 862. 3, 2005.

[33] J. Tribolet, P. Noll, B. McDermott, and R. E. Crochiere, "A study of complexity and quality of speech waveform coders," in *Proc.* IEEE Int.Conf. Acoust., Speech, Signal Process., 1978, pp. 586–590.

[34] J. H. Friedman, "Multivariate adaptive regression splines (with discussion),"Ann. Statist., pp. 1–141.

[35] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low bit-rate speech coding systems," IEEE J. Sel. Areas Commun.,vol. 6, no. 2, pp. 262–273, Mar. 1988.

[36] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC, 2007.

[37] B. Grundlehner, J. Lecocq, R. Balan, and J. Rosca, "Performance assessment method for speech enhancement systems," in Proc. 1st Annu. IEEE BENELUX/DSP Valley Signal Process. Symp., 2005.

[38] The NOIZEUS database (2009). Available: http://www.utdallas.edu/~loizou/speech/noize