Improved Packet Loss Recovery using Interleaving for CELP-type Speech Coders in Packet Networks

Fatiha Merazka

Abstract— **In VoIP applications, packet loss is a major source of speech impairment. In this paper, a packet loss concealment scheme based interleaving is presented to improve speech quality deterioration caused by packet losses for code-excited linear prediction (CELP) based coders. We applied the proposed scheme to the ITU-T G729 8 kb/s speech coding standard to evaluate the performance of the proposed method. The perceptual evaluation of speech quality (PESQ) and enhanced modified bark spectral distortion (EMBSD) tests under various packet loss conditions confirm that the proposed algorithm is superior to the concealment algorithm embedded in the G729. The spectral distortion measure is also used as an objective distortion measure; the obtained results prove that the interleaving method is better at the expense of extra delay.**

Index Terms— **VoIP, ITU G729, interleaving concealment, Spectral distortion measure, EMBSD, PESQ**

I. INTRODUCTION

Packet switched telephony, voice over IP (VoIP) in particular, has gained great popularity over recent years essentially due to its low cost and relative ease of deployment. Unfortunately, the quality of service (QoS) has not yet reached a level equivalent to that offered by the traditional public switched telephone network (PSTN). One of the most difficult problems inherent in such networks is the packet loss issue (also known as frame erasure). Even a single missing packet may generate an audible artifact in the decoded speech signal. To reduce the effect of packet loss on perceived speech quality, the missing packets have to be regenerated at the receiver using packet loss concealment (PLC) algorithms. For pulse code modulation (PCM), specifically the G.711 coder and decoder (codec), techniques based on waveform substitution have been used with fair success. Those techniques were first proposed by Goodman *et al.* [1]. Algorithms that were originally designed for time-scale modification were also adapted for packet loss concealment [2]. The standardized PLC algorithm in G.711 [3, App. I] is undoubtedly one of the most successful implementations of PLC for PCM coders.

Packet loss for code-excited linear prediction (CELP) codecs is of more concern because of the extensive use of prediction in such codecs. The PLC algorithms designed for CELP-based codecs are generally able to conceal the frame erasure relatively well. However, when the decoder starts receiving good frames again, the decoder is no longer synchronized with the encoder. Specifically, for predictive codebooks, since the memories are corrupted, the decoded parameters will be erroneous even if the received codebook indices are correct. This can cause the error to propagate over several frames before the decoder retrieves its synchrony with the encoder.

In addition to existing packet loss concealment procedures in standardized CELP-based speech codecs [4], [5] several techniques, such as the algorithms proposed in [6] and [7], have been proposed to improve the concealment.

Many error concealment algorithms for CELP type coders were proposed in order to minimize the quality degradation and the error propagation problem. Some of them tried to accurately estimate the excitation signal of the missing packets by a voicing classification [8][9]. Others efficiently estimated the gain parameters of the lost and successive frames [10][11]. However, few works have been dedicated to reducing the error propagation caused by the adaptive codebook.

In this paper, we present an interleaving concealment scheme for CELP based coders. We apply this method to the ITU-TG729 Conjugate-Structure Algebraic CELP (CS-CELP) speech coder [12] that is widely used in VoIP applications.

We compare the performance of the proposed algorithm with embedded standard method by measuring the average spectral distortion of Line Spectrum Frequencies (LSF) [13][14] before interleaving and after frame concealment.

We also use objective quality estimation algorithms among them the perceptual evaluation of the speech quality (PESQ) [14] and the enhanced modified bark spectral distortion (EMBSD) [15].

The remainder of the paper is organized as fellows. In section 2, we briefly review frame erasure concealment algorithm embedded in the ITU-T G729 standard speech coder. The proposed method is presented in section 3. Simulation of the packet loss is presented in section 4. Comparison and evaluation results are presented in section 5. Section 6 concludes the paper.

II. FRAME ERASURE CONCEALMENT OF G729

In the G729 speech coder, an erased frame is reconstructed using the speech coding parameters of the previous received good frame [12]. Once frame erasure is detected, the new

Manuscript received Dec 10, 2008. Dr. Fatiha. Merazka Author is with the Electronic & Computer Engineering Faculty University of Science & Technology Houari Boumediene, P.O.Box 32, El Alia, 16111 Algiers, Algeria phone: 213-21-247187; fax: 213-21- 247187; e-mail: fmerazka@hotmail.com).

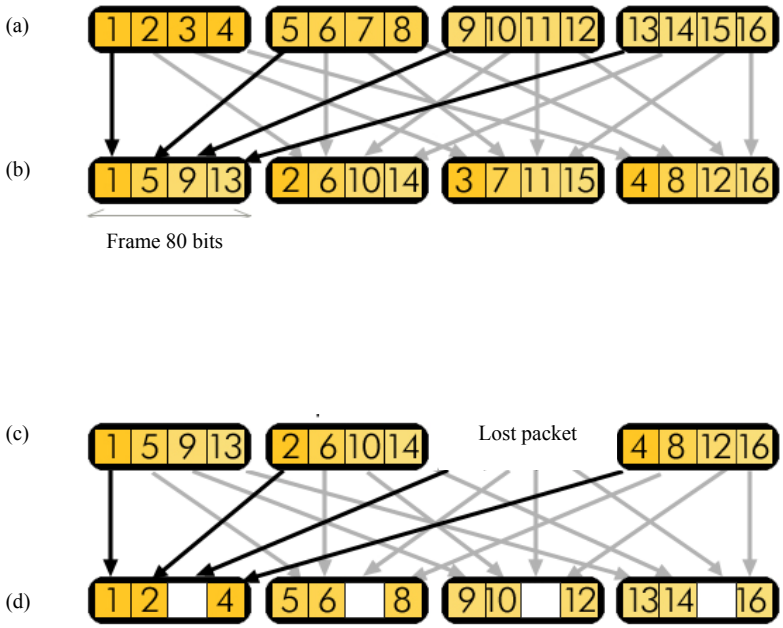

Fig. 1. Interleaving packet concealment scheme, (a) original frames, (b) frames interleaved, (c) frame loss, (d) reconstructed frames.

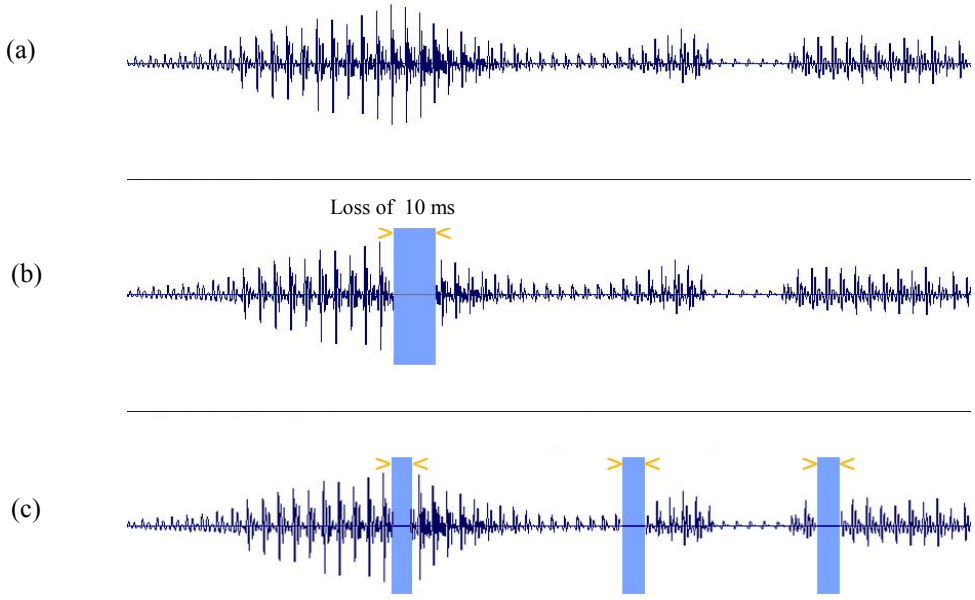

Fig. 2. Example of speech quality degradation due to frame loss, (a) original speech, (b) frame loss with G729, (c) reconstructed speech by interleaving

parameters are generated by analyzing the spectral parameters of the last good speech frame. The method replaces the missing excitation signal of the erased frame by taking one of the similar characteristics, while gradually decaying its energy.

If n-th frame is detected as an erased frame, the G.729 repeats the spectral parameters of the last received good frame to the erased frame. In addition, an adaptive codebook gain and a fixed codebook gain are obtained by multiplying predefined attenuation factors by the gains of the previous frame.

To avoid excessive periodicity a long term prediction lag is increased by one to the value of the previous frame. The main reason that the speech coding parameters of the erased frame are basically assigned with slightly different or scaled down values from the previous good frame is to prevent from generating a reverberant sound. However this simple scaling down approach causes a fluctuation of an energy trajectory for the decoded speech and brings a more annoying affect to the listeners when longer frames are erased [10].

III. Description of the Interleaving Concealment Method

In interleaving method, the data in N consecutive frames can be mixed together before transmission [16]. This way, the loss of a packet destroys only a few bits from each frame. Assuming the coder is more robust to bit errors than frame erasures (which is generally true), this approach may reduce the effect of loss. However, it does so at the expense of the substantial delays. Fig.1 shows 4 interleaved frames before and after transmission. At the coder side, we have 4 frames of 80 bits. Each frame is divided into 4 sub-frames of 20 bits each and interleaved as shown in Fig.1.

The first sub-frames of each frame are grouped to form the fist frame. The second sub-frames of each frame are concatenated to form the second frame. The third sub-frames of each frame are concatenated to form the third frame. Finally, the fourth sub-frames of each frame are concatenated to form the fourth frame. By doing so, we prevent a loss of a sequence of samples in case of a single packet loss. After transmission, the loss of a single packet from an interleaved stream results in multiple small gaps in the reconstructed stream.

Fig. 2 shows an example of speech quality degradation when frame loss is occurred. A frame of 10 ms is erased. Interleaving method spread the loss in small gaps. The proposed method gives an improved waveform shape.

IV. Simulating Packet Loss

We simulate real-time voice over packet networks where each packet contains one frame. Packet losses are not independent on a frame-by-frame basis, but appear in bursts. Bolot [17] studied the distribution of packet loss in the Internet and concluded that this could be approximated by a Markovian loss model such as the Gilbert or Elliott models. Thus, we have simulated the IP network by using a 2-state Markov model, known also as a Gilbert model as in Fig. 3.

Let state "0" stand for a packet being correctly received and "1" be a packet being erased. Let the p be the transition probability from "0" to "1" and q be the probability from "1" to "0" and five loss rates are simulated as given in Table I.

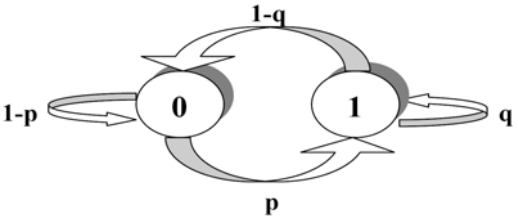

Fig. 3. Two-state Markov model.

Table I. Simulated loss rates

rate(%)	p	q
00	0.00	0.00
10	0.10	0.15
20	0.20	0.30
30	0.30	0.35
40	0.30	0.40

V. Experimental Results

In this section we compare the performance of the proposed method with that of the embedded method in the G729 standard.

We use the spectral distortion (SD) measure as an objective distortion measure expressed in dB and given by the following equation:

$$SD=\frac{1}{N_f}\sum_{n=1}^{N_f}\left(\frac{1}{\pi}\int_0^{\pi}\left[\log S_n(w)-\log\hat{S}_n(w)\right]^{1/2}dw\right),$$

where $S_n(w)$ and $\hat{S}_n(w)$ are the spectra of the nth speech frame without quantization and with quantization, respectively and N_f is the total number of frames. The spectral distortion measure is known to have a good correspondence with subjective measure [18].

To achieve transparent quality quantization average SD must be about 1 dB with less than 2% outliers in the range 2-4 dB, and no outlier with SD greater than 4 dB [19].

All original speech for male and female speakers is taken from TIMIT database [20].

Figs. 4 and 5 show the line spectrum frequencies (LSF) [21] [22] performance under several loss rates for female and male speakers respectively [10].

Outliers are tabulated in Tables II and III for male and female speakers respectively.

Table II. Outliers of LSF spectral distortion with packet loss male speakers

Loss Rate (%)	Original G729			4 frames interleaved		
	Av. Spec. Dist (dB)	Outliers (%)		Av. Spec Dist (dB)	Outliers (%)	
		2-4 (dB)	>4 (dB)		2-4 (dB)	>4 (dB)
0	1,22	6,95	0,05	1,22	6,95	0,05
10	2,94	8,51	2,54	2,54	7,14	1,95
20	3,96	35,05	23,85	3,63	28,80	18,94
30	4,63	45,65	26,25	4,25	40,34	26,25
40	5,05	47,80	26,30	4,67	44,11	26,30

Table III. Outliers of LSF spectral distortion with packet loss female speakers

Loss Rate (%)	Original G729			4 frames interleaved		
	Av. Spec. Dist. (dB)	Outliers (%)		Av. Spec Dist. (dB)	Outliers (%)	
		2-4 (dB)	>4 (dB)		2-4 (dB)	>4 (dB)
0	1,21	6,70	0,20	1,21	6,70	0,20
10	2,72	23,85	9,95	2,52	19,63	8,83
20	3,94	39,40	14,80	3,73	32,27	13,96
30	4,85	45,85	17,85	4,59	41,38	16,48
40	5,34	47,80	21,30	5,06	44,18	19,71

These results from tables II and III show that up to 0.4 dB and 0.28 dB improvements are obtained on average spectral distortion over the original G729 for male and female speakers respectively.

The number of outliers is substantially reduced under frame erasures for both male and female speakers.

We notice that the average SD is greater for female than male speakers under the same lost rates.

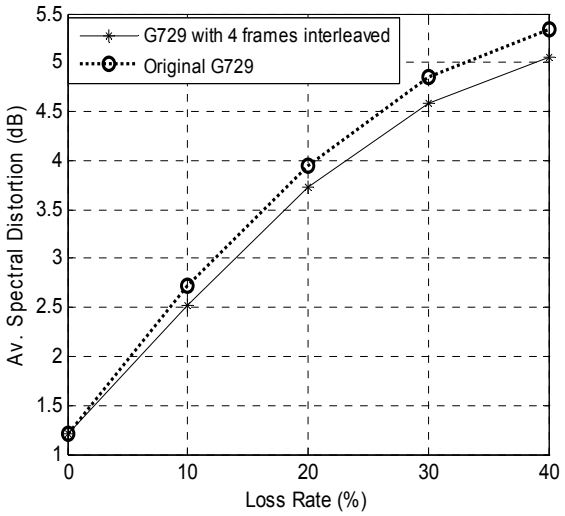

Fig. 4. Comparison of Average Spectral distortion for G729 decoded female speakers with embedded method (dash line) and the proposed method (4 fames interleaved) (solid line) under different loss rates.

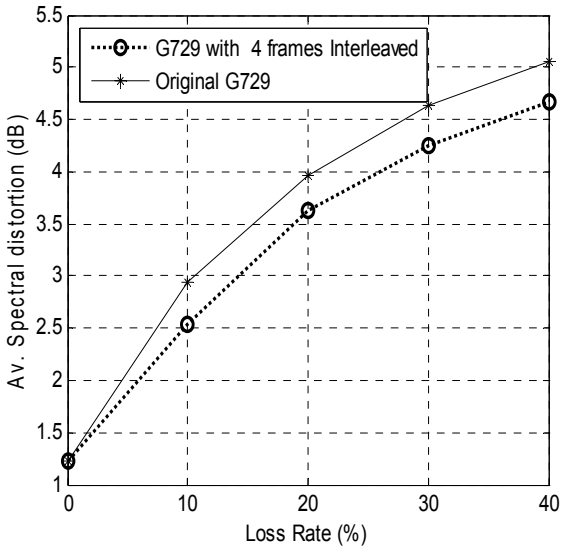

Fig. 5. Comparison of Average Spectral distortion for G729 decoded male speakers with embedded method (dash line) and the proposed method (4 fames interleaved) (solid line) under different loss rates.

We use PESQ for an objective quality measure. The PESQ gives a value between 0 (no similarity) and 4.5 for two speech file similar.

Figs. 6 and 7 show comparison results obtained for female and male speakers respectively.

As the packet loss rate increases, the PESQ scores of the two algorithms decrease. The scores of the proposed algorithm are higher than the embedded method in the G729 standard coder.

We notice that the PESQ score is less than 1.5 from 15% of loss rate for female speakers but for male speakers it reaches this score from 20% to 25% of loss rate.

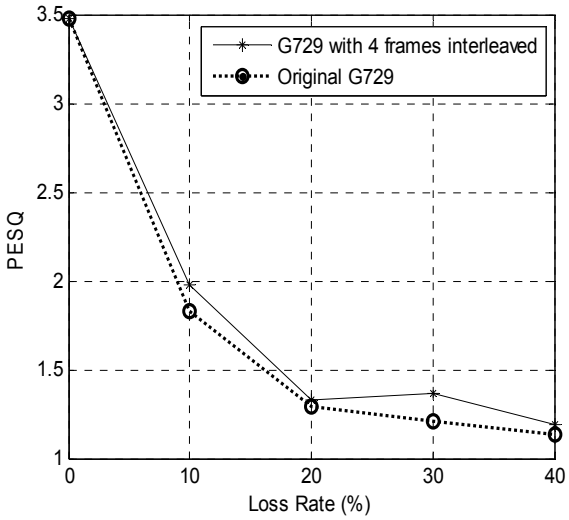

Fig. 6. Comparison of PESQ for female speakers decoded with original G729 (dash line) and the proposed method (4 fames interleaved) (solid line) under different loss rates.

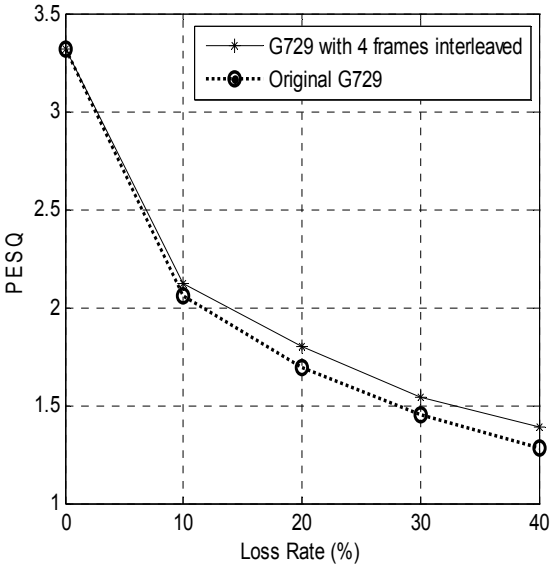

Fig. 7. Comparison of PESQ for male speakers decoded with original G729 (dash line) and the proposed method (4 fames interleaved) (solid line) under different loss rates.

We also performed EMBSD tests and the results are depicted in Figs. 8 and 9 for female and male speakers respectively. The EMBSD is value of 0 for two similar speech files and a greater value as the distortion increases.

We can see from Figs 8 and 9 that as the packet loss rate increases, the EMBSD of the two methods increase.

For female speakers the improvement is up to 3.96 but for male speakers it is up to 4.06.

It is shown that the proposed algorithm is always better than the embedded method in the G729 standard coder.

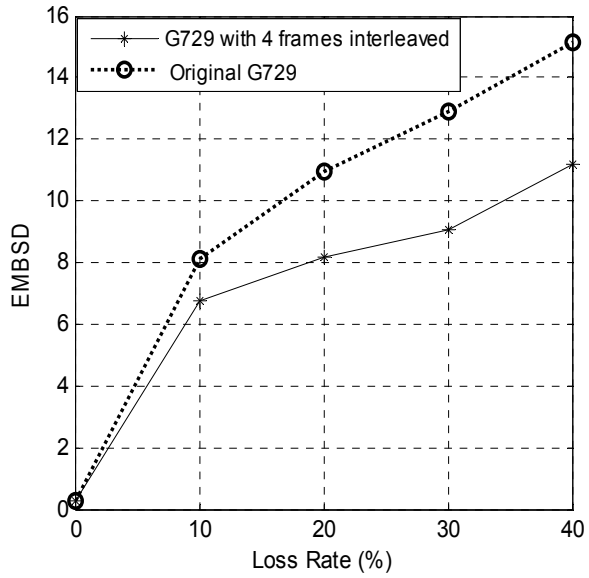

Fig. 8. Comparison of EMBSD for female speakers decoded with original G729 (dash line) and the proposed method (4 fames interleaved) (solid line) under different loss rates.

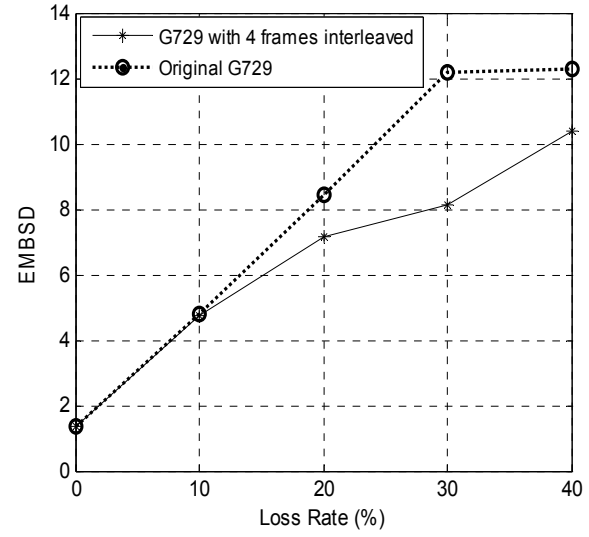

Fig. 9. Comparison of EMBSD for male speakers decoded with original G729 (dash line) and the proposed method (4 fames interleaved) (solid line) under different loss rates.

VI. CONCLUSION

In this paper, an efficient method for reconstructing the missing frames for CELP based coders is presented. The method consists in interleaving 4 speech frames in order to spread out the error. Its performance was compared with the embedded algorithm in the standard G729 coder.

The objective measures given by the average spectral distortion measure verify that the interleaving method is better at the expense of extra delay.

From PESQ measurement and EMBSD tests under a variety of frame erasure conditions, we found that the proposed method, improved significantly the speech quality compared to the embedded algorithm in the standard G729 coder.

We found that the distortion obtained with female speakers is greater than that obtained with male speakers. Effectively, female speakers are characterized by higher frequency than male speakers.

REFERENCES

[1] D. J. Goodman, G. B. Lockhart, O. J.Wasem, and W.-C.Wong, "Waveform substitution techniques for recovering missing speech segmentsin packet voice communications," IEEE Trans. Acoust., Speech, Signal *Process.*, vol. ASSP-34, no. 6, pp. 1440–48, Dec. 1986.

[2] H. Sanneck, A. Stenger, K. B. Younes, and B. Girod, J. Crowcroft and H. Schulzrinne, Eds., "A new technique for audio packet loss concealment,"in Proc. IEEE Global Internet 1996, London, U.K., Nov. 1996,pp. 48–52.

[3] A High "Quality low-complexity algorithm for packet loss concealment with G.711," Int. Telecom. Union (ITU), Geneva, Switzerland, 1999, Rec. ITU-T G.711, App. I.

[4] R. Salami, C. Laflamme, J. P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and description of CS-ACELP: A toll quality 8 kb/sspeech coder," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp.116–130, Mar. 1998.

[5] "Adaptive multi-rate-wideband (AMR-WB) speech codec: Error concealment of lost frames," Jun. 2007, 3GPP Tech. Spec. 3GPP TS 26.191.

[6] J.-F. Wang, J.-C. Wang, J.-F. Yang, and J.-J. Wang, "A voicing-driven packet loss recovery algorithm for analysis-by-synthesis predictive speech coders over internet," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 98–107, Mar. 2001.

[7] J.-H. Chen, "Frame Erasure Concealment for Predictive Speech Based on Extrapolation of Speech Waveform," U.S. Pub. No. US 2003/0078769 A1, 2003, U.S. Patent Application Publication.

[8] A. Huaain and V. Cupeman, "Reconstruction of missing packets for CELP-based speech coders," in Proc. ICASSP-95, vol. 1, 1995, pp.245-248.

[9] Jhing-Fa Wang, Jia-Ching Wang, Jar-Ferr Yang and Jian-Jia Wang, "A voicing-driven packet loss recovery algorithm for analysis-by-synthesis predictive speech coders over Internet."in Multimedia, IEEE Transaction , vol. 3, pp.98-107, March 2001

[10] Hong Kook Kim and Hong-Goo Kang, "A Frame Erasure Concealment Algorithm Based on Gain Parameter Reestimation for CELP coders," in 1EEE Signal Processing Letters, vol. 8, pp.252-256, Sept 2001.

[11] De Martin, J.C, Unno, T. and Viswanathan, V. ''Improved frame erasure concealment for CELP-based coders," in Proc.ICASSP'00, vol. 3, pp.1483-1486.

[12] ITU, ITU-T G.729: CS-ACELP Speech Coding at 8 kbit/s, ITU 1998.

[13] F.Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", J.Acoust. Soc. Amer., vol. 57, suppl. 1, p. S35(A),1975.

[14] W. Yang, "Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measurement Based on Audible Distortion and Cognition Model," Ph.D Dissertation, Temple University, USA, May 1999.

[15] ITU-T Draft Rec P.862 "Perceptual evaluation of speech quality (PESQ), an objective method of end-to-end speech quality assessment of narrowband telephone networks and speech codecs," May. 2000.

[16] J. L. Ramsey, "Realization of optimum interleavers," IEEE Trans. Info. Theory, vol. IT-16, May 1970, pp. 338–45.

[17] J.C. Bolot, "End-to-end frame delay and loss behavior in the Internet," in *Proc. ACM SIGCOMM*, Sept. 1993, pp. 289–298.

[18] Y. Kitawaki, K. Itoh. and K. Kakehi, "Speech quality measurement methods for synthesized speech," *Review of ECL.,*. Vol. 29 no. 9-1. NTT Japan Sept-Dec. 1981.

[19] K. K. Paliwal and B.Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *ICASSP*, pp. 661-664, Mar. 1991.

[20] NIST,Timit Speech Corpus, NIST 1990.

[21] F.Itakura, \Line spectrum representation of linear predictive coefficients of speech signals", J.Acoust. Soc. Amer., vol. 57, suppl. 1, p. S35(A),1975.

[22] F.K. Soong and B. Juang, \Line spectrum pair(LSP) and speech data compression", Proc.IEEE Int. Conf. Acoustics, Speech, Signal Processing,San Diego, CA, 1984, pp. 1.10.1-1.10.4.