Intra-frame Quantization of LSF Parameters FOR CELP-Based Coders in Packet Networks

Fatiha Merazka

Abstract— This paper presents an intra-frame quantization of Line Spectrum Pairs (LSP) parameters with interpolative concealment method to improve frame erasures for ITU-T G729 coder. The standard ITU-T G729 coder uses an inter-frame quantization of the LSF parameters which causes error propagation to the next frames. First a comparison study is performed in order to find the suitable splitting in terms of spectral distortion measure and bit rate. The intra-frame 4-6 splitting is applied to the G729 with an interpolative concealment method since in voice over internet protocol (VoIP) applications one or more frame, at least most of time, are present in the so called playout buffer. Comparison is performed with the embedded method of the G729. Simulations results show that the intra-frame quantization with interpolative concealment achieve smaller average spectral distortions than that of the embedded in the G729. Enhanced modified bark spectral distortion (EMBSD) tests under various packet loss conditions confirm that the proposed method is superior to the concealment algorithm embedded in the G729.

***Index Terms*—VoIP, ITU G729, Error propagation, Intraframe quantization, Interpolative concealment, Spectral distortion measure, EMBSD.**

I. Introduction

When sending real-time speech packets through IP networks, there is no guarantee of receiving the transmitted packets in a timely manner due to the best effort nature of the networks. When one or more packets are missing and no effort is made to recover those packets, the perceptual quality of the received speech can significantly get worse. A lot of schemes have been proposed to improve this effect and they are often classified as encoder-based or decoder-based [1]-[6].

Most packet loss concealment (PLC) algorithms embedded in the standards speech coders are based on an extrapolation method or a repetition method in| which the speech coding parameters are extrapolated or repeated from the parameters of the last good frame received. Since the lost packet causes the corruption of the long term prediction memory, extra performance degradation may occur from the use of the incorrect memory even at the received frames in the future.

Forward Error Concealment schemes are effective when the network loss is predictable and extra bandwidth is provided. Decoder based concealment is of relevance for bandwidth limited applications. Coded linear prediction (CELP) coded speech frames are adequate for this technique since many coding parameters show good smoothness between frames. Some ITU speech coders [7]-[8] have built-in mechanisms that process the erased frames based on predictive recovery. However, both of the coders G723.1 and G729 quantize Line Spectrum Frequencies (LSF) parameters via predictive methods. Predictive concealment can cause error propagation to subsequent frames; this phenomenon is illustrated in Figure 1, where LSF spectral distortions of a coded speech with and without frame erasure are plotted for comparison.

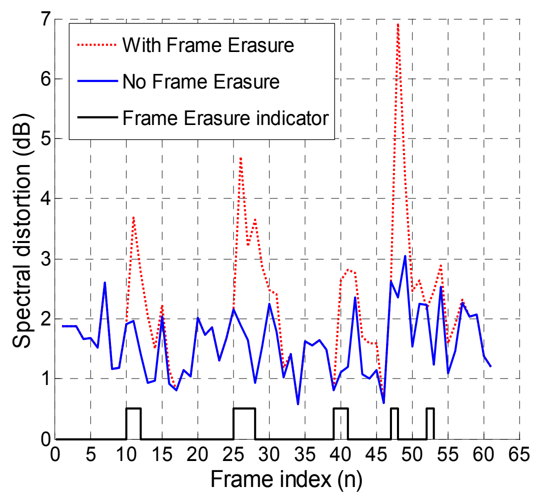

Figure 1. LSF spectral distortion error propagation of G.729 coded speech

From figure 1, we observe that for several frames following an erasure, the two spectral distortion curves diverge from each other, indicating propagated distortion error.

We propose improvements over standard practice for G729 by the use of intraframe quantization of LSF parameters and interpolative concealment that uses past and future information.

This paper is organized as follows. In section 2, an intrafame coding method is presented. In section 3 the performance results of the interpolative concealment with intraframe quantization are presented. The conclusion is given in section 4.

II. Intraframe Quantization

LSF parameters are well known for their ordering property [9], which states that within each frame, LSF's are strictly in

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- 24 71 87; e-mail: fmerazka@hotmail.com).

ascending order with their indexes as shown in figure 2. We can see from this figure that LSF in medium frequencies are more variable than the LSF at high and low frequencies. They are also known for their intraframe and interframe correlation.

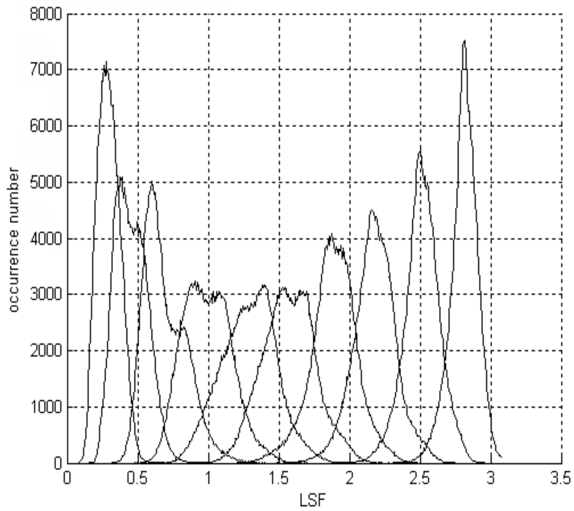

Figure 2.Histograms of LSF parameters from 1 to 10.

The localized sensitivity property of the LSFs makes them ideal for split vector quantization as the individual parts of an LSF vector can be independently quantized without a leakage of quantization distortion from one spectral region to another [10].

In order to find the optimal partition of the LSP vectors, an intraframe correlation has been calculated for 229829 LSP vectors [11], namely the correlation between LSF_i a nd LSF_j within the same frame for i, j= 1, 2, …., 10.

The intraframe correlation coefficients are represented in Table I. These results show that the correlation between consecutive LSF is considerable. The common division method as reported in the literature [11] is the split 3-3-4 VQ in which the first subvector contains the first three lowest LSF components of the vector the second subvector contains the three middle LSF's while the final subvector contains the four highest LSF's. However, according to table I, the fourth LSF is more correlated with the third one rather than the fifth. Moreover, we have identical correlation between the fourth LSP and the first one as well as between the fourth LSF and the sixth. From these correlations it can be deduced that the fourth LSF must be displaced from the second subvector to the first one. Furthermore, the correlation between the eighth LSF and ninth is feeble, consequently the eighth LSF must be displaced from the third subvector to the second one.

From the above analysis, it can be deduced that the division of the form of 4-4-2 is better than that of 3-3-4.

TABLE I. CORRELATION BETWEEN LSF_i AND LSF_j OF THE SAME FRAME

j	i x 10^{-3}									
	1	2	3	4	5	6	7	8	9	10
1	1000	721	427	472	69	15	94	104	95	-9
2	721	1000	772	576	323	274	325	364	276	195
3	427	772	1000	745	480	491	450	509	411	300
4	472	576	745	1000	728	512	490	432	441	259
5	69	323	480	728	1000	775	586	491	335	279
6	15	274	491	512	775	1000	757	629	456	301
7	94	325	450	490	586	757	1000	740	525	399
8	104	364	509	432	491	629	740	1000	606	398
9	95	276	411	441	335	456	525	606	1000	533
10	-9	195	300	259	279	301	399	398	533	1000

The quantizers of the subvectors are trained using the GLA for different bit rates. As the ITU G.729 uses 18 bits/frame for the quantization of the LSF parameters, we started our simulation with 18 bits/frame in order to find a average spectral distortion (Av. SD) that is equal or better than the standard's one. The Av. SD computed for the test database is 1.543 dB, but for 18 bits/frame and 19 bits/frame we found that the spectral distortions are greater than 1.8 dB. The results obtained from our simulations in terms of spectral distortion which are tabulated in Table II are close to the spectral distortion of the G.729.

TABLE II. AV. SD FOR DIFFERENT SPLITTING AND DIFFERENT BIT ALLOCATIONS

Number of bits	Splitting	Bit allocation	Av. SD (dB)
20	3-3-4	7-6-7	1.679
20	4-4-2	9-9-2	1.645
20	4-4-2	9-8-3	1.556
20	4-6	10-10	1.503

From the above results, we can see that division 4-6 with a bit allocation of 10-10 bits respectively, is the best one since it gives an Av. SD of 1.503 dB which is better than the standard's at 18 bits/frame with additional 2 bits/frame

III. INTERPOLATIVE CONCEALMENT METHOD WITH SVQ FOR G729

In consideration of a simple characterization of the behavior of the network, the well-known Gilbert model is used to approximate packet loss [4] that is shown in figure 3 and it emphasizes the 'bursty' nature of the Internet packet loss.

Let state "0" represents a packet that is being correctly received and state "1" a packet that is being erased. The probability p corresponds to the transition from state "0" to state "1" while probability q is the transition from state "1" to state "0". We have simulated four loss rates as listed in table III.

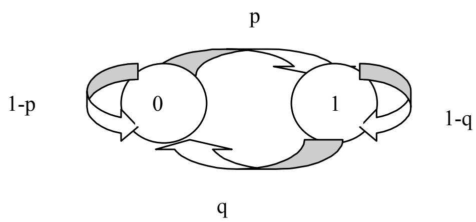

Figure 3. Two-state Gilbert model.

TABLE III. SIMULATED LOSS RATES

rate(%)	*p*	*q*
0	0	0
10	.1	.85
20	.2	.77
30	.3	.65
40	.3	.50

An interpolative concealment reconstruction was applied to G729 with the SVQ 4-6 as described above. The decoder was modified so that if a frame erasure occurs, and if the next frame is not lost as well, interpolative concealment is applied instead of the embedded method in G729.

The LSF parameters are linearly interpolated from previous and next good frames. The pitch lag and gain information are linearly interpolated from previous and next good frames. Voicing decision is used same as in the standard G729.

Figure 4 shows the performance of the interpolative reconstruction method with intraframe coding which is compared with G.729 predictive coding. The Av.SD outliers which are important parameters that affect the perceptual quality of the decoded speech are tabulated in Table VI. We noticed that with extra 0.2 kb/s which corresponds to 2.5 % of the total bit rate, the LSF coding method with interpolative concealment achieve 0.3-1.1 dB lower Av. SD. The percentage of the outliers is also much smaller and yields significant perceptual improvement in the occurrence of erasures.

The histograms corresponding to the spectral distortions of the standard G.729 and G729 with interpolative concealment and SVQ are plotted in figure 5. It can be noticed from figure 5 that most of the lost frames are interpolated with small distortions while using the intraframe SVQ.

We have used Enhanced modified bark spectral distortion (EMBSD)[12] tests under various packet loss conditions for perceptual speech quality measurement. Figure 6 shows the perceptual distortions evaluated by EMBSD for different loss rates.

It is certain that with extra 0.2 kb/s our proposed method achieves up to 2.5 perceptual distortions (EMBSD) better than the original G.729. The results from figure 6 are compared with those of table VI.

It can be concluded that our method gives a “good” quality up to 30% of the loss rate which corresponds to a fair quality of the G.729 standard.

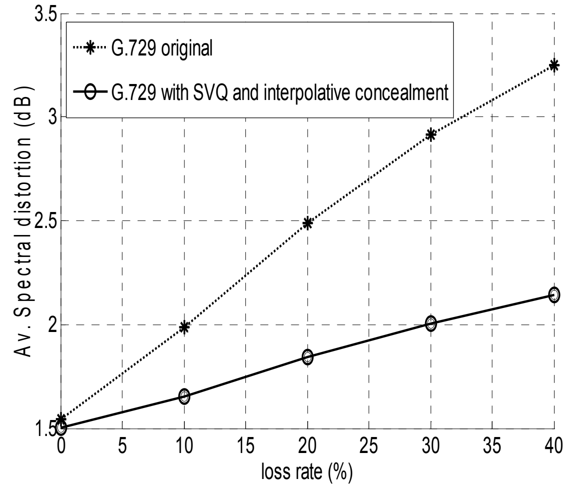

Figure 4. Average LSF spectral distortion with SVQ and interpolative concealment.

TABLE VI: FRAME ERASURE PERFORMANCE OF SVQ 4-6 FOR DIFFERENT LOSS RATES

frame loss (%)	Original G729			SVQ 4-6 with interpolative concealment		
	Av. SD (dB)	Outliers (%)		Av. SD (dB)	Outliers (%)	
		2-4 dB	> 4dB		2-4 dB	> 4dB
0	1.543	19.60	0.62	1.503	13.89	0.02
10	1.954	26.21	6.59	1.685	18.36	2.54
20	2.359	32.90	12.43	1.861	22.94	4.90
30	2.777	39.63	18.43	2.045	27.45	7.55
40	3.249	56.15	23.72	2.141	32.55	8.02

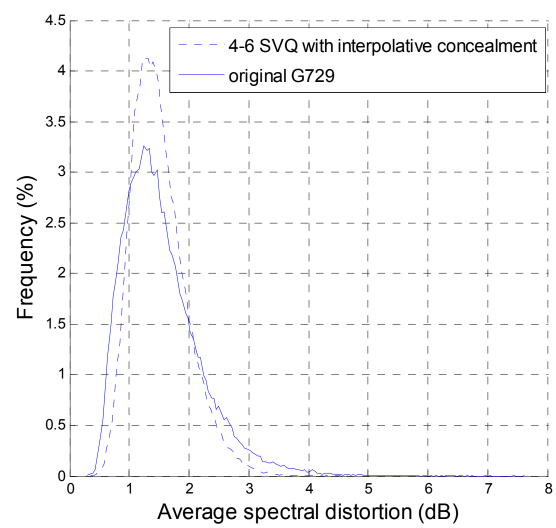

Figure 5. Histograms of interpolated LSF spectral distortion

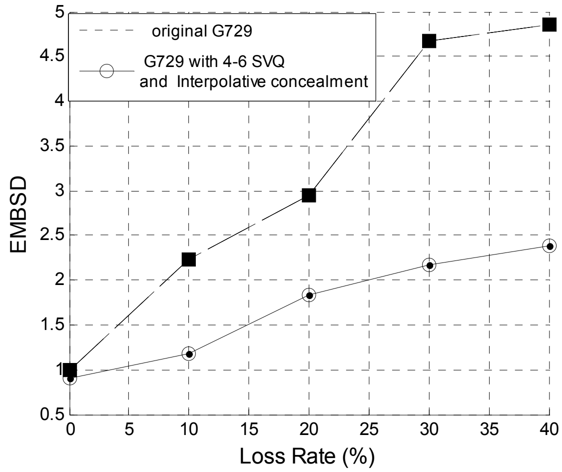

Figure 6. EMBSD for different lost rates.

IV. CONCLUSION

In this paper we have presented an efficient method for reconstructing the missing frames for G729. We used an intraframe 4-6 SVQ quantization scheme for the LSF parameters and interpolative concealment method. We noticed that with an extra 0.2 kb/s due to using intraframe quantization, which corresponds to 2.5 % of the total bit rate, we can improve the standard G729 .

REFERENCES

[1] N. Jayant and S.W. Christensen, "Effect of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure," IEEE Trans. Commun. Vol. 29 NO. 2, Feb. 1981.

[2] J-C. Bolot, "Analysis and control of audio packet loss in the Internet," NOSSDAV 95.

[3] M. Podolsky, C. Romer and S. McCanne, "Simulation of FEC-based error control for packet audio on the Internet," Proceedings IEEE Infocom, vol.2, pp. 505-515. April 1998.

[4] J. C. Bolot, S. Fosse-Parisis, and D. Towsley, "Adaptive FEC-Based Error Control for Interactive Audio in the Internet," Proceedings IEEE Infocom 1999, New York, NY, March 1999.

[5] D. Goodman and G. Lockhart etc, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," IEEE Trans. On Acoustics, Speech, and Signal Processing, vol. ASSP-34, No. 6, pp. 1440, Dec. 1986.

[6] W. R. Erhart and J. D. Gibson, "A speech packet recovery technique using a model based tree search interpolator," IEEE workshop of speech coding for telecommunications, Sainte-Adele, Quebec, Canada pp. 13-15, Oct. 1992.

[7] ITU, ITU-T G.723.1: Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, ITU 1996.

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

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

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

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

[12] S. Voran, "Objective estimated of perceived speech quality-part II: evaluation and measuring normalized block technique," IEEE Trans. Speech and audio processing, vol. 7, NO 4, pp. 371-382, July 1999.