Proceedings of the World Congress on Engineering 2009 Vol I
WCE 2009, July 1 - 3, 2009, London, U.K.

LSF Quantization to Enhance the Frame Erasure Robustness of CELP Type Coders in Packet Networks

Fatiha Merazka

Abstract— **Line Spectrum Frequencies (LSF) have been the current parameter set to represent LPC coefficients in speech coding. Extensive research has been performed to exploit their interframe and intraframe correlations and quantize them more efficiently. Interframe coding of LSF's can cause error propagation when frame erasures occur. Since most LSF quantizers were designed with the primary concerns of bit-rate and complexity, less attention was paid to error propagation. In this paper, we consider the erasure performance of LSF differential scalar quantizer (DSQ) and compare it with the interframe coding method embedded in the standard G723.1 of the ITU. Our results show that with only 5% extra bit-rate, DSQ algorithm is much more robust to frame erasures and improvements in terms of spectral distortion and Enhanced modified bark spectral distortion (EMBSD) tests under various packet loss conditions are obtained.**

Index Terms—**Error propagation, VoIP, ITU G723.1, LSF, quantization, EMBSD, Spectral Distortion.**

I. INTRODUCTION

The quality of real time voice communications in packet switched transmission or mobile links is degraded by frame erasures. In voice communication over IP networks, the packet loss is caused by the transmission impairments such as the process of the transmission capacity and congestion. Since even a single missing packet may generate an audible artifact in the decoded speech signal, the receiver needs a packet loss concealment method to minimize the quality degradation at the packet loss regions.

With the emergence of voice over packet networks, erasure-robustness has become an important problem for coder performance. Coding standards designed for these applications are represented by the ITU dual rate voice coder G.723.1 [1] and ITU toll quality coder G.729 [2]. Especially G.723.1 has been built into many Internet applications. However, both coders inherited the interframe predictive split vector quantization (PSVQ) coding of line spectrum pairs (LSF's) [3] from previous development where the major concerns were on bit-rate and complexity. With interframe coding, when a frame erasure occurs, the decoder states change and thus cause error propagation. Although most coders have the forgetting ability to smooth out the erased frames, at least 2-3 frames are affected. This occurrence is illustrated in Figure 1, where LSF spectral distortions of a coded speech signal for the standard G723.1 with and without frame erasure are plotted for comparison. 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.

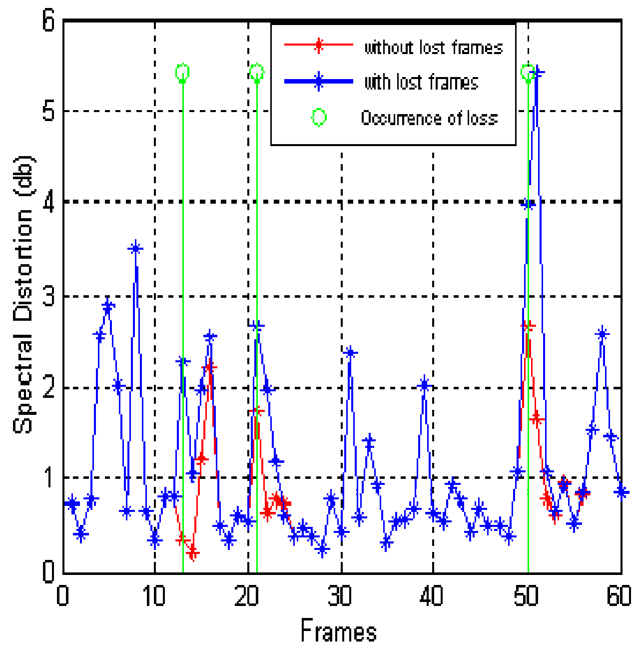

Figure 1. LSF spectral distortion error propagation of G.723.1 coded speech

This paper is organized as follows. In section 2, LSF statistics are presented. In section 3, we present DSQ for LSF parameters. Performance comparison and results are given in section 4 and section 5 concludes the paper.

II. LINE SPECTRUM FREQUENCIES

LSF parameters are popularly known for their ordering property, which states that within each frame, LSF's are mainly in ascending order with their indexes [4]. They are also known for their intraframe and interframe correlation.

We have investigated the statistics of the LSF and differential LSF namely ΔLSF parameters and results are reported in table I and II respectively. The corresponding

Manuscript received March 23, 2009. 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).

histograms for LSF and ΔLSF respectively and shown by figure 2 and 3. The ΔLSF are obtained as fellows:

$$\begin{cases} \Delta LSF[1] = LSF[1] \\ \Delta LSF[i] = LSF[i] - LSF[i-1] \ldots for\ i \neq 1 \end{cases} \quad (1)$$

TABLE I. STANDARD DEVIATION AND DYNAMIC OF LSF PARAMETERS.

LSF	Standard deviation*10^{-4}	Dynamic* 10^{-4}
1	267	2001
2	353	2852
3	465	3458
4	547	3753
5	645	3890
6	603	4022
7	526	3961
8	469	3707
9	386	2982
10	285	2277

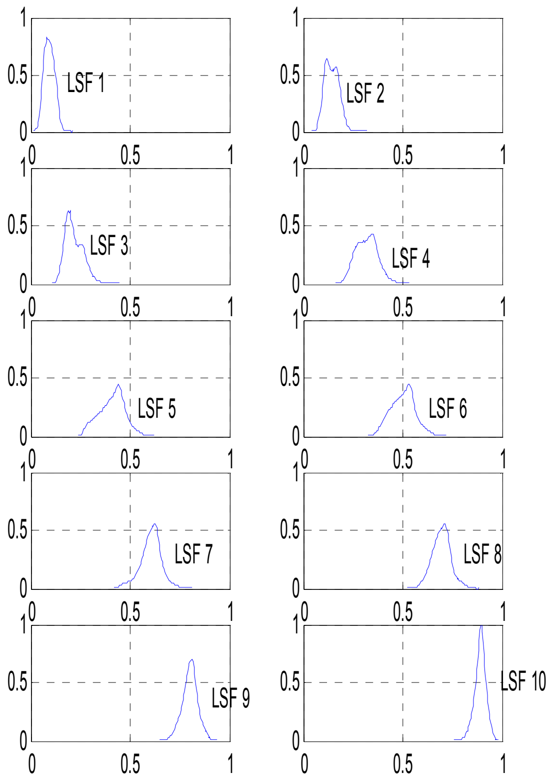

Figure 2. Histograms of LSF parameters

From figure 2, we can see that the distribution of LSF 1, 2, 9 and 10 is very close to a Gaussian distribution. The LSF 3, 4, …, 8 have wider dynamic range.

From figure 3 and table II the ΔLSF are observed to be less divergent with smaller dynamic ranges compared to the absolute LSF themselves, so it is better to quantize the ΔLSF rather than the LSF parameters.

TABLE II. STANDARD DEVIATION AND DYNAMIC OF ΔLSF PARAMETERS.

ΔLSF	Standard deviation*10^{-4}	Dynamic* 10^{-4}
1	267	2001
2	293	2468
3	307	2655
4	429	3257
5	472	3256
6	466	3291
7	436	3102
8	397	3129
9	410	2981
10	351	2624

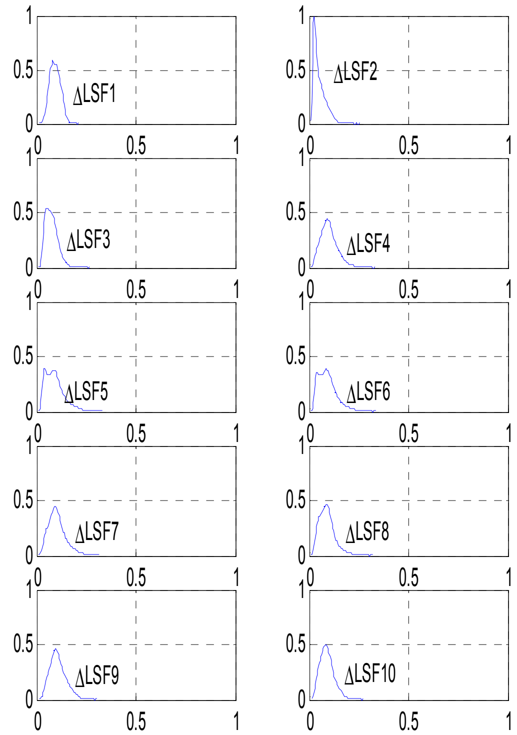

Figure 3. Histograms of ΔLSF parameters

III. LSF DIFFERENTIAL SCALAR QUANTIZATION

In this section we describe the DSQ used to quantize the LSF parameters for the standard ITU G 723.1. Figure 4 show DSQ method.

The procedure of DSQ is :

1. Quantize $\Delta LSF[1] \xrightarrow{Q_1} \hat{LSF}[1]$
2. Compute
3. Quantize $\Delta LSF[i] \xrightarrow{Q_i} \hat{LSF}[i]$
4. If i >10 $\Delta LSF[i] = LSF[i] - \hat{LSF}[i-1] \ldots for\ i \neq 1$
   stop,
   otherwise reconstruct
   $\hat{LSF}[i] = \hat{LSF}[i-1] + \Delta \hat{LSF}[i]$, go to 2.

ISBN: 978-988-17012-5-1

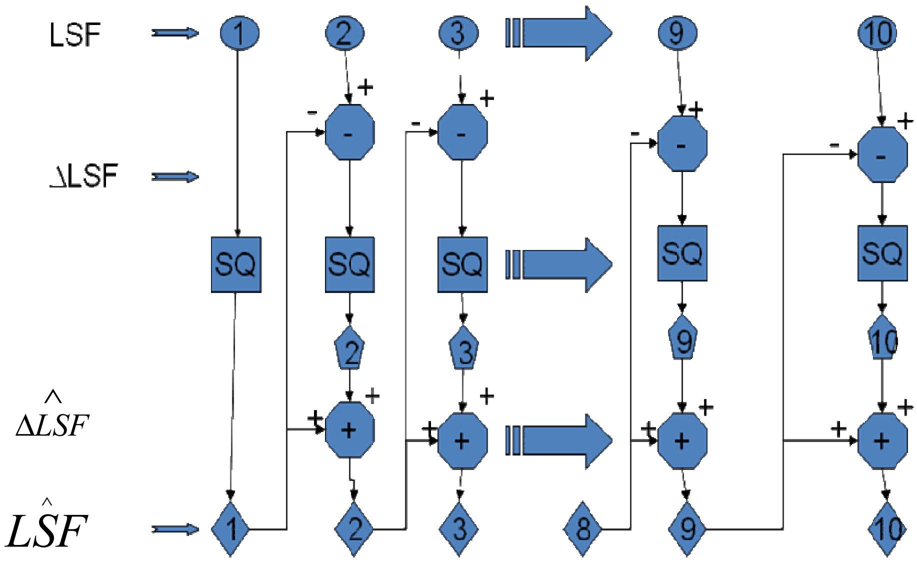

Figure 4 LSF Differential Scalar Quantization

IV. PERFORMANCE COMPARAISON AND RESULTS

In this section we compare the performance of DSQ method with that of the PSVQ embedded method in the G723.1 for speech from TIMIT database [5]. The performance comparison between DSQ and PSVQ are evaluated using the average spectral distortion (Av. SD) measure [4] and enhanced modified bark spectral distortion EMBSD [6], the results are depicted in table III.

TABLE III. PERFORMANCE COMPARISON BETWEEN DSQ AND PSVQ

	Av. SD (dB)	2–4 dB	> 4dB	EMBSD
PSVQ	1.84	32.99	4.67	1.551
DSQ	1.81	32.99	4.44	1.498

We simulate real-time voice over packet networks where each packet contains one frame. Packet loss is approximated by a Gilbert random process which emphasizes the bursty nature of Internet packet loss as in Figure 5. 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"

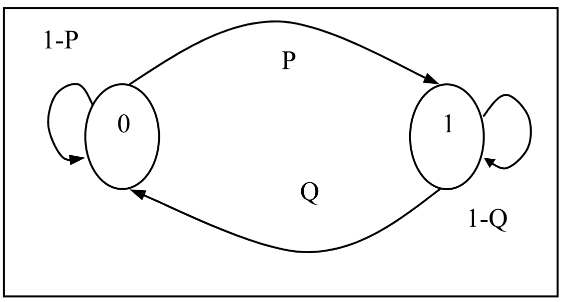

Figure 5. Two-state Gilbert model.

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 G723.1.

The bit allocation for DSQ is given in Table VI.
The LSF parameters are linearly interpolated from previous and next good frames. For the frame recovery method from [7] is used. The obtained results are tabulated in Table V for EMBSD and Table VI for spectral distortion.

Table IV. Bit allocation for DSQ at 27 bits

Index	1	2	3	4	5	6	7	8	9	10	Total
Bits	2	3	3	3	3	3	3	3	2	2	27

Table V. Performance of PSVQ method for different loss rates at 24 bits

Loss Rate (%)	Av. SD (dB)	2<SD<4 (%)	Sd>4 (%)	EMBSD
0	1.84	32.99	4.67	1.551
10	2.07	36.95	7.71	2.567
20	2.31	41.28	10.98	3.688
30	2.54	45.76	13.93	4.610
40	2.77	48.97	17.44	4.944

Table VI. Performance of DSQ method for different loss rates at 27 bits

Loss Rate (%)	Av. SD (dB)	2<SD<4 (%)	Sd>4 (%)	EMBSD
0	1.81	32.99	4.44	1.498
10	1.93	34.85	6.23	1.933
20	2.05	36.47	8.14	2.676
30	2.17	38.11	9.85	3.227
40	2.30	40.08	11.73	3.417

Figures 6 and 7 show the performance of DSQ and PSVQ respectively for different loss rates.

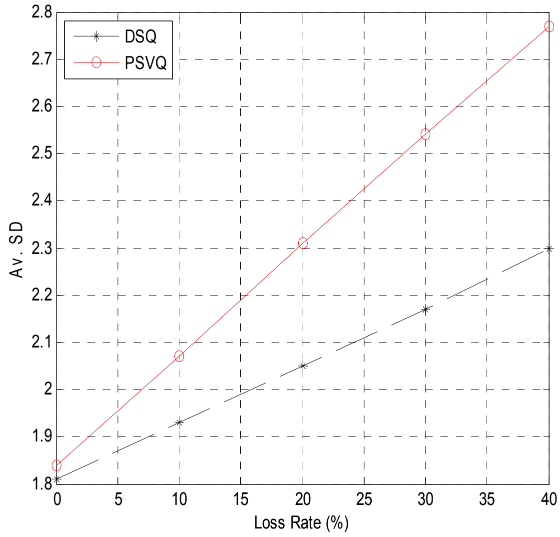

Figure 6. Average LSF spectral distortion for DSQ and PSVQ.

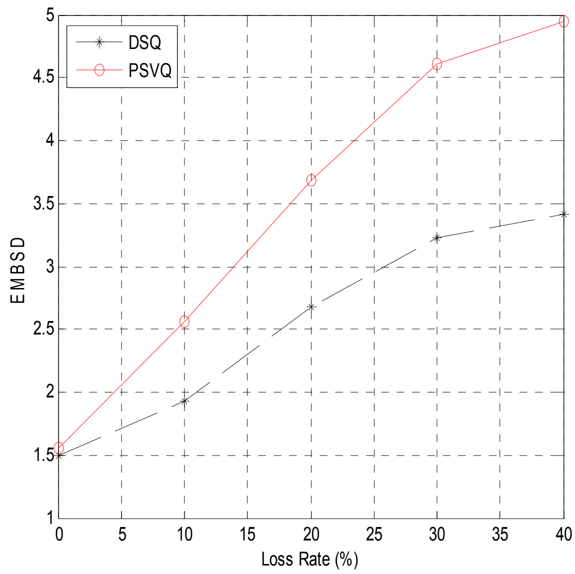

Figure 7. EMBSD for different lost rates for DSQ and PSVQ.

These results show that the DSQ obtains 0.14-0.47 dB improvement on average spectral distortion and the number of outliers is substantially reduced under frame erasures compared to PSVQ. The spectral distortion distributions show that more lost frames are interpolated with small distortion with DSQ.

The EMBSD tests indicate significant quality improvement with erasure-interpolated speech from DSQ coding.

The total rate for the G.723.1 speech coder is 6.3 or 5.27 kbit/s depending on the excitation method; therefore 3 bits/frame will add a 0.1 kbit/s extra rate on top of this, which is 5% of the total rate. Since packet loss is unavoidable, trading a small percent of rate for much needed reliability may be an attractive solution.

V. Conclusion

In this paper we have presented an efficient method for reducing error propagation for CELP based coders the standard G723.1.

Our results show by adding 3bits due to using DSQ rather than PSVQ, we can substantially improve the interpolation accuracy of LSF parameters under frame erasures.

References

[1] ITU, ITU-T G.723.1: Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, ITU 1996.
[2] ITU, ITU-T G.729: CS-ACELP Speech Coding at 8 kbit/s, ITU 1998.
[3] F.Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", J.Acoust. Soc. Amer., vol. 57, suppl. 1, p. S35(A),1975.
[4] K. K. Paliwal and B.Atal, "Efficient Vector quantization of LPC Parameters at 24 bits/frame," ICASSP, pp. 661-664, Mar. 1991.
[5] NIST,Timit Speech Corpus, NIST 1990.
[6] 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
[7] 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.