DYNAMIC SCALAR QUANTIZATION OF LSP (Line Spectral Pair)

Gao Yang[*,**] and H. Leich[*]

[*] Lab. T.C.T.S., Faculté Polytechnique de Mons,Belgium

[**] Lernout & Hauspie Speechproducts n.v.,Belgium

ABSTRACT - On the scalar quantization (SQ) of LSP, the interframe and intraframe correlations can be utilized to reduce the bit rate; however, the interframe correlation is generally weak for a longer frame shift which seems to be necessary for speech coding at a low bit rate below 4k bps; besides the influence of channel errors is not local if the interframe correlation is used. In this paper, a dynamic SQ method of LSP without using the interframe correlation is proposed; a simple and efficient distortion measure is presented. The experiments show that a satisfied perceptual quality can be achieved by using 25 bits per frame with the dynamic SQ.

INTRODUCTION

The spectral envelope of speech signal can be represented by various synthesis-filter parameters such as cepstral coefficients, reflection coeffients or Line Spectrum Pairs. Until now the Line Spectrum Pairs seem to be the most efficient parameters whose quantization and interpolation are robust and physical meaning is obvious. LSP was first introduced by Itakura (1975) as an alternative LPC spectral representation. It was found that this representation has such interesting properties as (1) all zeros of LSP polynominals are on the unit circle, (2) the corresponding zeros of the symmetric and anti-symmetric LSP polynominals are interlaced, (3) the reconstructed LPC all-pole filter preserves its minimum phase property if (1) and (2) are kept intact through a quantization procedure.

For speech coding at a bit rate as low as 3 kbps, it seems necessary to reduce the bit rate for coding LSP's to 25 bits/frame or below. VQ (Vector Quantization) of LSP's was found to be a good way to arrive at this purpose, but the required computation load and storage are usually much higher than SQ. That is why the SQ methods are also attractive for some practical applications. On SQ of LSP, the interframe and intraframe correlations can be utilized to reduce the bit rate. The two-dimensional differential coding (Chih-Chung Kuo & al.,1992) is an example for using the interframe and intraframe correlations in the same time. However, the interframe correlation is usually weak for a longer frame shift which seems to be necessary for speech coding at a low bit rate as low as 4k bps; besides the influence of channel errors is not local if the interframe correlation is used. An efficient method of coding differential LSP (d-LSP) paramters (Frank,K.S. & Juang,B.H.,1984) by using a DPCM framework is a typical example to use only the intraframe correlation of LSP's. The high quality and low computation load of d-LSP coding at a bit rate from 30 to 32 bits/frame was later confirmed by Sugamura and Farvardin (1988). Neverthless, the coding of d-LSP parameters, due to its differential coding nature, occasionally leads to large spectral distortions as described by Frank K.S.& Juang,B.H.(1990). These spectral "glitches," if not corrected, can be subjectively disturbing and significantly degrade the perceived quality of a coder. On the other hand, it is impossible to reduce the bit rate to 25 bits/frame with the d-LSP coding while maintaining the high quality. In this paper, we propose a dynamic SQ method using only the intraframe correlation of LSP's. For evaluating the performance of the quantization methods, a simple and efficient distortion measure is presented.

A SIMPLE AND EFFICIENT DISTORTION MEASURE

In order to limit the computation load, It is obligatory for VQ to define a simple and efficient objective measure which is also usefull for estimating the performance of SQ method. Some researches have been done by Frank,K.S. & Juang,B.H. (1990) and Hagan,R. & Hedelin,P. (1990). Indeed, the objective distortion measure should combine two features: it must be able to predict subjective tests accurately and, secondly, it should be computationally efficient. Good predictors of the performance of the human ear tend to be complex as regards computation.

One of the best available suggestions for a measure (of moderate complexity) of spectral distortion is SD, defined by

$$SD^2 = \frac{1}{\pi}\int_0^{\pi}\left[10\cdot\log\frac{S(\omega)}{\hat{S}(\omega)}\right]^2 d\omega \qquad (1)$$

This is a direct evaluation of the RMS dB-distance along the frequency axis between the original spectrum and its quantized version for one frame of speech signal. Average SD (over some available database) has become accepted for the evaluation of spectral coders. Experimental results from several studies indicate that whenever the SD-value drops below 1 dB, the quantizer has introduced distortion which can be considered as negligible.

As described by Hagan,R. & Hedelin,P. (1990), it is easy to criticize the SD-measure. For instance, it does not account for masking effects, neither in the frequency-domain, nor in the time-domain. Frank,K.S. & Juang,B.H. (1990) defined a spectral distortion measure using a weighted Euclidean distance (WED):

$$WED^2 = \sum_{i=0}^{p-1} w_i^2 (f_i - \hat{f}_i)^2 \qquad (2)$$

where p is the order of the LPC all-pole model and the weighting coefficient w_i of the spectral sensitivity with respect to the i-th LSP frequency, f_i, is defined as

$$w_i = \sqrt{\int_0^1 \left|\frac{\partial \log S}{\partial f_i}\right|^2 df} \qquad (3)$$

In this formulation $\log S$ is the log power spectrum of the p-th order LPC all-pole model and f is the normalized frequency. In this paper we will use also the weighted Euclidean distance of (2) but define a simple and efficient weighting coefficient w_i as

$$w_i = \left[\frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i}\right]^2 \cdot ham(i) \qquad (4)$$

where ham(i) = $\cos[(3\pi/8)\cdot i/(p-1)]$ is a part of Hamming window. The formulation (4) reflects three perceptual characters of the short-term spectral envelope: (a) when the LSP

frequencies are closer each other (a formant exists), the spectral sensitivity with respect to the LSP frequency is higher; (b) the coding errors in the lower frequency region is more perceptual than the higher frequency region; (c) the auditory quality of voiced speech is more important than unvoiced speech. We would not say that the objective measure by the combination of (2) and (4) can replace the subjective tests, but indeed it can well represent the perceptual performance obtained by the subjective tests.

DYNAMIC SCALAR QUANTIZATION OF LSP

Suppose LSP frequencies of 10-th order f_i (Hz), i=0,..,9, and f_{10} = 4000 Hz are defined. The dynamic SQ of LSP without using the interframe correlation is based on the following properties: (a) $f_i > f_{i-1}$, (b) the quantization precision should be higher when two LSP frequencies are closer each other, (c) the perceptual quality of quantization of LSP for voiced speech is more important than unvoiced speech, (d) $f_{2i} - f_{2i-2} > 200$ Hz for i=2,3,4.

First let us suppose f_{2i} (i=0,..,4) have been quantized to $\hat{f}_{2i}$. f_{2i+1} (i=0,..,4) are then quantized to $\hat{f}_{2i+1}$ by minimizing the distance measure

$$d(j) = f_{2i+1} - [\hat{f}_{2i} + \alpha_i(j)\cdot(\hat{f}_{2i+2} - \hat{f}_{2i})], \quad j = 1,..,M_i \qquad (5)$$

where the constants $\{\alpha_i(j)\}$ satisfy

$$0 < \alpha_i(1) < \ldots\ldots \alpha_i(M_i) < 1 \qquad (6)$$

M_i is the number of quantization levels which depends on the bit number (e.g. M_i=8 for 3 bits); usually $\alpha_i(M_i/2)$ or $\alpha_i(M_i/2+1)$ are close or equal to 0.5; we make always

$$0 < \alpha_i(1) < \alpha_i(2) - \alpha_i(1) < \ldots\ldots < \alpha_i(M_i/2) - \alpha_i(M_i/2-1) \qquad (7)$$

the similar principle is used for $\alpha_i(j)$, $j > M_i/2$:

$$1 - \alpha_i(M_i) < \alpha_i(M_i) - \alpha_i(M_i-1) < \ldots\ldots < \alpha_i(M_i/2+2) - \alpha_i(M_i/2+1) \qquad (8)$$

(6) means $\hat{f}_{2i} < \hat{f}_{2i+1} < \hat{f}_{2i+2}$. (7) and (8) signify that the more precise quantization is given when f_{2i+1} is close to $\hat{f}_{2i}$ or $\hat{f}_{2i+2}$, which takes place usually for voiced speech. If d(j) has been minimized with $j=J_{opt}$, we have the quantized value of f_{2i+1}:

$$\hat{f}_{2i+1} = \hat{f}_{2i} + \alpha_i(J_{opt})\cdot(\hat{f}_{2i+2} - \hat{f}_{2i}) \qquad (9)$$

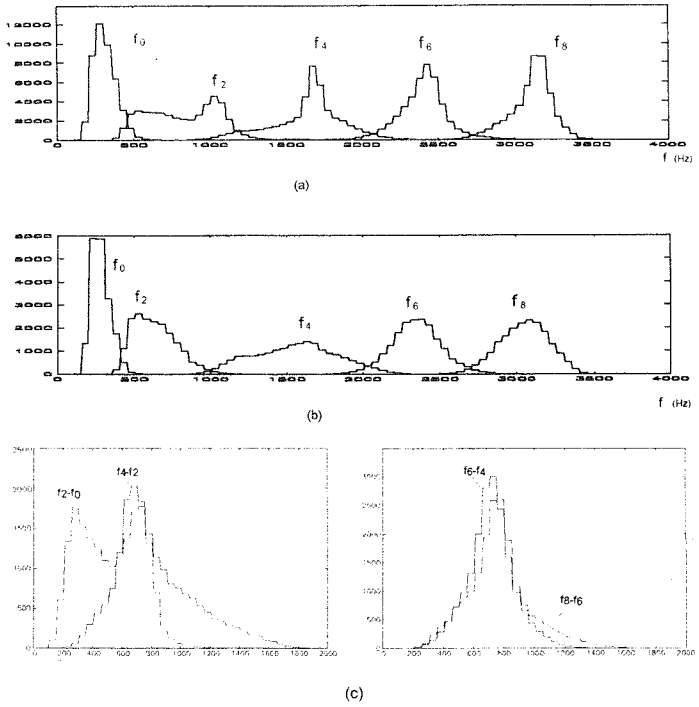


Fig.1 (a) Histograms of LSP frequencies: f_{2i} (i=0,..,4).
(b) Histograms of LSP frequencies f_{2i} for voiced speech.
(c) Histograms of LSP frequency differences: $f_{2i+2} - f_{2i}$ (i=0,..,3)

Fig.1(a) and (b) show the histograms of LSP frequencies f_{2i} (i=0,..,4) respectively for usual speech and for voiced speech. Fig.1(c) verifies the histograms of LSP frequency differences $f_{2i+2} - f_{2i}$ (i=0,..,3) for usual speech. We can see that the histogram of LSP frequencies for usual speech is different from that for voiced speech, comparing Fig.1(a) and (b). The quantization levels will be determined mainly according to the histograms for voiced speech, as the perceptual quality of voiced speech is more important than that of unvoiced speech.

Now the quantization method of f_{2i} (i=0,..,4) will be given. f_0 is directly quantized with usual SQ. $f_2 - \hat{f}_0$ is then quantized instead of directly coding f_2 because the range of $f_2 - f_0$ distribution is smaller than that of f_2. f_{2i} (i=2,3,4) can be coded as follows. Suppose the possible maximun and minimum of f_4 are respectively f_4^{MAX} and f_4^{MIN} according to the distribution of f_4. For a known $\hat{f}_2$, the practical possible minimum of f_4 is

$$f_4^{min} = \hat{f}_2 + 200 \qquad (10)$$

where $f_{2i} - f_{2i-2} > 200$ Hz for i=2,3,4. is used. f_4^{min} is often greater than f_4^{MIN}, so $f_4^{MAX} - f_4^{min}$ is often smaller than $f_4^{MAX} - f_4^{MIN}$. If $f_4^{min} < f_4^{MIN}$, we set $f_4^{min} = f_4^{MIN}$. In this way we have always

$$f_4^{MAX} - f_4^{min} < f_4^{MAX} - f_4^{MIN} \qquad (11)$$

f_4 is coded by minimizing the error distance

$$e(i) = f_4 - [f_4^{min} + \beta(i)(f_4^{MAX} - f_4^{min})], \quad i=1,..,N_4 \qquad (12)$$

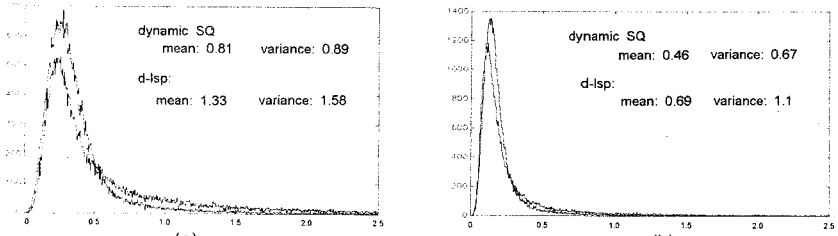
where the constants $\{\beta(i)\}$ satisfy $0<\beta(1)<\cdots<\beta(N_4)<1$. We can define several sets of $\{\beta(i)\}$ for different values of f_4^{min} mainly according to the statistical characteristics of f_4 for voiced speech. If e(i) has been minimized with i= I_{opt}, we have

$$\hat{f}_4 = f_4^{min} + \beta(I_{opt})(f_4^{MAX} - f_4^{min}) \qquad (13)$$

f_6 and f_8 are similarly coded as f_4.

PERFORMANCE OF THE DYNAMIC SCALAR QUANTIZATION OF LSP

The same data base used for calculating the histograms in Fig.1 is used here for testing the performance of the dynamic scalar quantization of LSP. The speech signals were digitized at 8 kHz sampling rate. A 10-th order LPC analysis, based on the autocorrelation method, was performed by using a 30 ms Hamming window shifted every 20 ms. There are about 50000 frames of LSP vectors used for obtaining all the histograms in this paper. The performance of the dynamic SQ is compared with that of the typical SQ named d-LSP (Frank,K.S. & Juang,B.H.,1984) without using interframe correlations.



(a) (b)

Fig.2 Coding performance comparison between the dynamic SQ and d-LSP SQ.
(a) distortion histograms of the two coding schemes with 25 bits:(3,2,4,2,4,2,3,1,3,1) for the dynamic SQ and (3,3,3,3,3,2,2,2,2,2) for the d-LSP SQ.
(b) distortion histograms of the two coding schemes with 30 bits:(3,3,4,3,4,3,3,2,3,2) for the dynamic SQ and (3,3,3,3,3,3,3,3,3,3) for the d-LSP SQ.

Fig.2 gives the distortion histograms with the measure of (2) and (4) for the dynamic SQ and the d-LSP SQ. The auditory results by the subjective tests applying the quantized LSP's to a 2.4 kbps vocoder are basically consistent with those obtained by the objective

measure. We can see that the performance of the dynamic SQ is evidently better than the d-LSP SQ. A satisfied perceptual quality with the dynamic SQ of 25 bits being used to the 2.4 kbps vocoder is achieved, while it is impossible to reduce the bit rate to 25 bits/frame with the d-LSP SQ.

CONCLUSION

This paper proposes a dynamic SQ method of LSP without using the interframe correlation; a simple and efficient distortion measure is presented. The experiments show that the performance of the dynamic SQ is evidently better than the d-LSP SQ and a satisfied perceptual quality with the dynamic SQ of 25 bits is achieved.

ACKNOLEDGMENT

The authors wish to thank Mr. G. Zanellato and the other colleagues in the laboratory for their supports and helps.

REFERENCES

Chih-Chung Kuo & al.(1992) *Low bit-rate quantization of LSP parameters using two-dimentional differential coding*, IEEE ICASSP pp. I-97.

Frank,K.S. & Juang,B.H.(1984) *Line spectrum pair (LSP) and speech data compression*, IEEE ICASSP, pp 1.10.1-1.10.4.

Frank K.S.& Juang,B.H.(1990) *Optimal quantization of LSP parameters using delayed decisions*, IEEE ICASSP, S4a.5, pp.185-188.

Hagan,R. & Hedelin,P.(1990) *Low bit-rate spectral coding in CELP,a new LSP-method*, IEEE ICASSP, S4a.6, pp.189-192.

Itakura,F.(1975) *Line spectrum representation of linear predictive coefficients of speech signals*, J. Acoust. Soc. Am., 57, 535(A).

Sugamura,N. & Itakura,F.(1981) *Speech data compression by LSP speech analysis-synthesis technique*, Trans. IECE'81/8 Vol. J 64-A, No.8, pp. 599-606.

Sugamura,N. & Farvardin,N.(1988) *Quantizer design in LSP analysis-synthesis*, IEEE Selected Areas in Communications, Vol.6, No.2, pp.432-440, Feb. 1988.