Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 49(63), Fascicola 1, 2004

Efficient vector quantization of speech spectral parameters

Cornel Balint[1]

Abstract – Most low bit rate speech coders employ linear predictive coding (LPC) which models the short-term spectral information within each speech frame as an all-pole filter. In this paper, we examine various methods in order to efficiently encode spectral parameters.

Two efficient methods for LSF coding are proposed, that exploit the interframe correlation exiting between LSF parameters. Using vector predictive coding techniques, transparent coding quality was obtained at 24 – 26 bit/frame.

Keywords: speech coding, LSF, vector coding, predictive co ing

I. INTRODUCTION

Various methods for speech analysis-synthesis techniques based on linear prediction are known [1]. Among these, Code-Excited Linear Prediction (CELP) is the most widely studied and the most efficient algorithm for high quality speech at low and very low bit rate (4 – 8 kbps) [2].

In the source-filter model of speech, the speech signal is modeled as the response of a time-varying linear filter to a voice source excitation signal. The production filter is estimated by linear predictive analysis (LPC), where the parameters of the filter are chosen to minimize a quadratic error criterion.

Figure 1 illustrates the source-filter model for speech production.

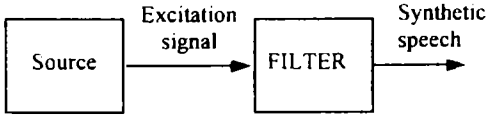

Fig. 1 Source filter model

In speech coding applications that use CELP, two major categories of information need to be quantized and transmitted to the receiver, and then used to produce reconstructed speech:

- excitation parameters,
- filter parameters.

Quantization of excitation and filter parameters is the key in CELP speech coders and this paper investigate some vector quantization techniques for coding parameters that describe the filter.

The schematic diagram of the simulation model for LPC coefficients quantization is presented in figure 2.

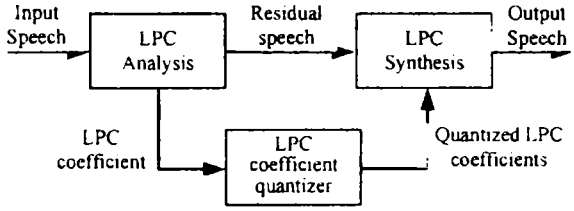

Fig. 2. Model for quantization of LPC coefficients

For every 20 ms frame interval the LPC analysis is performed in order to compute the prediction filter with transfer function:

$$\frac{1}{A_p(z)} = \frac{1}{1+\sum_{i=1}^{p} a_i z^{-i}} \qquad (1)$$

Only the quantization of LPC coefficients a_i are investigate in this work and the speech residual signal is transmitted to the LPC synthesis filter without any modification. Efficient representation of the LPC coefficients is an important part of the speech coding architecture.

Line spectral frequencies (LSF) are considered to be the most advantageous parametric representation of the LPC for coding [3].

Vector quantization (VQ) techniques ensure higher performance at lower bit rate [4]. Typically, about 30 bits are assigned to code 10-dimensional LSF vector, with spectral distortion less than 1 dB. However, 30 bit full search VQ is impractical in terms of computational complexity.

Various sub-optimal VQ techniques have been used for LSF quantization. The most commonly used are product code techniques, split VQ (SVQ) and multi-stage VQ (MSVQ), employed in intraframe spectral parameters coding, where each frame vector is encoded independently from other frames [5].

[1] Department of Communications, "Politehnica" University, Timisoara, Blvd V. Parvan 2, 300223 Timisoara, Romania, Phone: +40256-403310, E-mail: cbalint@etc.utt.ro

Depending on the code structure, "transparent coding" quality is achieved for SVQ at 26-30 bits/frame and for MSVQ at 25-28 bits/frame [3], [8].

Because speech signal is quasi-stationary, the LSF parameters denote an interframe correlation.

In this paper, two different efficient methods for LSF coding are investigated, in order to exploit the interframe correlation: predictive split VQ (PSVQ) and nonlinear PSVQ. In the nonlinear PSVQ, a nonparametric and nonlinear predictor replaces the linear predictor used in PSVQ.

Comparing with a simple split VQ, the proposed PSVQ method offer a performance gain of 4-6 bits/frame, regardless of predictor type.

Using intraframe SVQ for odd frame and interframe PSVQ for even frame, quantization error propagation is limited to at most one frame.

The proposed particular form of nonlinear prediction involves a no significant additional encoding computational complexity.

II. THE SPECTRAL FREQUENCIES

The LPC coefficients a_k determined according (1) represent the short-term speech signal spectrum and are completely equivalent in mathematical sense to other linear predictive coefficients, as LSF or PARCOR coefficients [1], [2].

The polynomial $A_p(z)$ corresponding to p-order LPC analysis, satisfies the recurrence relationship:

$$A_p(z) = A_{p-1}(z) - k_p z^{-p} A_{p-1}(z^{-1}), \quad p = 1,2,...P \qquad (2)$$

where $A_0(z) = 1$. Parameters k_p are called PARCOR (PARtiall CORellation) coefficients.

For $p = P + 1$ (2) becomes:

$$A_{P+1}(z) = A_{P-1}(z) - k_{P+1} z^{-(P+1)} A_P(z^{-1}) \qquad (3)$$

Considering in (3) two extreme artificial boundary condition: $k_{P-1} = 1$ and $k_{P+1} = -1$, that correspond to a complete closure respective to a complete opening of the glottis in the acoustic tube model of the vocal tract [2], the polynomial (3) can be expressed:

$$P(z) = (1 - z^{-1}) \prod_{i=2,4,..p} (1 - 2z^{-1}\cos\omega_i + z^{-2}) \qquad (4)$$

for $k_{P+1} = 1$, and:

$$Q(z) = (1 + z^{-1}) \prod_{i=3,5,..p-1} (1 - 2z^{-1}\cos\omega_i + z^{-2}) \qquad (5)$$

for $k_{P-1} = -1$, where is assumed that:

$\omega_1 < \omega_3 < ... < \omega_{P-1}$, $\omega_2 < \omega_4 < ... < \omega_P$ and P is an even integer.

The roots of the polynomials $P(z)$ and $Q(z)$ are $e^{j\omega_i}$, $i = 1,2...,P$ and the parameters ω_i are defined as Line Spectrum Pair (LSP) or Line Spectrum Frequencies (LSF). LSF parameters can be interpreted as natural resonant frequencies of the vocal tract in two extreme artificial boundaries conditions at the glottis. In figure 3 are illustrate the LPC short-term power spectrum and the LSF frequencies of the model presented in fig. 2, for a fragment of voiced sound.

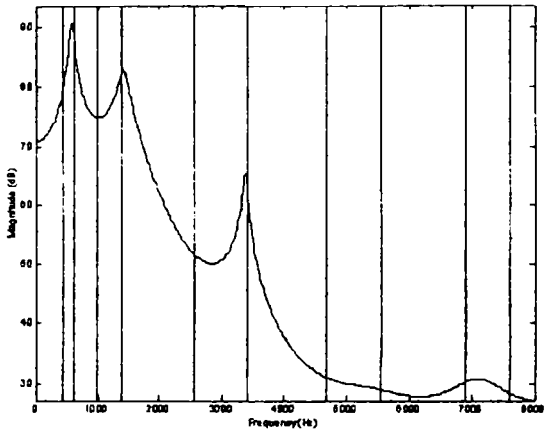

Fig. 3 Power spectrum and LSF for voiced sound

The most important properties of LSF coefficients are the ordering and boundary property, that is very useful in coding:

$$0 = \omega_0 < \omega_1 < \omega_2 < ...\omega_P < \omega_{P+1} = \pi \qquad (6)$$

The ordering property denotes that the LSF parameters within a frame are correlated and use of this correlation can improve their quantization.

Due to the fact that the vocal tract varies slowly in time, we expect to find a strong correlation between the LSF parameters in consecutive frame. In figure 4, some consecutive spectra of the same situation depicted in fig. 3, are represented together, in order to illustrate the interframe correlation between LSF parameters.

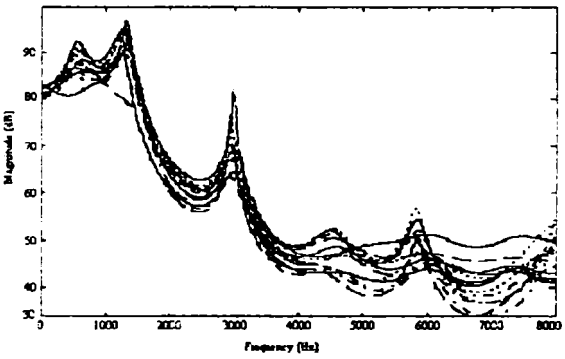

Fig. 4. Consecutive LPC power spectra

To investigate the interframe correlation of the LSF in figure 5 are represented the time evolution of the first 3 LSF parameters.

The correlation coefficients between $\omega_{n,i}$ and $\omega_{n-k,i}$, where n denotes the frame and i denote the

order of LSF, was computed for k = 10 consecutive frames. In table 1 was indicate the correlations coefficients for first 7 frames that confirm the strong correlations between consecutive frames.

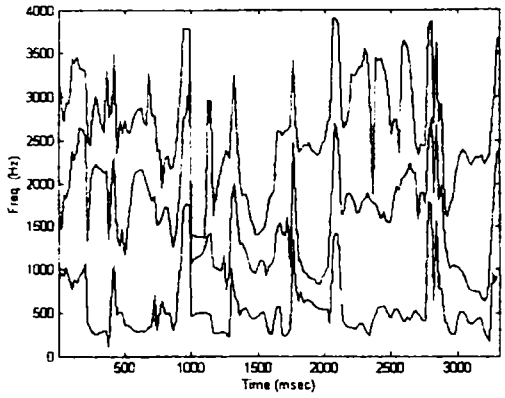

Fig. 5 Traces of LSF parameters

Table 1. Interframe correlation coefficients

l\k	1	2	3	4	5	6	7
1	0.94	0.85	0.76	0.69	0 61	0.56	0.50
2	0.90	0.77	0.63	0.54	0 47	0.38	0.32
3	0.93	0.79	0.71	0.62	0.50	0.44	0.35
4	0.91	0.83	0.73	0.64	0.55	0.48	0.43
5	0.96	0.87	0.81	0.75	0 66	0.60	0.53
6	0.95	0.85	0.76	0.68	0.61	0.55	0.48
7	0.92	0.82	0.76	0.65	0.59	0.50	0.42
8	0.91	0.82	0.72	0.64	0.56	0.48	0.42
9	0.89	0.74	0.65	0.56	0.47	0.42	0.36
10	0.80	0.70	0.59	0.50	0.44	0.38	0.33

III. LINEAR PREDICTIVE VECTOR LSF QUANTIZATION

The predictive vector quantizer adapted to exploit inteframe correlation of LSF is presented in figure 6.

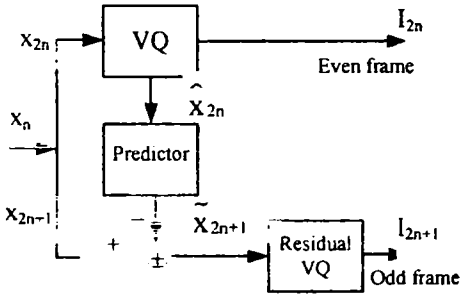

Fig. 6. Predictive VQ

Let $\{x_n\}$ denote the input sequence of 10-dimensional LSF vector to be encoded. Grouping together two consecutive vectors, obtain contiguous pair (x_{2n}, x_{2n+1}). The even vector x_{2n} is independent encoded in VQ quantizer, resulting the quantized vector $\hat{x}_{2n}$ and the index I_{2n} that are transmitted to the decoder.

From $\hat{x}_{2n}$ the predictor compute the prediction value $\tilde{x}_{2n+1}$ of the LSF odd vector x_{2n+1}.

The prediction error:

$$e_{2n+1} = x_{2n+1} - \tilde{x}_{2n+1} \tag{7}$$

's quant'ze 'us'ng res 'ua ' ' ' 'ock 'n fig. 6, o' ta'n the index I_{2n+1} for odd frame. The quantized LSF vector of odd frame (predicted frame) can be reconstructed according:

$$\hat{x}_{2n+1} = \tilde{x}_{2n+1} + e_{2n+1} \tag{8}$$

This encoding process is repeated by applying alternately VQ in even frame and predictive VQ in odd frame.

For a LSF vector sequence $\{x_n\}$ the M-th order vec or linear predic ion compu e a pre 'ic ion $\tilde{x}_n$ o' the actual vector x_n based on M preceding vectors:

$$\tilde{x}_n = \sum_{i=1}^{M} P_i \hat{x}_{n-i} \tag{9}$$

where P_i, $i = 1, 2, \ldots, M$ are 10x10 prediction matrices.

If all M prediction matrices are diagonal matrices, vector linear prediction reduces to particular case of scalar linear prediction in which each LSF vector component in the predicted frame is predicted using only the corresponding vector component from preceding frame.

When the prediction matrices are not diagonal, we have vector linear prediction that exploits the intercomponent correlation between adjacent frames, if that correlation exists.

Exploiting the interframe redundancy of LSF parameters, fewer bits are required for coding the prediction residual vector in predicted frames that required to coding the LSF vector. An additional delay of one frame will be introduced.

IV. NONLINEAR PREDICTIVE VECTOR LSF QUANTIZATION

Nonlinear predictive LSF quantization is based on the diagram of fig. 6, when the block predictor was replace with a nonlinear predictor. The nonlinear predictor is based on nonlinear interpolative vector quantization proposed in [7] and [10] for split and multistage vector quantizers [6]. The minimum mean-square error prediction $\hat{\mathbf{Y}}$ of a random vector $\mathbf{Y}$ given another random vector $\mathbf{X}$ is the conditional expectation of $\mathbf{Y}$ given $\mathbf{X}$:

$$\hat{\mathbf{Y}}(\mathbf{X}) = E[\mathbf{Y} \mid \mathbf{X}] \tag{10}$$

If the joint probability distribution of $\mathbf{X}$ and $\mathbf{Y}$ is unknown, we can assume that the conditional expectation is a nonlinear function. If the observation of $\mathbf{X}$ is quantized to a finite set of possible values

$\{\tilde{\mathbf{X}}^{(i)}\}$, there is also a finite number of possible conditional expectation values $\{\hat{\mathbf{y}}^{(i)}\}$:

$$\tilde{\mathbf{y}}^{(i)} = E[Y \mid \hat{\mathbf{x}}^{(i)}] \qquad (11)$$

Without knowing the functional form of the mean-square error estimator, we can found a table of conditional expectation values in the designing and training process of quantizer [10].

The nonlinear predictor is constructed as a codebook of conditional expectation, one for each value of the quantized value $\hat{\mathbf{x}}_{2n}$.

V. EXPERIMENTAL RESULTS

Experimental results are based on a training set and a test set of LFS vectors. The training set consists of 45.000 LSF vectors, obtained from approximately 15 minutes of speech. For tests were used similar sequences of 4.000-6-000 LSF vectors, different from training set. The speech signal was recorded from FM radio (Internet), low-pass-filtered at 3.4 kHz and sampled at 8 kHz. For the experimental results illustrated in fig. 3, 4 and 5 the speech signal was sampled at 12.5 kHz. Using a sound card and a wave editing software the silence period from speech signal was removed.

A 10 order LPC analysis using the stabilized covariance method with high-frequency compensation and error weighting was performed every 20 ms using a 25 ms analysis window. A fixed 10 Hz bandwidth expansion was applied to each pole of the LPC vector.

For subjective listening tests on reconstructed speech signal, a synthesis filter with quantized coefficients was used. The excitation signal for the synthesis filter is the unquantized linear prediction residual signal.

For measuring the quantization performance we calculate the log spectral distance (SD) in the 0-3 kHz band according to:

$$SD^2 = \frac{1}{N}\sum_{n=0}^{N-1}\left(20\log\frac{|S(n)|}{|\hat{S}(n)|}\right)^2 \qquad (12)$$

All vector quantizers are designed using the GLA algorithm [4]. For the codebook search we employ the weighted Euclidean measure that has shown to improve both the objective quality and the subjective quality of the coded speech.

In table 2 are presented the experimental results for spectral distortion SD for linear prediction, as function of prediction order (M = 1, 2, 3), for different codebook sizes. According to table 2, greater prediction order has no practical effect. The number of vectors with SD in the range 2-4 dB, which can be considered as a measure of the number of outliers is reported in table 2 as percent (column %). This number is a very important measure since very disturbing distortion can result from a vector with high SD. In these experiments, no quantizers have vectors with SD greater than 4 dB. For linear prediction VQ, a split VQ technique was tested in order to reduce the coding complexity. The used split scheme was 3-3-4 according more importance to first LSF [4] and [9]

Table 2. Spectral distortions for linear prediction

Bits/ frame	M = 1		M = 2		M = 3	
	SD	%	SD	%	SD	%
20	1.11	4.0	1.11	3.7	1.11	3.9
21	1.03	2.6	1.02	2.51	1.03	2.6
22	0.96	1.8	0.95	1.7	0.95	1.9
23	0.90	1.4	0.89	1.3	0.89	1.3
24	0.83	0.8	0.83	0.9	0.83	0.9
25	0.77	0.6	0.77	0.6	0.77	0.6
26	0.73	0.5	0.73	0.4	0.73	0.4
27	0.68	0.3	0.67	0.3	0.67	0.3

For the nonlinear predictor case, the experimental results are presented in table 3.

Table 3. Spectral distortions for nonlinear prediction

Bits/frame	20	24	27
SD	1.08	0.80	0.61
%	2.2	0.5	0.3

As future work, the presented method will be combined with a classification of speech frame in voicing or unvoicing. Some subjective test suggest that for the unvoice frame transparent quality can be achieved at 10-12 bit/frame. Using classification, intraframe vector coding becomes a classified vector quantizer [4], using different sets of codebook for voiced and unvoiced frames.

REFERENCES

[1] L.R. Rabiner, R. W. Shafer, *Digital Processing of Speech Signal*, Prentice-Hall, Englewood Clifs, NJ 1978

[2] J. R. Deller, Jr., J. G. Proakis, J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, New York: MacMillan, 1993.

[3] K. K. Paliwal, B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Trans. Speech and Audio Proc.*, vol. 1, January 1993.

[4] A. Gersho, R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, Boston, 1992.

[5] J. Grass, P. Kabal, "Methods of improving vector-scalar quantization of LPC coefficients", *Proc. Int. Conf. Acoustics, Speech and Signal Proc.*, (Toronto), May 1991.

[6] E. Paksoy, W.-Y. Chan, A. Gersho, "Vector quantization of speech LSF parameters with generalized product codes", *Proc. Int. Conf. Spoken Language*, October 1992.

[7] J. H. Y. Loo, W.-Y. Chan, "Nonlinear predictive vector quantization of speech spectral parameters", *Proc. IEEE Workshop on Speech Coding for Telecom.*, September 1995.

[8] F. Soong, B.-H. Juang, "Optimal quantization of LSP parameters", *IEEE Trans. Speech and Audio Proc.*, vol. 1, January 1993.

[9] R. Laroia, N. Phamdo, N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers", *Proc. Int. Conf. Acoustics, Speech and Signal Proc.*, Toronto, May 1991.

[10] A. Gersho, "Optimal nonlinear interpolative vector quantization", *IEEE Trans. on Comm.*, vol. COM-38, no. 9-10, sept. 1990.