INTERFRAME DIFFERENTIAL VECTOR CODING OF LINE SPECTRUM FREQUENCIES

Engin Erzin and A. Enis Çetin

Dept. of Electrical and Electronics Engineering,
Bilkent University
Bilkent, Ankara, 06533, Turkey

ABSTRACT

Line Spectrum Frequencies (LSF's) uniquely represent the Linear Predictive Coding (LPC) filter of a speech frame. In many vocoders LSF's are used to encode the LPC parameters. In this paper, an interframe differential coding scheme is presented for the LSF's. The LSF's of the current speech frame are predicted by using both the LSF's of the previous frame and some of the LSF's of the current frame. Then, the difference vector resulting from prediction is vector quantized.

I. Introduction

Linear Predictive modeling techniques are used in various speech coding, synthesis and recognition applications. In many vocoders the sampled speech signal is divided into frames and in each frame a 10-th order Linear Predictive Coding (LPC) filter is estimated. The LPC parameters can be represented by the Line Spectrum Frequencies (LSF's) which were first introduced by Itakura [1]. It is desirable to code the LSF parameters accurately by using as few bits as possible without degrading the speech quality.

The LSF representation provides a robust representation of the LPC synthesis filter with the following properties : (i) All of the zeros of the so-called LSF polynomials are on the unit circle, (ii) the zeros of the symmetric and anti-symmetric LSF polynomials are interlaced, and (iii) the reconstructed LPC all-pole filter maintains its minimum phase property, if the properties (i) and (ii) are preserved during the quantization procedure.

For a given m-th order LPC inverse filter $A_m(z)$, the LSF polynomials $P_{m+1}(z)$ and $Q_{m+1}(z)$ are defined as follows :

$$P_{m+1}(z) = A_m(z) + z^{-(m+1)}A_m(z^{-1}), \qquad (1)$$

and

$$Q_{m+1}(z) = A_m(z) - z^{-(m+1)}A_m(z^{-1}). \qquad (2)$$

It can be shown that the roots of $P_{m+1}(z)$ and $Q_{m+1}(z)$ uniquely characterize the LPC filter, $A_m(z)$. All of the roots are on the unit circle. Therefore, the roots of $P_{m+1}(z)$ and $Q_{m+1}(z)$ can be represented by their angles with respect to the positive real axis. These angles are called the Line Spectrum Frequencies (LSF's). In order to represent m-th order filter, $A_m(z)$, m suitably selected roots or equivalently LSF's are enough [2].

In a typical 8 kHz sampled speech waveform the LSF's of consecutive frames slightly vary [3]-[4]. By taking advantage of this fact we develop an interframe differential vector coding scheme for the LSF's in this paper.

In Section II we describe the new coding method and in Section III we present simulation examples.

II. Differential Coding of LSF's

In this section, we present the new LSF coding

method. The key idea of our scheme is to predict the LSF's of the current frame by using *both the LSF's of the previous frame and some of the LSF's of the current frame*. The prediction error vector between the true LSF's and the predicted LSF's is vector-quantized.

Let $A_{10}^n(z)$ be the 10-th order LPC filter of the nth speech frame. Corresponding to $A_{10}^n(z)$, 10 LSF's are defined. Let us denote the ith LSF of the nth frame by f_i^n, $i = 1, 2, ..., 10$. Our differential vector coding scheme estimates the current LSF, f_i^n, from i^{th} LSF of the $(n-1)^{th}$ frame, f_i^{n-1}, and $(i-1)^{th}$ LSF of the nth frame, f_{i-1}^n. In this way, we not only take advantage of the relation between neighboring LSF's but the relation between the LSF's of the consecutive frames as well. The estimate, $\hat{f}_i^n$, of the LSF, f_i^n, is predicted as follows,

$$\hat{f}_i^n = \begin{cases} a_i^n \Delta_i + b_i^n f_i^{n-1} & i = 1 \\ a_i^n(f_{i-1}^n + \Delta_i) + b_i^n f_i^{n-1} & i = 4, 8 \\ b_i^n f_i^{n-1} & \text{otherwise} \end{cases} \quad (3)$$

where a_i^n's and b_i^n's are the predictor coefficients and Δ_i is an offset factor which is the average angular difference between the ith and $(i-1)^{th}$ LSF's. The parameter, Δ_i, is experimentally determined. The set of offset factors that are used in our simulation examples are listed in Table 1. Predictor coefficients a_i^n's and b_i^n's are adapted by the Least Mean Square (LMS) algorithm as follows,

$$\begin{bmatrix} a_i^n \\ b_i^n \end{bmatrix} = \begin{bmatrix} a_i^{n-1} \\ b_i^{n-1} \end{bmatrix} + \alpha_i^{n-1} \begin{bmatrix} f_{i-1}^{n-1} + \Delta_i \\ f_i^{n-2} \end{bmatrix} d_i^{n-1} \quad (4)$$

where d_i^{n-1} is the quantized error value between the true LSF, f_i^{n-1}, and the predicted LSF, $\hat{f}_i^{n-1}$, and the adaptation parameter, α_i^{n-1} is given as:

$$\alpha_i^{n-1} = \frac{\lambda_i}{(f_{i-1}^{n-1} + \delta_i)^2 + (f_i^{n-2})^2}, \quad 0 < \lambda_i < 2. \quad (5)$$

The parameters, λ_i's, are also experimentally determined.

The error vector whose entries are, $f_i^n - \hat{f}_i^n$, $i = 1, 2, ..., 10$, is divided into three subvectors containing the first three LSF's, the middle four LSF's and the last three LSF's, respectively. We experimentally observed that choosing the LSF subvectors with the above partition produces better results than any other grouping. Due to the fact that there are three subvectors, only quantized f_{i-1}^n, $i = 4, 8$, are available in the predictor. Therefore, the predictor described in (3) uses only f_{i-1}^n, $i = 4, 8$. This intraframe information improves the performance of the predictor.

Each subvector is quantized using different vector quantizers. The codebook sizes that are used in simulation examples are shown in Table 2. For example, in the coding of the first (second) [third] subvector a codebook of size 128 (1024) [128] is used for 24 bits/frame case.

Recently, simulated annealing based quantizer design algorithms were developed [5]-[6], and it was observed that globally optimal solutions can be reached. In this paper we use the stochastic relaxation algorithm [6] in the design of the above three vector quantizers. We observed that stochastic relaxation algorithm produces better results than the generalized-Lloyd algorithm.

A weighted Euclidean distance measure [7] is used in quantizer design. The weights (ω_i) are proportional to the value of LPC power spectra at a given LSF, f_i^n:

$$\omega_i = [P(f_i^n)]^r \quad (6)$$

where $P(f)$ is the LPC power spectra of the n-th frame and r is an empirical constant which is chosen to be equal to 0.15 in our simulation examples.

III. Simulation Examples

In this section we present simulation examples and compare our results to other LSF coding schemes, including the vector quantizer based methods of Atal [7] and Farvardin [8].

The weighted M.M.S.E quantizers are trained in a set of 15000 speech frames containing six male and six female persons. The performance of the interframe LSF coding scheme is measured in a set of 11000 speech frames obtained from utterances of three male and three female persons. Lowpass filtered speech is digitized at a sampling rate of 8 kHz. A 10-th order LPC analysis is performed by using stabilized covariance method with high frequency compensation [4]. During the analysis a 30-ms Hamming window is used with a frame update period 20 ms. In order to avoid sharp spectral peaks in the LPC spectrum, a fixed bandwidth of 10 Hz is added uniformly to each LPC filter by using a fixed bandwidth-broadening factor, 0.996.

In our simulation examples we use the following spectral distortion measure

$$d'(A(\omega), A'(\omega)) = [\frac{1}{2\pi} \int_{-\pi}^{\pi} [10 log \frac{|A'(\omega)|^2}{|A(\omega)|^2} d\omega]^{1/2} \quad (7)$$

which is also used in [7] and [8]. The methods described in [7] and [8] reach 1.0 dB spectral distortion and an acceptable percent of outliers (less than 2% outliers with spectral distortion greater than 2 dB, [7]) at 24 and 25 bits/frame, respectively. Our method also reaches this LPC quantization level at 24 bits/frame. Our simulation results and the results of [7] and [8] are summarized in Table 3 and Table 4, respectively. Although we use different evaluation data sets than [7] and [8] (the sets used in [7] and [8] are also different from each other), we observe that our method produces comparable results to [7].

Interframe differential vector coding of LSF's is more advantageous than direct vector quantization of LSF's. Since the overall codebook size of our coder is much smaller than the ones used in [3] and [7] (e.g., 6.4 times lower than [7]), our method is computationally more efficient than [3] and [7], and it requires smaller storage space.

Recently, other interframe differential coding schemes are also described in [9] and [10]. In [9] the

scalar quantization is used and the prediction coefficients are fixed. In [10] the predictor does not utilize the angular offset factor, Δ_i, and a 1900 bits/sec (with a comparable distortion level) transmission rate is reported. In this paper better results than [9] and [10] are obtained. With our coding scheme a transmission rate of 1200 bits/sec with 1 dB average spectral distortion can be achieved. This is because of the fact that in this paper an adaptive predictor is used and the difference vector resulting from the prediction is vector quantized.

IV. Conclusion

In this paper, an interframe differential vector coding scheme is presented for the LSF's. The new scheme is computationally efficient and easy to implement. It can be used in low bit rate vocoders.

Recently, LSF's are used in speech recognition and synthesis. The new scheme can be also utilized in residual-excited PSOLA [11] type text-to-speech synthesizers.

References

1. F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals," *J. Acoust. Soc. Am.*, 57, 535(a), s35(A), 1975.

2. F. Soong and B.H. Juang, "Line Spectrum Pair and Data Compression," *Proc. of the Int. Conf. on Acoustic, Speech and Signal Processing 1984 (ICASSP '84)*, pp. 1.10.1-1.10.4, 1984.

3. M. Yong, G. Davidson, and A. Gersho, "Encoding of LPC Spectral Parameters Using Switched-Adaptive Interframe Vector Prediction," *Proc. of the Int. Conf. on Acoustic, Speech and Signal Processing 1988 (ICASSP '88)*, pp. 402-405, 1988.

4. B. S. Atal, "Predictive Coding of Speech at Low Bit Rates," *IEEE Trans. on Communications*, Vol. COM-30, No. 4, pp. 600-614, April 1982.

5. A. E. Çetin, and V. Weerackody, "Design of Vector Quantizers Using Simulated Annealing,"

IEEE Trans. Circ. Syst., CAS-35, pp. 1550, 1988.

6. K. Zeger, A. Gersho, "Stochastic Relaxation Algorithm for Improved Vector Quantiser Design", *Electronics Letters*, Vol. 25, No. 14,pp. 96-98, July 1989.

7. K.K. Paliwal and B.S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *Proc. of the Int. Conf. on Acoustic, Speech and Signal Processing 1991 (ICASSP '91)*, pp. 661-664, May 1991.

8. N. Phamdo, R. Laroia and N. Farvardin, "Robust and Efficient Quantization of LSP Parameters Using Structured Vector Quantizers," *Proc. of the Int. Conf. on Acoustic, Speech and Signal Processing 1991 (ICASSP '91)*, pp. 641-645, May 1991.

9. E. Erzin and A.E. Çetin, "Interframe Differential Coding of Line Spectrum Pairs," presented in *26-th Conference on Information Sciences and Systems, Princeton*, March 1992.

10. C.C. Kuo, F.R. Jean and H.C. Wang, "Low Bit-Rate Quantization of LSP Parameters Using Two-Dimentional Differential Coding," *Proc. of the Int. Conf. on Acoustic, Speech and Signal Processing 1992 (ICASSP '92)*, pp. 97-100, March 1992.

11. E. Moulines, F. Charpentier, "Pitch-Synchronous Waveform Processing Tecniques for Text-to-Speech Synthesis using Diphones," *Speech Comm.*, Vol. 9, No. 5/6, pp. 453-467, Dec. 1990.

Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
0.22	0.12	0.24	0.37	0.32

Δ_6	Δ_7	Δ_8	Δ_9	Δ_{10}
0.26	0.37	0.23	0.29	0.28

Table 1: *The angular offset factors which are used in simulations*

Rate	Codebook Sizes		
bits/frame	first 3	middle 4	last 3
22	128	256	128
23	128	512	128
24	128	1024	128

Table 2: *Codebook sizes for each subvector at different rates*

Rate bits/frame	Av. SD (dB)	outliers > 2dB
22	1.02	3.66%
23	0.92	2.26%
24	0.86	1.55%

Table 3: *Spectral Distortion (SD) Performance of our method*

Rate bits per frame	[7]		[8]	
	Av. SD (dB)	outliers > 2dB	Av. SD (dB)	outliers > 2dB
22	1.17	2.73%	-	-
23	1.10	1.60%	-	-
24	1.03	1.03%	1.11	1.50%
25	0.96	0.61%	1.02	0.20%

Table 4: *Spectral Distortion (SD) Performance of the Vector Quantizers [7] and [8]*