A new intraframe LSP interpolation technique for low bit rate speech coding

J. S. Mao, S. C. Chan and K. L. Ho
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong

Abstract

This paper presents a linear LSP interpolation between neighboring frames. Usually there are two to four formants in speech spectrum envelope, and the LSF(Line Spectral Frequency) parameters have an order property. In this paper, we divide the LSF parameters into three sub-vectors, and predict them by the vectors of previous and next frames. An efficient and reliable LSP vector distance measure is proposed for this interpolation algorithm. The interpolated LSP vectors are utilized in our mixed excitation LPC vocoder, which is operated at 1.5 kbps. Informal listening tests indicate that the synthesized speech sounds natural and intelligible.

1 Introduction

In low bit rate speech coding, the spectrum envelope is regarded as one of the most important characteristics of the speech signal. Line Spectral Frequency (LSF) are frequently used in the quantization of LPC parameters. Nearly transparent quantization of the LSF parameters can be achieved by scalar quantization, with 34 bits per frame [1]. Recently, more efficient vector quantization technique has been proposed, which requires only 24 bits per frame, with performance comparable to the 34 bits scalar quantizer [2]. For low bit rate coding, usually both frame interpolation and frame rate reduction have to be applied. Low bit rate speech coding at 1.2 kbps using frame interpolation technique has been reported in [3]. However, the large frame size (such as 25 ms) introduces larger time delay in frame interpolation, and the performance of the LPC synthesis filter will degrade considerably. Also, the computation complexity becomes much higher if frame interpolation is employed. The later problem is partially solved by employing a new distortion measure reported in [4].

This paper proposes a new LSF parameters interpolation technique for low bit rate speech coding. In the proposed scheme, the 10 LSF parameters are grouped into several sub-vectors. They are then predicted with minimum vector distance between original and interpolated individually. The interpolated LSF parameters are utilized in our mixed excitation LPC vocoder which is operated at very low bit rate. It generates natural sounding speech at the expense of slightly increased bit rate (from 1.2 kbps to 1.5 kbps).

The paper is organized as follows: section 2 briefly reviews the principle of frame parameters interpolation in very low bit rate speech coding. In section 3, the proposed sub-vector LSF parameters interpolation technique is discussed. Section 4 gives the experiment results and the performance of the proposed algorithm in the mixed excitation LPC vocoder. Finally, we summarize the results in the conclusion.

2 Frame parameters interpolation

To reduce the bit rate below 2.0 kb/s, we use both frame interpolation and frame rate reduction. Since short time speech is a non-stationary signal, analysis with larger frame size may not be able to track the pitch of the speech signal. For the same reason, the 10th short-term LPC filter cannot model the speech accurately. The larger the window or frame size, the larger the error between the original and the synthesized speech signal. Therefore we choose the frame size to be 22.5 ms. In a previous work [4], we have developed a 1.2 kb/s mixed excitation LPC vocoder [5][6] using the frame interpolation technique and a new spectral distortion measure. In the odd frames, all the parameters (LSFs, pitch, energy and v/uv decision) are quantized and transmitted. While in even frames, only the pitch and the v/uv decision are directly quantized and transmitted to the receiver. The LSF parameters and the energy are interpolated from the previous and next frames. The interpolation index is transmitted to the receiver instead of the LSFs and energy. In the receiver, the LSFs and energy are reconstructed by interpolation. The accumulative spectral dis-tortion (dB) is frequently used as the performance measure:

$$SD = \sqrt{\frac{1}{F_s}\int_0^{F_s}\left|10\log_{10}\frac{P(f)}{\hat{P}(f)}\right|^2 df} \quad , \qquad (1)$$

where F_s is the sampling rate, $P(f)$ and $\hat{P}(f)$ represent the power spectra of the original and the interpolated LPC filters, respectively. Here $\hat{P}(f)$ is linearly interpolated by the previous and the next LSFs vectors as follows:

$$lsf_j(i) = lsf_{j-1}(i) + \left[lsf_{j+1}(i) - lsf_{j-1}(i)\right]\frac{l}{K-1}$$

$$l = 0,1,2,\ldots,K-1 \ , \qquad (2)$$

where K is an integer which is a power of 2, $lsf_j(i)$ is the i^{th} LSF in the j^{th} frame. The index with minimum SD in LSFs interpolation is selected as the interpolation index $k_{\min}$, which is also used to reconstruct the energy E_j:

$$E_j = E_{j-1} + (E_{j+1} - E_{j-1})\frac{k_{\min}}{K-1} \quad . \qquad (3)$$

Since (1) has to be calculated K times per frame, it requires a lot of computation. In [4], we have replaced (1) by a more efficient LSFs distance measure:

$$Dist(f,\hat{f}) = \sum_{i=1}^{10}\left[g_i w_i (f_i - \hat{f}_i)\right]^2 \ , \qquad (4)$$

where $f_i, \hat{f}_i$ are the i^{th} LSFs in the original and interpolated LSFs vectors, respectively, w_i is a weighting function derived from the LPC spectral envelope:

$$w_i = \left|P(f_i)\right|^{0.2} \qquad (5)$$

and g_i is the perceptual weight given by:

$$g_i = \begin{cases} 1.0 & 1 \le i \le 8 \\ 0.8 & i = 9 \\ 0.4 & i = 10 \end{cases} \qquad (6)$$

Thus, the LSFs which are near the spectral formants will have more weights than those that are in the spectral valleys. The perceivable distortion will therefore be greatly reduced.

3 Proposed sub-vector interpolation

Since (2) is a global LSF interpolation, it cannot always generate satisfactory interpolation parameters when neighboring frame parameters change rapidly. As a shorter frame size generates much smaller distortion in the interpolation, we choose the frame size to be 20 ms. Because the formants of speech spectrum envelope determine the intelligibility of the synthesized speech, it is more important to predict the formants than the spectrum valleys. Note that formants usually depend on the closeness of LSP line pairs. Statistical results indicate that there are up to 3 or 4 formants in speech spectrum envelope. Therefore, it is reasonable to divide the 10 LSF parameters into three groups or sub-vectors and perform the interpolation individually. The grouping that we have used are: the first three LSFs, the next three LSFs and the last four LSFs. Figure 1 displays the LPC spectrum envelope and their LSPs.

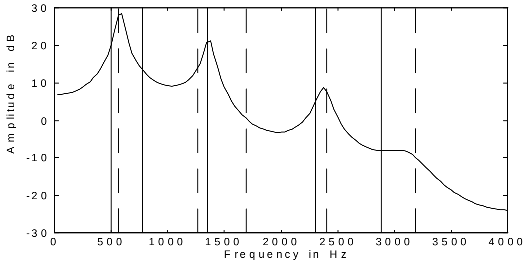

Fig 1 LPC spectral envelope with LSPs

—— odd ······ even

The proposed sub-vector interpolation is performed as follows: in odd frames, the codebook indexes of the vector quantized LSF parameters are transmitted to the decoder. In even frames, three current LSF sub-vectors are predicted individually, and the three interpolation indexes are transmitted to decoder. In the decoder, each grouped LSF sub-vector is reconstructed from its interpolation index. The accumulative spectral distortion (dB) measure used is the same as equation (1). The $\hat{P}(f)$ is linearly interpolated from the parameters in the previous and next frames:

$$\tilde{F}_m^c = F_m^p + \alpha_m * (F_m^n - F_m^p) \qquad m = 0,1,2 \quad (7)$$

where F_m^p, F_m^c, F_m^n are the m^{th} LSF sub-vectors of the previous, current and next frames, respectively. $\tilde{F}_m^c$ is the current LSF sub-vector to be interpolated. And α_m is the m^{th} interpolation coefficient:

$$\alpha_m = \frac{l}{K_m - 1} \qquad l = 0,1,2,\ldots K_m - 1\,, \qquad (8)$$

where K_m is an integer which is a power of 2. The selection of index α_m satisfies two conditions: first it satisfies the minimum spectral distortion in equation (1). Since α_m may not be the same in the three LSF sub-vectors, the selection of α_m has to satisfy the order property of LSFs:

$$0 < lsf_1 < lsf_2 < \ldots < lsf_{10} < \pi \qquad (9)$$

We propose a simple vector distance measure which utilizes the closeness of neighboring LSFs:

$$Dist_m = (\tilde{F}_m - F_m)^T W_m (\tilde{F}_m - F_m) \quad m = 0,1,2 \qquad (10)$$

Here, $Dist_m$ is the m^{th} vector distance between the predicted vector $\tilde{F}_m$ and the original vector F_m. W_m is a diagonal weighting matrix, whose coefficients depend on the closeness of neighboring LSF parameters:

$$W_{1,1} = \frac{1}{lsf_2 - lsf_2}$$

$$W_{i,i} = \frac{1}{\min(lsf_i - lsf_{i-1}, lsf_{i+1} - lsf_i)} \qquad 2 \le i \le 9$$

$$W_{10,10} = \frac{1}{lsf_{10} - lsf_9} \qquad (11)$$

It should be noted that equation (10) requires the least computation, and is simple to be implemented than equations (1) and (4).

4 Experiment results

The new LSFs intraframe interpolation technique is tested on 8000 frames of both clean and telephone speech. We choose 8000 Hz as the sampling frequency and 20ms as the frame size. We calculate the average SD and the percent of outliers (>2dB). The two distortion measures for frame interpolation equations (1) and (10) are compared in Table 1:

database	method	SD (dB)	>2 dB (%)
clean speech	original	1.06	6.85
	proposed	1.18	7.07
telephone speech	original	0.962	4.95
	proposed	0.987	5.25

Table 1 Spectral distortion in LSFs interpolation

It can be seen that intraframe interpolation introduces extra spectral distortion compared with direct quantization. Also, there are no significant differences between the two distortion measures. The slight increase of average SD in the proposed measures is due to the limitation of the linear interpolation. We also found that the LSFs interpolation is sensitive to frame size. When the frame size is enlarged to 25ms or more, the interpolation will introduce large spectral distortion. The proposed scheme offers another trade-off between bit rate and speech quality in very low bit rate speech coding. Figure 2 shows an example of the LSFs interpolation between the previous and the next frames.

The LSP intraframe interpolation technique has been utilized in our mixed excitation LPC vocoder which is operated at very low bit rate (1.2 kbps) [4]. In odd frames, we allocate 24 bits to the 10 LSF parameters. While in even frames, we use sub-vector LSFs interpolation, and allocate 3 bits to each of the three interpolation indexes. We can design a 1.5 kb/s mixed excitation LPC vocoder.

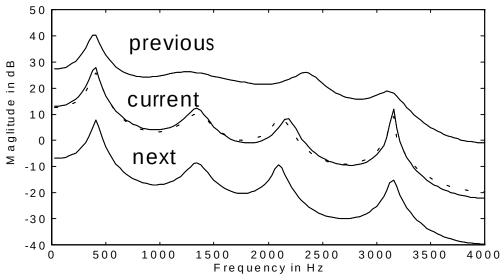

Fig 2 LPC spectral envelope interpolation
—— original - - - - interpolated

Furthermore, informal listening tests are conducted to evaluate the performance of the proposed intraframe interpolation. Listeners are required to give their preferences after they compare the speech quality of the 1.5 kb/s mixed excitation LPC vocoder and the 2.4 kb/s LPC-10e vocoder. The results indicate that our proposed vocoder generates natural sounding and intelligible speech, and is better than that of the 2.4 kb/s LPC-10e vocoder.

5 Conclusion

In this paper, we present a new LSP intraframe interpolation technique for low bit rate speech coding. The LSP parameters are grouped into three sub-vectors of size 3, 3 and 4. They are then predicted and interpolated. A fast and reliable distance measure is also proposed to reduce the complexity of the frame interpolation. The new technique has been applied to the mixed excitation LPC vocoder [4] to obtain a coder working at 1.5 kb/s. Informal listening tests indicate that the synthesized speech sounds natural and intelligible. The speech quality is improved at the expense of slightly increased bit rate.

Reference

[1] Sugamura and N. Farvardin, “ Quantizer design in LSP speech analysis-synthesis,” IEEE Journal on Selected Areas in Communications, Vol.6, No. 2, pp432-440, Feb. 1988.

[2] K. K. Paliwal and B. S. Atal, “ Efficient vector quantization of LPC parameters at 24 bits/frame,” IEEE Trans. Speech and Audio Processing, Vol.1, No.1, pp.15-14. Jan. 1993.

[3] S. Yeldener, A. M. Kondoz and B. G. Evans, “Multiband linear prediction speech coding at very low bit rates,” IEE Proc. -Vis. Image Signal Process., Vol. 141, No.5, pp.289 -296, October 1994.

[4] J. S. Mao, S. C. Chan and K. L. Ho, “ A mixed excitation vocoder operating at very low bit rate”, Proc. of IEEE Conference on Universal Personal Communications, Oct. 1997

[5] J. Makhoul , R. Viswanathan, R. Schwartz, and A. W. F. Huggins, “ A Mixed -Source Model for Speech Compression and Synthesis,” J. Acoust. Soc. Amer., Vol.64, pp.1577-1581, Dec 1978.

[6] A. V. McCree and T. P. Barnwell III, “A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding,” IEEE Trans. Speech and Audio processing, Vol. 3, pp.242-250, July 1995.