INTERPOLATION PROPERTIES OF LINEAR PREDICTION PARAMETRIC REPRESENTATIONS

K.K. Paliwal

School of Microelectronic Engineering
Griffith University
Brisbane, QLD 4111, Australia

ABSTRACT

In this paper, interpolation of linear predictive coding (LPC) parameters in terms of the following representations is investigated: linear prediction coefficient representation, reflection coefficient representation, log-area-ratio representation, arc-sine reflection coefficient representation, cepstral coefficient representation, line spectral frequency representation, autocorrelation coefficient representation and impulse response representation. Though each of these representations provides equivalent information about the LPC spectral envelope, their interpolation performance is found to be different. It is shown that the line spectral frequency representation results in the best interpolation performance.

1. INTRODUCTION

Linear predictive coding (LPC) parameters are widely used in various speech coding applications for representing the short-time spectral envelope information of speech [1]. In these applications, these parameters are obtained from the speech signal on a frame-by-frame basis and are quantized prior to their transmission. For low bit rate speech coding applications, it is important to devise methods for efficient quantization and interpolation of these parameters. Quantization of LPC parameters has been extensively studied in literature [2, 3]. In the present paper, we are interested in the interpolation of LPC parameters. Interpolation of LPC parameters helps in transmitting them less often (i.e., at a lower frame rate), thus reducing the bit rate[1]. However, it introduces distortion in LPC spectral envelope. This distortion increases as the frame rate is decreased.

In a low bit rate speech coder, LPC analysis is generally carried out at frame intervals of 10 to 25 ms, which means that the LPC parameters are transmitted at a rate of 40 to 100 frames/s. The lower frame rate can lead to large changes in LPC parameter values in adjacent frames which may introduce undesired transients (or, clicks) in the reconstructed (or, synthesized) speech signal. Therefore, to support lower frame rate, interpolation of LPC parameters is used at the receiving end to get smooth variations in their values. Usually, interpolation is done linearly at a few equally-spaced time instants (called subframes) within each frame. The LPC parameters can, in principle, be interpolated on a sample-by-sample basis. However, it is not necessary to perform such a fine interpolation. In addition, it is computationally expensive. Linear interpolation is generally done at a subframe interval of every 5 ms.

---

[1]Note that interpolation of LPC parameters is useful for speech synthesis, too [4, 5].

There are a number of LPC parametric representations proposed in the literature for quantizing LPC information. These representations are: linear prediction (LP) coefficient representation, reflection coefficient representation, log-area-ratio representation, arc-sine reflection coefficient representation, cepstral coefficient representation, line spectral frequency representation, autocorrelation coefficient representation and impulse response representation. Though each of these representations provides equivalent information about the LPC spectral envelope, their quantization performance is different. For example, it has been observed that the line spectral frequency representation performs best for scalar as well as vector quantization of LPC information [2, 3].

The aim of the present paper is to study these LPC parametric representations for interpolation. Though a few studies have been reported in the literature [6, 7, 8, 9, 10] where some of these representations are investigated for interpolation. For example, Itakura et al. [6, 7] and Atal et al. [8] have studied log-area-ratio, arc-sine reflection coefficient and line spectral frequency representations for interpolation and found the line spectral frequency representation to be the best. We investigate in this paper the interpolation performance of all of these LPC parametric representations and report their results in terms of spectral distortion measure which is defined as the root mean square difference between the original LPC log-power spectrum and the interpolated LPC log-power spectrum.

Some of the LPC parametric representations (LP coefficient, cepstral coefficient and impulse response representations) may result in an unstable LPC synthesis filter after interpolation. If these representations are used for interpolation, the LPC parameters after interpolation must be processed to make the resulting LPC synthesis filter stable. This processing is computationally expensive and, hence, these unstable representations should not be used for interpolation, if possible. However, some of the popular speech coding systems reported in the literature [11] have used the unstable LP coefficient representation for interpolation. Therefore, we use the number of unstable subframes resulting from the interpolation process as another measure of interpolation performance.

2. DATA BASE

The speech data base used in this study consists of about 2 minutes of speech from 32 speakers (16 male and 16 female). The speech signal is analyzed every subframe (at the rate of 200 subframes per second) using a 10th order LPC analysis. Duration of the LPC analysis window is 20 ms. For LPC analysis, the autocorrelation method with

20 ms Hamming window is used and 10 LPC coefficients are computed for every subframe.

3. OBJECTIVE CRITERIA USED FOR PERFORMANCE EVALUATION

Spectral distortion measure is commonly used for evaluating the performance of LPC quantizers. We use this measure for evaluating the interpolation performance of different LPC representations. As done in LPC quantization studies, we provide here results in terms of average spectral distortion and percentage of outlier subframes having spectral distortion greater than 2 dB. The outlier frames are divided into the following two types: Type 1 consists of outlier subframes in the range 2–4 dB, and Type 2 consists of outlier subframes having spectral distortion greater than 4 dB.

Spectral distortion (D_i) for the ith subframe is defined (in dB) as follows:

$$D_i^2 = \frac{1}{F_s} \int_0^{F_s} [10 \log_{10}(P_i(f)) - 10 \log_{10}(\hat{P}_i(f))]^2 df, \quad (1)$$

where F_s is the sampling frequency in Hz, and $P_i(f)$ and $\hat{P}_i(f)$ are the LPC power spectra of the i-th subframe. These are given by

$$P_i(f) = 1/|A_i(\exp(j2\pi f/F_s))|^2, \qquad (2)$$

and

$$\hat{P}_i(f) = 1/|\hat{A}_i(\exp(j2\pi f/F_s))|^2, \qquad (3)$$

where $A_i(z)$ and $\hat{A}_i(z)$ are the original and interpolated LPC polynomials, respectively, for the ith subframe. Spectral distortion is evaluated for all the subframes and its average value and percentage of outlier subframes are computed.

As mentioned earlier, we also use number of unstable subframes resulting from the interpolation process as another criteria for performance evaluation. For this, we perform stability check for each interpolated LPC polynomial $\hat{A}_i(z)$ and count the percentage of unstable subframes.

3. EXPERIMENTS AND RESULTS

In this section, we study the interpolation properties of different LPC parametric representations and provide results in terms of average spectral distortion, and percentage of outlier and unstable subframes.

As mentioned earlier, interpolation of LPC parameters is done linearly. Let $\mathbf{a}(i)$ and $\hat{\mathbf{a}}(i)$ be the original and interpolated LP coefficient vectors for the ith subframe. Also, let M be the number of subframes in a frame. Then linear interpolation of LP coefficients between two frames is achieved using the following equation:

$$\hat{\mathbf{a}}(i + j) = \frac{(M - j)\mathbf{a}(i) + j\mathbf{a}(i + M)}{M}, \qquad (4)$$

for $1 \leq j \leq (M - 1)$. If N is the total number of subframes in the speech data base, then Eq. (4) results in $(N-M)(M-1)$ interpolated LP coefficient vectors. These are used to measure interpolation performance of different LPC parametric representations in terms of spectral distortion and number of unstable subframes.

We have studied interpolation performance for frame intervals of 10, 15, 20, 25 and 30 ms. Interpolation results are listed in Table 1, 2, 3, 4 and 5, respectively. We can see from these tables that the line spectral frequency representation provides the best interpolation performance in terms of spectral distortion. In addition, it always results in stable LPC synthesis filters after interpolation. The LP coefficient representation also provides good interpolation performance in terms of spectral distortion measure. But, since it causes some unstable subframes, it is not a good choice for interpolation.

4. SUMMARY

We have studied a number of LPC parametric representations for linear interpolation and investigated their performance in terms of spectral distortion and number of unstable subframes. The line spectral frequency representation has been found to provide the best interpolation performance.

REFERENCES

[1] P. Kroon and B.S. Atal, "Predictive coding of speech using analysis-by-synthesis techniques", in *Advances in Speech Signal Processing*, S. Furui and M.M. Sondhi, Eds. New York, NY: Marcel Dekker, 1991, pp. 141-164.

[2] K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Trans. Speech Audio Process.*, Vol. 1, No. 1, pp. 3–14, 1993.

[3] K.K. Paliwal and B.W. Kleijn, "Quantization of LPC parameters", in *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal, Eds. Amsterdam:Elsevier Science B.V., 1995.

[4] J.P. Olive and N. Spickenagel, "Speech resynthesis from phoneme-related parameters", *J. Acoust. Soc. Am.*, Vol. 59, pp. 993-996, 1976.

[5] V.H. Shadle and B.S. Atal, "Speech synthesis by linear interpolation of spectral parameters between dyad boundaries", *J. Acoust. Soc. Am.*, Vol. 66, pp. 1325-1332, 1979.

[6] T. Umezaki and F. Itakura, "Analysis of time fluctuating characteristics of linear predictive coefficients", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1986, pp. 1257-1260.

[7] N. Sugamura and F. Itakura, "Speech analysis and synthesis methods developed at ECL in NTT — From LPC to LSP", *Speech Commun.*, Vol. 5, pp. 199-215, June 1986.

[8] B.S. Atal, R.V Cox and P. Kroon, "Spectral quantization and interpolation for CELP coders", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1989, pp. 69-72.

[9] J.S. Erkelens and P.M.T. Broersen, "Analysis of spectral interpolation with weighting dependent on frame energy", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1994, Vol. 1, pp. 481-484.

[10] V. Goncharoff and M.K. Krolak, "Interpolation of LPC spectra via pole shifting", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1995, Vol. 1, pp. 780-783.

[11] I.A. Gerson and M.A. Jasiuk, "Vector sum excited linear prediction(VSELP)", in *Advances in Speech Coding*, B.S. Atal, V. Cuperman and A. Gersho, Eds. Boston: Kluwer Academic, 1990, pp. 69-80.

Table 1: Interpolation performance of different LPC parametric representations. Interpolation is done from LPC parameters computed from speech at a frame interval of 10 ms.

Parametric representation	Av. SD	Outliers (in %)		No. of unstable
LP coefficient representation	0.88	5.3	0.5	0.0
Reflection coefficient representation	1.00	7.5	1.4	0.0
Log area ratio representation	0.99	7.3	1.5	0.0
Arc-sine reflection coefficient representation	0.99	7.4	1.4	0.0
Cepstral coefficient representation	0.91	6.1	0.3	4.5
Line spectral frequency representation	0.86	5.0	0.3	0.0
Autocorrelation coefficient representation	0.92	6.1	0.7	0.0
Impulse response representation	0.97	6.5	0.3	11.1

Table 2: Interpolation performance of different LPC parametric representations. Interpolation is done from LPC parameters computed from speech at a frame interval of 15 ms.

Parametric representation	Av. SD (in dB)	Outliers (in %)		No. of unstable subframes (in %)
		2-4 dB	>4 dB	
LP coefficient representation	1.20	11.1	1.1	0.0
Reflection coefficient representation	1.35	14.0	2.8	0.0
Log area ratio representation	1.34	13.9	2.6	0.0
Arc-sine reflection coefficient representation	1.34	13.9	2.6	0.0
Cepstral coefficient representation	1.22	12.0	0.7	5.8
Line spectral frequency representation	1.17	10.5	0.7	0.0
Autocorrelation coefficient representation	1.30	13.1	1.7	0.0
Impulse response representation	1.30	14.5	0.8	14.2

Table 3: Interpolation performance of different LPC parametric representations. Interpolation is done from LPC parameters computed from speech at a frame interval of 20 ms.

Parametric representation	Av. SD (in dB)	Outliers (in %)		No. of unstable subframes (in %)
		2-4 dB	>4 dB	
LP coefficient representation	1.45	19.0	2.1	0.1
Reflection coefficient representation	1.63	21.5	4.6	0.0
Log area ratio representation	1.61	21.2	4.3	0.0
Arc-sine reflection coefficient representation	1.61	21.3	4.4	0.0
Cepstral coefficient representation	1.46	20.3	1.5	6.9
Line spectral frequency representation	1.41	18.4	1.4	0.0
Autocorrelation coefficient representation	1.58	21.7	3.5	0.0
Impulse response representation	1.54	23.0	1.9	16.8

Table 4: Interpolation performance of different LPC parametric representations. Interpolation is done from LPC parameters computed from speech at a frame interval of 25 ms.

Parametric representation	Av. SD (in dB)	Outliers (in %)		No. of unstable subframes (in %)
		2-4 dB	>4 dB	
LP coefficient representation	1.66	25.1	3.7	0.3
Reflection coefficient representation	1.87	27.0	6.9	0.0
Log area ratio representation	1.84	26.7	6.4	0.0
Arc-sine reflection coefficient representation	1.84	26.9	6.5	0.0
Cepstral coefficient representation	1.66	26.8	2.8	7.8
Line spectral frequency representation	1.61	24.6	2.9	0.0
Autocorrelation coefficient representation	1.84	27.7	6.2	0.0
Impulse response representation	1.74	29.4	3.3	19.1

Table 5: Interpolation performance of different LPC parametric representations. Interpolation is done from LPC parameters computed from speech at a frame interval of 30 ms.

Parametric representation	Av. SD (in dB)	Outliers (in %)		No. of unstable subframes (in %)
		2-4 dB	>4 dB	
LP coefficient representation	1.83	29.4	5.3	0.5
Reflection coefficient representation	2.06	30.9	9.2	0.0
Log area ratio representation	2.03	30.6	8.5	0.0
Arc-sine reflection coefficient representation	2.03	30.8	8.9	0.0
Cepstral coefficient representation	1.82	31.4	4.1	9.2
Line spectral frequency representation	1.77	28.9	4.2	0.0
Autocorrelation coefficient representation	2.06	31.6	9.0	0.0
Impulse response representation	1.91	34.0	4.9	22.2