Proc. IEEE Workshop Speech Coding (Delavan, WI), pp. 105-107, Sept. 2000

PARTIAL-ENERGY WEIGHTED INTERPOLATION OF LINEAR PREDICTION COEFFICIENTS

Tamanna Islam and Peter Kabal

Electrical & Computer Engineering
McGill University, Montreal, Canada

Abstract

This paper discusses the interpolation of linear prediction (LP) coefficients. The performance of LP analysis using different numbers of subframes and the choice of representation for the LP coefficients are studied. Interpolation is done by converting the LP coefficients in one of the following representations: line spectral frequencies, reflection coefficients, log area ratios, and autocorrelations. It is shown that good performance is obtained for line spectral frequencies and five subframes per frame. A new interpolation technique which incorporates partial frame energy is introduced. This technique generalizes the concept of energy weighting to different LP coefficient representations.

1 INTRODUCTION

Many low bit rate speech coders employ linear prediction (LP) to model the short-term spectrum for speech. The LP coefficients are obtained from standard LP analysis, based on blocks of input samples, typically representing 20 ms to 30 ms of speech. The LP spectrum is the spectrum of the all-pole synthesis filter generated by the LP coefficients.

In transition segments, a large change in spectral characteristics can occur in a short time interval. Abrupt changes in the LP parameters between frames can introduce artefacts in the reconstructed speech. To follow the changes in spectra or to smooth the spectral transition, the LP coefficients can be updated more frequently (decreasing the frame length). However, this can increase the bit rate needed to code the coefficients. Smoothness can also be achieved by interpolating the LP coefficients. Interpolation of the LP coefficients can provide improved speech quality without requiring the transmission of additional information.

In this paper we are interested in improving the performance of interpolation of linear prediction coefficients. Three aspects of interpolation method are studied: the effect of varying the number of subframes per frame, the choice of representation for the LP coefficients, and partial energy weighting for LP coefficient interpolation.

2 INTERPOLATION AND ITS PERFORMANCE

Let $\boldsymbol{a}^{(i)}$ be the LP coefficient vector for the ith frame. Linear interpolation can be performed as follows

$$\boldsymbol{b}^{(i)}(\alpha) = (1-\alpha)\,\boldsymbol{a}^{(i)} + \alpha\,\boldsymbol{a}^{(i+1)}, \qquad (1)$$

where the parameter α measures the relative position of the interpolated subframe, $0 \le \alpha < 1$. Usually a frame is divided into several equally sized subframes, and interpolation is done at the subframe level, i.e., the LP coefficients are held constant for each subframe. In that case, $\alpha = j/M$, where j is the subframe number (0 to $M-1$) and M is the number of subframes per frame. This is shown in the figure below for $M = 4$.

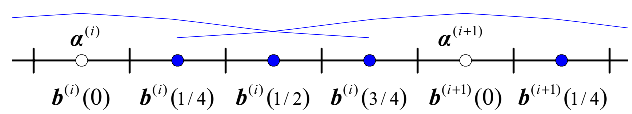


Fig. 1 Interpolation between consecutive frames

In Fig. 1, an open circle represents a reference subframe using the original LP coefficients. Superimposed on this figure is a window shape showing that the calculation of the LP coefficients, for instance $\boldsymbol{a}^{(i)}$, depends on samples to either side of the reference point. The filled circles represent subframes that use interpolated coefficients. These interpolated coefficients are formed as the linear combination of the LP coefficients of the reference subframes. The interpolated coefficients are kept fixed for the duration of each subframe.

To evaluate the effect of interpolation on the LP spectrum, LP coefficients for the intermediate subframes are also calculated. These serve as the "original" spectrum for calculating the spectral distortion. In addition the prediction gain is measured. This is the ratio of original signal energy to the prediction residual energy. The residual signal is calculated using interpolated LP filter coefficients.

Our baseline system uses the following parameters: 8 kHz sampling, 240 sample Hamming window, 160 sample frame size, and 10'th order autocorrelation analysis. Different LP coefficient representations are considered: line spectral frequencies (LSF's), reflection coefficients (RC's), log-area ratios (LAR's), and autocorrelation coefficients (AC's). In all of these representations, an appropriately weighted sum of LP coefficients corresponding to stable synthesis filters, gives an interpolated LP vector corresponding to a stable synthesis filter.

Our first inquiry is as to the choice of the number of subframes per frame. When the number of subframes is equal to 1 ($M = 1$), no interpolation is used. At the other extreme, if the number of subframes is equal to the number of samples in the frame, the LP coefficients change at every sample time. The table below shows the prediction gain that results for different numbers of subframes per frame for the different LP coefficient representations. These values are computed for a composite speech file containing both male and female speech. From the table, the prediction gain peaks at about 5 subframes per frame. The fact that the prediction gain is not monotonically increasing with the number of subframes may at first seem surprising. However, further examination of the configuration (see for instance Fig. 1) shows that for speech samples near the reference point, it is undesirable to contaminate the LP coefficients with coefficients from the next frame that have been calculated with a window that does not overlap the samples near the reference point.

Table 1 Prediction gain for different representations for LP coefficients for different numbers of subframes/frame.

M	LSF	RC	LAR	AC
1	16.45	16.45	16.45	16.45
2	16.47	16.44	16.45	16.44
4	16.50	16.46	16.47	16.46
5	16.53	16.49	16.50	16.48
8	16.52	16.48	16.50	16.48
10	16.52	16.48	16.49	16.47
16	16.52	16.48	16.49	16.47
20	16.52	16.48	16.50	16.48

The table of prediction gains shows that for conventional interpolation, LSF's outperform the other LP representations. This is consistent with previous results which have shown that LSF interpolation performs well in terms of spectral distortion [1, 2] and in listening tests [3].

3 THE EFFECT OF FRAME ENERGY

Our objective is to find how well the interpolation process models the intermediate subframes. In particular, we plot the spectral distortion (in this case for $M = 2$ and LSF parameters) below the subframe energy. Note that for

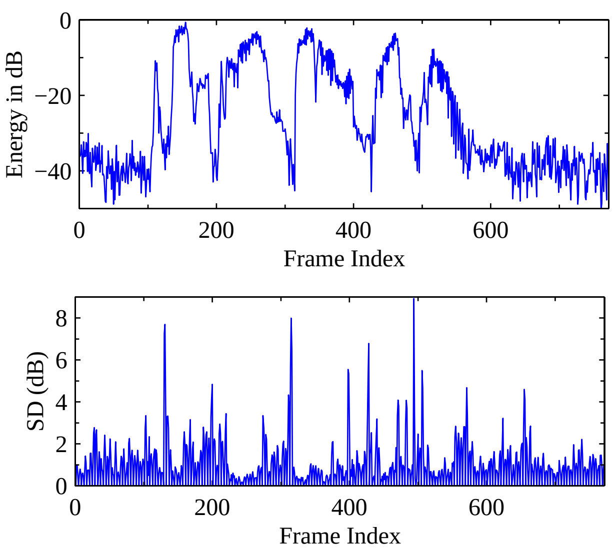


Fig. 2 Effect of change in subframe energy on spectral distortion

$M = 2$, for every second subframe the spectral distortion is zero (no interpolation is occurring). For the interpolated subframes, the spectral distortion increases at energy onsets. It has been previously suggested that interpolation should take into account the frame energy [4, 5]. Specifically, it has been suggested that interpolation should use the energy-weighted autocorrelation coefficients.

Interpolation of autocorrelation coefficients and interpolation of line spectral frequencies differ in quite a fundamental way. Consider two frames of speech, exhibiting sharp resonances at ω_1 and ω_2 respectively. Since the autocorrelation values are linearly related to the power spectrum, interpolation of the autocorrelation is equivalent to interpolation of the (linear domain) power spectrum. For the case cited, the interpolated spectrum will consist of the sum of the peaks. As the parameter α varies from 0 to 1, the contribution of the peak at ω_1 decreases while the contribution of the peak at ω_2 increases. Now consider, LSF interpolation. Peaks in the spectrum are generally signalled by closely spaced LSF's. Assuming the peaks are not too far apart (the same two LSF's contribute to the peaks for both frames), the interpolated coefficients will show a single peak at an intermediate frequency. The figure below shows the power spectrum for two frames. (This data has been contrived to show two separated peaks.)

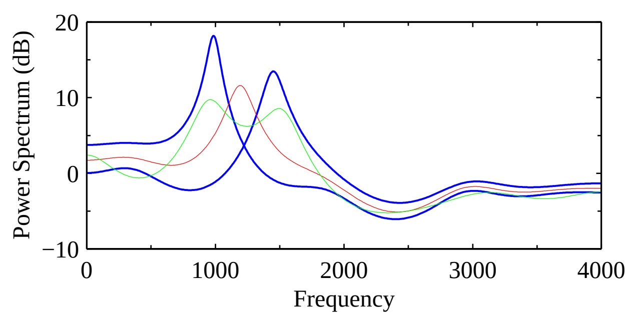


Fig. 3 Power spectrum for two frames (thick lines), along with LSF interpolation (single peak, thin line) and AC interpolation (double peak, thin line)

It seems clear that for steady voiced speech, LSF interpolation is to be preferred. The spectral peaks retain their bandwidth and shift smoothly in frequency. In contrast, autocorrelation interpolation, even for small shifts in the location of the peaks, will generate interpolated peaks which are wider in bandwidth. However, at voicing onsets where the position of excitation in the vocal tract changes, the type of modelling implied by autocorrelation interpolation may be preferred.

4 PARTIAL-ENERGY WEIGHTED INTERPOLATION

In this section, we generalize the concept of energy-weighting in the interpolation process. For the sake of exposition, consider interpolating autocorrelation coefficients. In this case, the autocorrelation coefficient vectors are normalized so that the first coefficient is unity. The first generalization is to allow partial-energy weighting,

$$E_\alpha \boldsymbol{R}^{(i)}(\alpha) = (1-\alpha)E_i^\gamma \boldsymbol{R}^{(i)} + \alpha E_{i+1}^\gamma \boldsymbol{R}^{(i+1)}. \qquad (2)$$

For $\gamma = 0$, we get interpolation of the normalized autocorrelation coefficients. For $\gamma = 1$, we get interpolation of the energy-weighted autocorrelation coefficients. Since the zero'th autocorrelation coefficient is normalized to unity,

$$E_\alpha = (1-\alpha)E_i^\gamma + \alpha E_{i+1}^\gamma. \qquad (3)$$

With this relation, the interpolation can be written in a simpler form,

$$\boldsymbol{R}^{(i)}(\alpha) = (1-\beta)\boldsymbol{R}^{(i)} + \beta \boldsymbol{R}^{(i+1)}, \qquad (4)$$

where

$$\beta = \frac{\alpha E_i^\gamma}{(1-\alpha)E_i^\gamma + \alpha E_{i+1}^\gamma}. \qquad (5)$$

In fact, this same interpolation can be applied to any LP representation,

$$\boldsymbol{b}^{(i)}(\alpha) = (1-\beta)\,\boldsymbol{a}^{(i)} + \beta\,\boldsymbol{a}^{(i+1)}. \qquad (6)$$

For $E_i = E_{i+1}$ or $\gamma = 0$, we get ordinary linear interpolation ($\beta = \alpha$). The non-linear effect of the energy differences is illustrated in the figure below. It plots β versus α for different values of the partial energy ratio $\rho = (E_i/E_{i+1})^\gamma$.

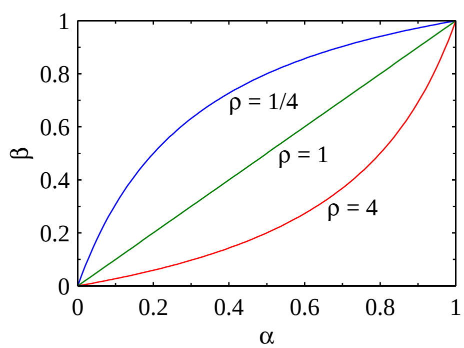
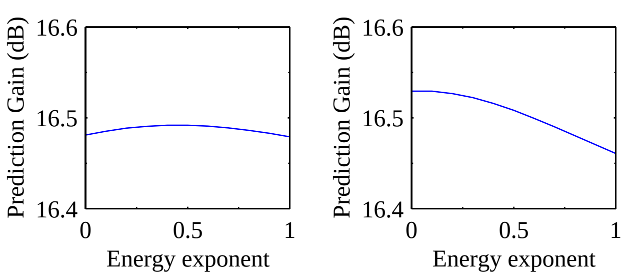


Fig. 4 Interpolation factor β as a function of subframe position α for different values of partial energy ratio.

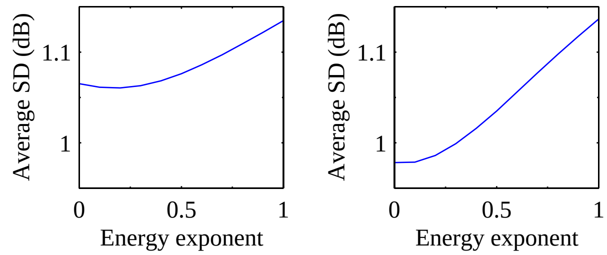
We now apply this new interpolation procedure to autocorrelation coefficients and to LSF's. The goal is to investigate the performance in terms of the partial-energy weighting exponent γ. The average prediction gain and spectral distortion are measured. For both prediction gain and



(a) AC interpolation (b) LSF interpolation

Fig. 5 Effect of the energy exponent γ on prediction gain

spectral distortion, no energy weighting seems to be best



(a) AC interpolation (b) LSF interpolation

Fig. 6 Effect of the energy exponent γ on the average spectral distortion

for LSF interpolation. For autocorrelation interpolation, the results are more mixed. In terms of prediction gains, an RMS weighting ($\gamma = 0.5$) is best. However, in terms of spectral distortion, a smaller energy weighting is best.

5 DISCUSSION

The results show that there is a best number of subframes per frame. Interpolation in the LSF domain produces the largest prediction gain and smallest average spectral distortion. Energy weighting is not particularly effective in terms of prediction gain or average spectral distortion. For autocorrelation interpolation, an RMS weighting ($\gamma = 0.5$) is better than full energy weighting. This study highlights the inadequacy of the performance measures used to assess LP spectral distortion. For a given level of spectral mismatch, the audible effect can be quite different depending on the relative energy of the frames and/or the surrounding context. An adaptive interpolation strategy (for instance, autocorrelation interpolation (with partial-energy weighting) at onsets and LSF interpolation in steady-state regions) would seem to be good solution.

References

[1] K. K. Paliwal, "Interpolation properties of linear prediction parametric representations," *Proc. European Conf. Speech Commun., Tech.* (Madrid), pp. 1029–1032, Sept. 1995.

[2] M. Yong, "A new LPC interpolation technique for CELP coders," *IEEE Trans. Commun.*, vol. 42, pp. 34–38, Jan. 1994.

[3] H. Choi, W. Wong, B. Cheetham, and C. Goodyear, "Interpolation of spectral information for low bit rate speech coding," *Proc. European Conf. Speech Commun., Tech.* (Madrid), pp. 1033–1036, Sept. 1995.

[4] J. Erkelens and P. Broersen, "LPC interpolation by approximation of the sample autocorrelation function," *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 569–573, Nov. 1998.

[5] J. S. Erkelens and P. M. T. Broersen, "Analysis of spectral interpolation with weighting dependent on frame energy," *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Adelaide), pp. I481–I484, April 1994.