

IMPROVED SPECTRAL TRACKING USING INTERPOLATED LINEAR PREDICTION PARAMETERS

Wesley Pereira

Peter Kabal

Department of Electrical & Computer Engineering
McGill University, Montreal, Canada

Abstract

Conventional interpolation of Linear Prediction (LP) parameters can provide a poor spectral match to the underlying speech signal for the intermediate subframes. In this paper, we present a method of modifying the interpolation endpoint LP parameters to improve the spectral tracking over *all* subframes. This ‘warping’ algorithm is based on minimizing the distortion between the interpolated LP parameters and those computed directly from the speech signal. The algorithm has been integrated into the Adaptive Multi Rate (AMR) speech codec. Our results show that this method enhances coder performance by smoothing out the LP parameter tracks and reducing coding distortion.

1 Introduction

LP parameters are used to represent the spectral envelope in many modern speech coding systems. These LP coefficients describe an all-pole synthesis filter and are computed for every frame (usually 20 ms) of speech data. For narrow-band speech (sampling rate of 8 kHz), a 10–12th order filter is typically used. Rapid fluctuations in the spectral envelope for transition segments result in large changes in the LP parameters which may produce distortion in the reconstructed speech. Therefore, LP parameters are generally interpolated for each subframe (typically 5 ms) to smooth out the variations. Line Spectral Frequencies (LSF’s) are often used to this end due to their desirable quantization and interpolation properties [1].

The interpolation endpoint or *terminal* LP parameters correspond to the subframe of speech for which the LP analysis was performed. Consequently, these LP coefficients yield a good match (given the limitations of the LP analysis technique) to the spectral envelope of the speech signal for that subframe. For the intermediate subframes however, the interpolated LP parameters can provide a poor spectral fit. By warping the terminal LP parameters, large spectral disparities in the intermediate subframes can be reduced. In addition, the frame-to-frame spectral dynamics will be improved.

Many methods to improve the dynamics of the spectral envelope described by the LP parameters have been proposed. Smoothing the quantized LSF vector at the decoder while keeping it in the same Voronoi region of the vector quantizer (VQ) reduces fluctuations in the LSF tracks and improves the perceptual quality [2]. Using a distortion measure with interframe memory for the VQ at the encoder can also reduce the unwanted variations present in

the quantized LSF tracks [3]. Both of these methods focus on smoothing the quantized LP parameters.

An LP analysis method which is based on maximizing the prediction gain when using the interpolated parameters to update the prediction filter at each subframe is presented in [4]. Since the derivation of the algorithm incorporates the interpolated parameters, a better spectral fit is obtained for the intermediate subframes. However, the method focusses on the use of direct form coefficients, which have poor quantization and interpolation properties.

The purpose of this paper is to introduce a method of modifying the terminal LP parameters in the LSF domain to improve the spectral tracking over the intermediate subframes. This warping algorithm is set up as an optimization problem in which the average distortion over all the subframes is minimized.

1.1 Distortion Measures

Spectral Distortion (SD) is a widely used measure to evaluate the performance of LP parameter quantizers. In [5], three conditions for transparent quantization of spectral information were derived based on experimental observations: 1) The average SD is less than 1 dB, 2) There is no outlier frame having an SD larger than 4 dB, and 3) The number of outlier frames having an SD in the range 2–4 dB is less than 2%. For this paper, the SD is computed using a 256 point FFT with 96 linearly spaced points between 125 Hz and 3.125 kHz. We used the SD to measure the performance of the warping algorithm and show a substantial decrease in the number of outlier frames.

Due to its non-linearities, the SD is not a practical distortion measure for optimization problems involving the LP parameters. Thus, we used the Weighted Euclidean LSF Distance given by:

$$d_{\text{LSF}}(\boldsymbol{\omega}, \hat{\boldsymbol{\omega}}) = \sum_{i=1}^{10} [c_i w_i (\omega_i - \hat{\omega}_i)]^2 \quad (1)$$

where $\boldsymbol{\omega}$ and $\hat{\boldsymbol{\omega}}$ are the reference and test LSF vectors, respectively, corresponding to the 10th order LP filter; and ω_i and $\hat{\omega}_i$ correspond to the i th LSF of the reference and test vectors, respectively, and are bounded by 0 and π . The fixed weights c_i are given by [5]:

$$c_i = \begin{cases} 1.0, & \text{for } 1 \leq i \leq 8, \\ 0.8, & \text{for } i = 9, \\ 0.4, & \text{for } i = 10. \end{cases} \quad (2)$$

We used the adaptive weights w_i proposed by Laroia

et al. [6]:

$$w_i = \frac{1}{\omega_i - \omega_{i-1}} + \frac{1}{\omega_{i+1} - \omega_i}, \quad (3)$$

with $\omega_0 = 0$ and $\omega_{11} = \pi$. The d_{LSF} has a high correlation with the SD [7] and is a linear function of the LSF's.

2 The Warping Method

To measure the spectral tracking efficiency using the interpolated LP parameters, an LP analysis was performed for each subframe. The resulting parameters are termed the *rapid analysis* parameters and served as the reference vector against which the interpolated parameters were compared.

For this section, a 10th order LP was performed using the autocorrelation method with a 25 ms Hanning window. The analysis was performed for 20 ms subframes, and the LSF's were linearly interpolated for each 4 ms subframe. A 60 Hz Gaussian lag window and a white noise correction factor of 1.001 were applied to the correlations.

Given the terminal LSF's for the previous frame (denoted $\tilde{\omega}^{(-1)}$), the goal of the warping algorithm is to determine the terminal LSF's for the current frame ($\tilde{\omega}^{(0)}$) to yield a good spectral match over all the subframes. Thus, a weighted sum of the d_{LSF} for every subframe is used:

$$d_{\text{TOT}} = \sum_{j=1}^I f_j d_{\text{LSF}}(\omega^{(j)}, \tilde{\omega}^{(j)}), \quad (4)$$

where $I = 5$ is the interpolation factor; f_j is the subframe weighting factor for the j th subframe; $\omega^{(j)}$ is the rapid analysis LSF vector for the j th subframe; and, $\tilde{\omega}^{(j)}$ is the interpolated LSF vector for the j th subframe and is given by:

$$\tilde{\omega}^{(j)} = (1 - \alpha_j)\tilde{\omega}^{(-1)} + \alpha_j\tilde{\omega}^{(0)}, \quad (5)$$

where $\alpha_j = j/I$. Minimization of d_{TOT} results in a set of $p = 10$ single variable quadratic equations. Equally weighting each subframe ($f_j = 1$ for $j = 1, \dots, I$) can yield large spectral mismatches in the following frame — modifying the terminal LSF's affects the interpolated LSF's for the previous and following frames. Thus, these weights were tuned over a large speech database to minimize the d_{LSF} and SD, and the optimized weights are shown in Table 1 (the first and third rows marked 'no lookahead'). The difference in the SD and the d_{LSF} optimized weights is due to the logarithmic relation between these two distortion measures [7].

Consider extending the d_{TOT} to include the case where rapid analysis LSF's for future subframes of the next frame are available:

$$d_{\text{TOT}} = \sum_{j=1}^I f_j d_{\text{LSF}}(\omega^{(j)}, \tilde{\omega}^{(j)}) + \sum_{j=1}^L l_j d_{\text{LSF}}(\omega_N^{(j)}, \tilde{\omega}_N^{(j)}), \quad (6)$$

where $L \leq I$ is the number of lookahead subframes; $\omega_N^{(j)}$ is the rapid analysis LSF vector for the j th subframe of the lookahead frame; l_j is the weighting factor for the j th subframe of the lookahead frame; and, $\tilde{\omega}_N^{(j)}$ is the interpolated LSF vector for subframe j of the lookahead frame and is given by:

$$\tilde{\omega}_N^{(j)} = (1 - \beta_j)\tilde{\omega}^{(0)} + \beta_j\tilde{\omega}^{(1)}, \quad (7)$$

where $\tilde{\omega}^{(1)}$ is the estimated LSF interpolation endpoint vector for the lookahead frame and β_j are the interpolation weights. Using $\beta_j = j/I$ and $\tilde{\omega}^{(1)} = \omega_N^{(L)}$ was found to be the most effective. This is equivalent to using the LSF's obtained from the last lookahead subframe (L th subframe of the lookahead frame) as the terminal LSF's for the lookahead frame, and minimizing the d_{LSF} between the interpolated and rapid analysis LSF's (over all the subframes in the current frame and the L subframes in the lookahead frame). Once again the minimization leads to a set of quadratic equations and the weights were tuned to reduce the average distortion (see the second and fourth rows in Table 1).

Table 1 Tuned subframe weights to minimize the average SD and d_{LSF} .

measure	f_1	f_2	f_3	f_4	f_5	l_1
d_{LSF}	0.15	0.55	0.41	0.02	1.00	no lookahead
	0.69	1.39	1.66	1.37	1.00	2.89
SD	0.06	0.11	0.14	0.00	1.00	no lookahead
	0.14	0.25	0.22	0.34	1.00	0.52

To provide a lower bound on the average SD and d_{LSF} , the optimal set of terminal LSF's was determined when there was infinite lookahead in the system. For the d_{LSF} , a closed-form solution exists whereas an iterative method must be used for the SD [7]. For the first iteration, the even frames are passed through in order and the terminal LSF's are adjusted to yield a minimum average distortion over the previous and following frames. For the next iteration, the process is repeated for the odd frames. By alternating between the even and odd frames, this iterative procedure necessarily converges to a local minimum.

Fig. 1 shows how the distortion decreases as the number of lookahead subframes increases. With only 1 lookahead subframe, the average distortion is significantly reduced relative to regular interpolation. The distortion results using the d_{LSF} and SD optimized weights with different amounts of lookahead are shown in Table 2. Warping the LSF's also improves the short-term and long-term prediction gains. The long-term gain was obtained by using a one-tap pitch filter updated every 5 ms using the covariance method.

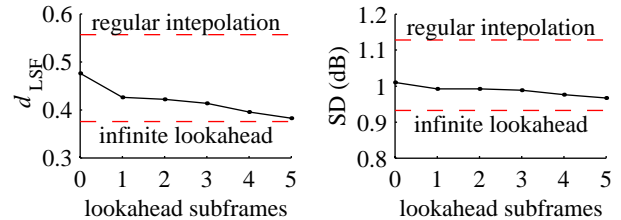


Fig. 1 The distortion performance of LSF warping as the number of lookahead subframes increases.

3 Simulation Results

This section details the implementation of the warping algorithm into the AMR speech codec. The simulations were performed using the 4.75 kbps mode of the coder.

Table 2 The prediction gains and distortion results using the warping algorithm.

		prediction gains			d_{LSF}	Spectral Distortion		
		short-term	long-term	overall		average	2–4 dB	> 4 dB
rapid analysis		11.26 dB	5.40 dB	16.66 dB	0.000	0.00 dB	0.00%	0.00%
regular interpolation		11.12 dB	5.19 dB	16.31 dB	0.595	1.02 dB	13.06%	1.05%
d_{LSF}	no lookahead	11.14 dB	5.18 dB	16.32 dB	0.477	1.03 dB	9.62%	0.57%
	one subframe lookahead	11.14 dB	5.18 dB	16.33 dB	0.427	1.02 dB	8.07%	0.34%
optimized	one frame lookahead	11.16 dB	5.21 dB	16.37 dB	0.383	0.99 dB	6.74%	0.23%
	infinite lookahead	11.15 dB	5.20 dB	16.34 dB	0.376	0.98 dB	6.50%	0.20%
SD	no lookahead	11.13 dB	5.20 dB	16.33 dB	0.526	1.01 dB	11.23%	0.85%
	one subframe lookahead	11.14 dB	5.20 dB	16.34 dB	0.465	0.99 dB	9.46%	0.58%
optimized	one frame lookahead	11.16 dB	5.21 dB	16.37 dB	0.407	0.97 dB	7.72%	0.38%
	infinite lookahead	11.14 dB	5.20 dB	16.34 dB	0.527	0.93 dB	7.01%	0.55%

3.1 Performance Measures

The following measures were used to evaluate the effect on performance of warping the LSF tracks in the AMR speech coder:

1. PWE_{tot} : The normalized perceptually weighted error energy (PWE_{tot}) is given by:

$$\text{PWE}_{\text{tot}} = \frac{\sum_n e_w^2[n]}{\sum_n s_w^2[n]}, \quad (8)$$

where $s_w[n]$ is the perceptually weighted speech signal and $e_w[n]$ is the perceptually weighted error. The PWE_{tot} was computed for each 5 ms subframe. Since the adaptive and fixed codebooks are searched by minimizing the energy of $e_w[n]$ over each subframe, a lower PWE_{tot} implies a higher coding efficiency.

2. Δw : The absolute difference between the terminal LSF's of successive frames is denoted by Δw . The difference is averaged over each of the 10 LSF's and over all the frames in units of Hz. A smaller Δw implies a smoother evolution of the LSF tracks and less quantization error when using appropriately optimized predictive quantizers for the LSF's.

SD and d_{LSF} were also used since the warping algorithm was derived by minimizing these distortion measures. Being a commonly used measure of speech quality, SNR_{seg} figures are also given.

3.2 AMR Setup

The 4.75 kbps mode of the AMR speech coder operates on 20 ms frames and has a lookahead delay of 5 ms. A 10th order LP analysis is performed for the fourth subframe using the hybrid Hamming-Cosine window given by:

$$w_d[n] = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{2L_1 - 1}\right), & n = 0, \dots, L_1 - 1, \\ \cos\left(\frac{2\pi(n - L_1)}{4L_2 - 1}\right), & n = L_1, \dots, L_1 + L_2 - 1, \end{cases} \quad (9)$$

where $L_1 = 200$ and $L_2 = 40$. A 60 Hz Gaussian lag window and a 1.0001 white noise correction factor are applied to the autocorrelations of the windowed speech. The direct

form filter coefficients are obtained using the autocorrelation method and are converted to LSF's. The LP filter is updated for every subframe by linearly interpolating the LSF's.

For the warping algorithm, the same window was used to obtain the rapid analysis parameters for the first three subframes ($L_1 = 200$ and $L_2 = 40$). For the lookahead subframe, the hybrid Hamming-Cosine window with $L_1 = 232$ and $L_2 = 8$ was used — no additional delay was thus incurred in obtaining the LP parameters for the lookahead subframe. The LP analysis window placement for all the subframes is shown in Fig. 2. The same lag window and white noise correction factor were applied to the autocorrelations for all the subframes.

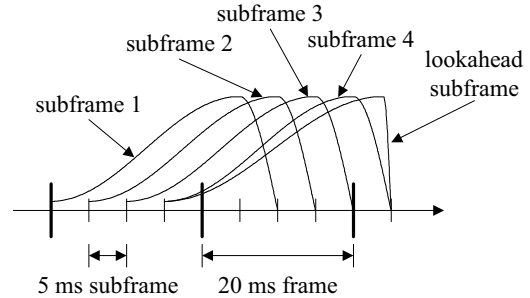


Fig. 2 The LP analysis window setup used to obtain the rapid analysis parameters in the AMR speech coder.

With this LP analysis setup, the subframe weighting factors were tuned to minimize the SD, d_{LSF} and PWE_{tot} , with and without using the lookahead subframe. These optimized weighting factors are given in Table 3. The optimization was done to within one decimal place. The SD optimized weights for the intermediate subframes are all zero when lookahead is used. Thus, the resulting LSF's are the same as those obtained with regular interpolation.

3.3 Results

Table 4 shows how the algorithm performed in the AMR speech coder. Using the PWE_{tot} optimized weights improved the performance of the coder in terms of the SNR_{seg} and the PWE_{tot} . In addition, the lower Δw associated with these weights implies a smoother evolution of the LSF

Table 4 Distortion results using different subframe weighting schemes in the AMR speech coder.

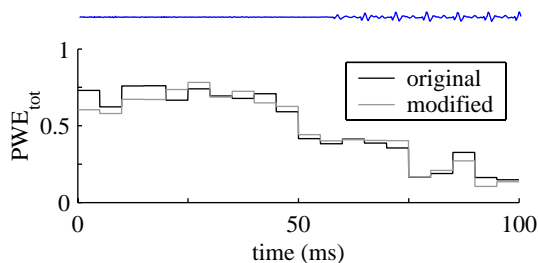
		d_{LSF}	Spectral Distortion			SNR _{seg}	PWE _{tot}	Δw
			average	2–4 dB	> 4 dB			
original AMR coder		0.70	1.06 dB	13.48%	1.95%	6.97 dB	0.476	73.8 Hz
no lookahead	d_{LSF} optimized	0.60	1.11 dB	10.96%	1.57%	6.98 dB	0.477	74.7 Hz
	SD optimized	0.70	1.06 dB	13.48%	1.95%	6.97 dB	0.476	73.8 Hz
	PWE _{tot} optimized	0.68	1.07 dB	12.84%	1.90%	7.00 dB	0.475	73.5 Hz
with lookahead	d_{LSF} optimized	0.57	1.10 dB	10.35%	1.39%	6.99 dB	0.476	72.9 Hz
	SD optimized	0.64	1.06 dB	12.05%	1.76%	7.00 dB	0.475	72.6 Hz
	PWE _{tot} optimized	0.64	1.09 dB	11.70%	1.66%	7.01 dB	0.474	72.4 Hz
infinite lookahead	d_{LSF} optimized	0.43	1.06 dB	8.40%	0.45%	7.03 dB	0.477	83.6 Hz
	SD optimized	0.55	1.00 dB	8.93%	1.18%	7.00 dB	0.476	80.4 Hz

Table 3 Optimal subframe weights to minimize the average SD, d_{LSF} and PWE_{tot} for the AMR speech coder.

measure	f_1	f_2	f_3	f_4	l_1
d_{LSF}	0.6	0.3	0.5	1.0	no lookahead
	1.2	0.8	1.0	1.0	2.0
SD	0.0	0.0	0.0	1.0	no lookahead
	0.2	0.0	0.2	1.0	0.4
PWE _{tot}	0.0	0.0	0.1	1.0	no lookahead
	0.0	0.0	0.6	1.0	2.0

tracks — the quantization error for the LSF’s could be reduced using predictive vector quantizers that are optimized accordingly. The largest improvement in the distortion measures shown was obtained when using the lookahead subframe. With the lookahead subframe, the increase in computational complexity relative to the original AMR coder was minimal — the execution time of the modified AMR coder was 7% longer.

Although the average PWE_{tot} is smaller using the warping scheme, there are large differences in the PWE_{tot} between the original and modified AMR coder for individual subframes. This is shown in Fig. 3. This suggests that extending the framework presented to yield a more robust approach could yield a more consistent improvement in performance from frame to frame.

**Fig. 3** The subframe to subframe fluctuations in the PWE_{tot} with and without warping the LSF’s in the AMR coder. The original speech segment (an unvoiced to voiced transition) is shown above.

4 Summary and Conclusions

In this paper, we introduced a novel method to warp the interpolation endpoint LSF’s in order to improve the spec-

tral match for the intermediate subframes. The warping algorithm is based on minimizing the Weighted Euclidean LSF Distance (d_{LSF}). Along with a substantial decrease in the average d_{LSF} , there were far fewer Spectral Distortion (SD) outlier subframes. Using the AMR speech codec as a testing platform, we showed an improvement in the SNR_{seg} for an increase of about 7% in computational complexity. The LSF tracks evolved more smoothly with the warping algorithm and the perceptually weighted error was reduced. Further improvements are expected if the coder were tuned to operate in conjunction with the algorithm.

References

- [1] T. Islam and P. Kabal, “Partial-energy weighted interpolation of linear prediction coefficients,” *Proc. IEEE Workshop on Speech Coding* (Delevan, Wisconsin), pp. 105–107, Sep. 2000.
- [2] H. Knagenhjelm and W. Kleijn, “Spectral dynamics is more important than spectral distortion,” *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Detroit, Michigan), pp. 732–735, May 1995.
- [3] F. Nordén and T. Eriksson, “A speech spectrum distortion measure with interframe memory,” *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Salt Lake City, Utah), 4 pp., May 2001.
- [4] T. Minde, T. Wigren, J. Ahlberg and H. Hermansson, “Techniques for low bit rate speech coding using long analysis frames,” *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Minneapolis, Minnesota), pp. 604–607, Apr. 1993.
- [5] K. K. Paliwal and B. S. Atal, “Efficient vector quantization of LPC parameters at 24 bits/frame,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-1, pp. 3–14, Jan. 1993.
- [6] R. Laroia, N. Phamdo and N. Farvardin, “Robust and efficient quantization of speech LSP parameters using structured vector quantizers,” *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Toronto, Canada), pp. 641–644, May 1991.
- [7] W. Pereira, *Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency*. Master’s thesis, McGill University, Montreal, Canada, 2001.
- [8] Global System for Mobile Communications (GSM). *Digital cellular telecommunications (Phase 2+); Adaptive Multi-Rate (AMR) speech transcoding (GSM 06.90 version 7.1.0 Release 1998)*, Jul. 1999.