A Noise-Robust Linear Prediction Analysis for Efficient Speech Coding

Aadel Alatwi[1], Stephen So[1], Kuldip K. Paliwal[1]

[1]School of Engineering, Griffith University, Brisbane, QLD 4111, Australia

aadel.alatwi@griffithuni.edu.au, s.so@griffith.edu.au, k.paliwal@griffith.edu.au

Abstract

In this paper, we propose a new linear prediction (LP) analysis method for estimating LP coefficients that are used in current speech coders. This method improves the robustness of LP coefficients computed from noise-corrupted speech as well as enabling better quantisation efficiency. The study compares the quantisation performance, as well as the noise-robustness between the proposed LP coefficients and the conventional LP coefficients, in terms of spectral distortion. The results indicate that the proposed LP coefficients were more robust to noise and also enabled transparent coding at lower bitrates (savings of up to two bits/frame) than the conventional LP coefficients.

Index Terms: linear prediction coefficients, linear prediction analysis, split vector quantisation, spectral distortion

1. Introduction

The linear prediction (LP) analysis method is extensively used as a basic technique for low bitrate speech coding applications [1]. In these applications, LP coefficients, which represent the short-time power spectrum of a speech signal using a low-order all-pole filter [2], are typically obtained using the autocorrelation method [3], converted into LP parameters (e.g. line spectral frequency (LSF)) and then quantised using as few bits as possible for transmission [4]. In noise-free environments, the performance of LP parameter-based speech coders is often satisfactory. However, in the presence of background noise, the LP analysis method yields a poor estimate of the LP spectrum of the input speech signal; hence, the variance of the estimated LP coefficients is increased [5], which results in an overall deterioration in the reconstructed speech quality [6][7].

A pth all-pole filter $H(z)$ that is driven by white Gaussian noise on the input is used to represent the speech production model:

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{1}$$

The filter coefficients (also known as LP coefficients) a_k and filter gain G are estimated by solving the Yule-Walker equations:[1]

$$\sum_{k=1}^{p} R(j-k) a_k = -R(j), \quad \text{for } k = 1, 2, \ldots, p \tag{2}$$

$$G = R(0) + \sum_{k=1}^{p} a_k R(k) \tag{3}$$

[1]It can be readily shown that all-pole modelling is equivalent to linear prediction analysis. Therefore, we will refer to this estimation process as *LP analysis* from now on.

where $R(k)$ are the autocorrelation coefficients estimated from a frame of N samples of the speech signal $x(n)$:

$$R(k) = \frac{1}{N} \sum_{k=0}^{N-1-k} x(n) x(n+k) \tag{4}$$

Another method of estimating autocorrelation coefficients utilises the Einstein-Wiener-Khintchine theorem, by taking the inverse discrete-time Fourier transform of the periodogram estimate of the power spectrum $P(\omega)$:

$$R(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) \mathrm{e}^{j\omega k} d\omega \tag{5}$$

$$\text{where} \quad P(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) \mathrm{e}^{-j\omega n} \right|^2 \tag{6}$$

This relationship between the periodogram and autocorrelation coefficients motivates our method for reducing the variance of the LP coefficient estimates, by smoothing the power spectrum before the LP analysis.

In this paper, we propose a modified procedure for performing LP analysis that reduces the variance of LP coefficient estimates in order to provide some noise robustness and also exploits the non-linear frequency selectivity of the human auditory system. The LP coefficients are fully compatible with current speech coding standards and the algorithm proposed can be easily incorporated into existing code implementations. We evaluate the quantisation performance of the proposed LP parameters using different quantisation schemes, as well as their robustness to noise in terms of the spectral distortion (SD). The results indicate that the proposed method provides a more accurate and noise-robust estimation of the LP parameters.

This paper is organized as follows: Section 2 describes the proposed LP analysis method. Section 3 outlines the experimental setup that we used to measure the quantization performance and the robustness of the LP parameters. Section 4 clarifies the details of each experiment and provides the results. Section 5 presents the conclusion of this study.

2. Proposed LP analysis method

The proposed LP analysis method uses two steps to compute the LP coefficients. The first step involves the manipulation of the periodogram of the input speech in order to reduce the variance of the spectral estimates and effects of noise. The second step applies the conventional autocorrelation method using the modified autocorrelation coefficients, which are computed by taking the inverse FFT of the processed power spectrum. The processed power spectrum is obtained through a smoothing operation using triangular filters that are spaced linearly on the Bark frequency scale [8]. These triangular filters simulate the human auditory system [9]. Also, there is a general downward spectral tilt in the speech power spectrum because there are more high energy formants located in the low frequencies and less energy peaks in the high frequencies. Therefore, the non-linear smoothing operation is performed less at low frequency regions and is performed more at high frequency regions [9], as shown in Figure 1.

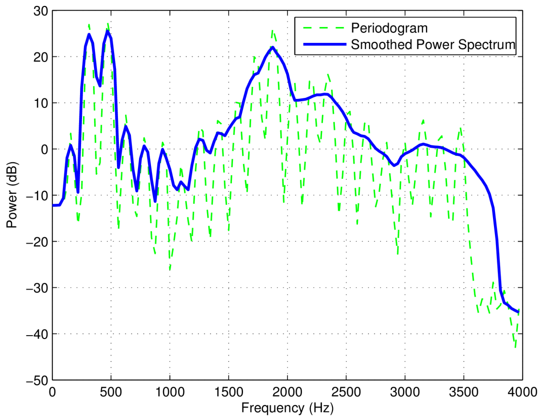

Figure 1: *Periodogram $P(k)$ and the smoothed power spectrum $\bar{P}(k)$ of speech sound (vowel / e / produced by male speaker).*

The proposed algorithm is described in the following steps:

Step 1: Compute the power spectrum $P(k)$ of a given frame $\{x(n),\quad n = 0,1,2,...,N-1\}$ of N samples from a speech signal using the FFT [10]:

$$P(k) = \frac{1}{N}\left|\sum_{n=0}^{M-1} x(n)w(n)\mathrm{e}^{-j2\pi kn/M}\right|^2, \quad 0 \leqslant k \leqslant M-1 \tag{7}$$

where $P(k)$ is the estimated power spectrum at the k^{th} normalized frequency bin, M is the FFT length where $M > N$, and $w(n)$ is a Hamming window.

Step 2: Determine the estimated smoothed power spectrum $\bar{P}(k)$ using a triangular filter at every frequency bin:

$$\bar{P}(k) = \sum_{i=-C(k)}^{C(k)} B(i)P(i-k) \tag{8}$$

where $B(i)$ is the triangular filter, which is spaced using the Bark frequency scale [8], and $C(k)$ is half the critical bandwidth of the triangular filter at frequency bin k.

Step 3: Take the inverse FFT of $\bar{P}(k)$ to compute the autocorrelation coefficients [10]:

$$\hat{R}(q) = \frac{1}{M}\sum_{k=0}^{M-1} \bar{P}(k)\mathrm{e}^{j2\pi kq/M} \quad , \quad 0 \leqslant q \leqslant M-1 \tag{9}$$

These autocorrelation coefficients $\hat{R}(q)$, $0 \leqslant q \leqslant p$, where p is the LP analysis order, are used in the Levinson-Durbin recursion algorithm [10] to compute the LP coefficients, which we called the Bark frequency Smoothed Linear Prediction (BS-LP) coefficients.

The behavior of the BS-LP analysis method in spectral modelling of speech signal is demonstrated in the example shown in Figure 2. In this figure, the periodogram $P(k)$ of a frame of the speech sound (vowel /e/) is shown together with the all-pole spectral models of order $p = 10$ that were computed with two techniques: the proposed BS-LP and the conventional autocorrelation LP analysis method. As illustrated in Figure 2, at high frequencies, the formants appear to have wider bandwidths due to the large critical bandwidths at these frequencies, where more smoothing is performed. Therefore, this added smoothing reduces the influence of noise components at higher frequencies.

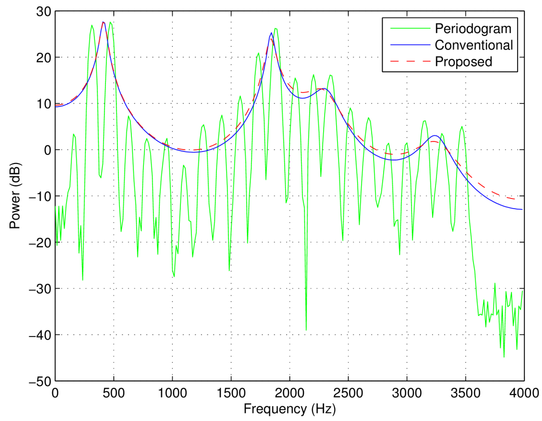

Figure 2: *Periodogram $P(k)$ with corresponding all-pole spectra of order $p = 10$ computed by the conventional LP and the proposed BS-LP analysis method of clean speech (vowel /e/ produced by male speaker).*

3. Experimental setup

3.1. Database

The TIMIT database was utilized for all of the simulations performed for this paper. It contains of 462 train speakers and 168 test speakers. The database was downsampled to 8 kHz. The estimation of the spectral envelope was carried out using the FFT length of 256 frequency samples. A 10^{th} order LP analysis was performed with high frequency compensation on a 20 ms analysis framework. A 10 Hz bandwidth expansion was applied because of sharp spectral peaks of LP spectrum that are caused by the underestimation of the formant bandwidths [4].

3.2. Performance evaluation criterion

In order to determine the quality of the power spectrum, the SD (spectral distortion) of the estimated spectral envelope was computed over the power spectrum of a frequency plane as an objective measure. It is defined as [4]:

$$SD = \sqrt{\frac{1}{F_s}\int_0^{F_s}[10\log_{10}P(f) - 10\log_{10}\hat{P}(f)]^2 df} \tag{10}$$

where F_s is the sampling frequency, and $P(f)$ and $\hat{P}(f)$ denote the true and estimated power spectra, respectively. As can be observed in Equation (10), a low SD indicates the reconstructed speech spectral envelope to be closer to that of the original speech, and therefore is of better quality. This distortion measure is carried out on the power spectrum produced from 20 – 30 ms frames of speech. The measure will be utilized to measure the accuracy and robustness of the proposed BS-LP analysis method in Section 4.1.

The number of bits allocated for quantisation influences the performance of the quantisation of the LP parameters. In many cases, the number of bits allocated for quantisation is found when the preferred rate of spectral accuracy has been achieved. This provides an equivalent basis of comparison between the proposed BS-LP analysis method and the conventional LP analysis method. To evaluate the quantisation process, the SD will be observed in two different classifications: the average SD for the whole data and the percentage of outlier frames. An outlier frame has an SD $\geq$ 2 dB. The outlier frames are divided into the following types: outlier frames with an SD in the range of 2 – 4 dB; and outlier frames with an SD greater than 4 dB. The preferred performance for the quantisation of the LP parameters is when transparent coding is achieved [4], which is defined by the following conditions:

1. The average SD is about 1 dB.
2. The number of outlier frames with an SD between 2 – 4 dB is less than 2%.
3. No outlier frames are greater than 4 dB.

4. Results and discussion

4.1. Noise robustness analysis

The robustness of the spectral envelope estimated using the BS-LP analysis method was compared with robustness found using the conventional LP analysis method. Simulations to find the robustness of the proposed method were carried out by measuring the SD between the power spectrum of the clean and noisy signal, respectively. The results that were obtained from this experiment are shown in Figure 3. The data shows that the BS-LP parameters appeared more robust to noise than the conventional LP parameters; the SDs of the BS-LP analysis method were consistently lower than the conventional LP analysis method for all Signal to Noise Ratios (SNRs). By referring to the example shown in Figure 2, this behaviour can be explained by the effectiveness of the smoothing operations in the computation of the BS-LP parameters.

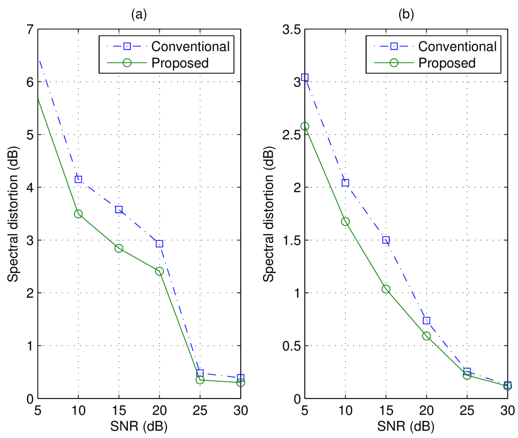

Figure 3: *Spectral distortion values (SD) between the conventional and proposed LP spectral envelopes of order $p = 10$ computed from clean and noisy speech (vowel / e / produced by male speaker). Speech was corrupted by two types of noise: (a) additive zero-mean Gaussian white noise, and (b) street noise, in six SNR categories.*

4.2. Quantisation performance of LP parameters

The full-search vector quantisation (VQ) has a high computational complexity that requires excessive memory space in order to perform the quantisation of the codebook. Though the split VQ mechanism is suboptimal, it lowers both the computational complexity and the required memory space to a manageable level without significantly affecting the performance of the VQ [11]. As a result, we used the split VQ to study the quantisation performance of the LP parameters.

In this split VQ, the LSF vector, which is a popular representation of the LP parameters, is separated into different parts of the lower order. The codebooks of the VQ were designed using the Linde-Buzo-Gray (LBG) algorithm [12] with the weighted LSF distance measure for every part, which is given by [4]:

$$d(f,\hat{f}) = \sum_{\mathbf{i=1}}^{\mathbf{10}} [c_i w_i (f_i - \hat{f}_i)]^{\mathbf{2}} \qquad (11)$$

where f_i and $\hat{f}_i$ are the i^{th} LSF representation in the approximated and original vector, respectively, and the weights w_i and c_i are assigned to the i^{th} LSF. The variable weight w_i is given by [4]:

$$w_i = [P(f_i)]^r \qquad (12)$$

where $P(f)$ is the LP power spectrum and r is equal to 0.15. The fixed weight c_i is given by [4]:

$$c_i = \begin{cases} 1.0 & \text{for } 1 \leq i \leq 8 \\ 0.8 & \text{for } i = 9 \\ 0.4 & \text{for } i = 10 \end{cases} \qquad (13)$$

We used two-split VQ (the first part had 4 LSFs and the second part had 6 LSFs) and three-split VQ (the first part had 3 LSFs, the second part had 3 LSFs, and the third part had 4 LSFs).

In all experiments, the quantisation performance for each method was evaluated using the SD measure, as given in Equation (10), where $P(f)$ and $\hat{P}(f)$ are the power spectrum of the original and reconstructed speech, respectively, and F_s covers the partial-band SD (i.e. 0 – 3kHz), which is used to evaluate the quantisation schemes that use a weighted distance measure [13].

The quantisation performances obtained from the two-split VQ experiments are listed in Tables 1 and 2 for the conventional LP and BS-LP analysis methods, respectively. The results indicate that the two-split VQ that uses the conventional LP analysis method required 24 bits/frame to achieve the transparent quantisation. However, the BS-LP analysis method required only 22 bits/frame, which saved 2 bits/frame from the conventional LP analysis method. The outlier frames percentage between 2 – 4 dB was much lower in favour of the proposed BS-LP analysis method.

Tables 3 and 4 illustrate the results obtained from the three-split VQ for the conventional LP and BS-LP analysis methods, respectively. It can be seen that the proposed BS-LP analysis method offers an advantage of 1 bit/frame over the conventional LP analysis method, which required 25 bits/frame to achieve the quantisation transparency.

5. Conclusion

This paper presented a modified method for estimating LP coefficients for current speech coders, by applying non-linear smoothing to the power spectrum. These BS-LP coefficients

Table 1: *Average SD of the two-split vector quantizer as a function of bitrate (using the conventional LP analysis method with the weighted LSF distance measure).*

Bits used	Av. SD (in dB)	Outliers (%) 2 – 4 dB	>4 dB
26	0.90	0.39	0.00
25	0.95	0.58	0.00
24	1.04	1.07	0.00
23	1.09	1.68	0.00
22	1.19	3.16	0.00
21	1.26	4.23	0.00
20	1.31	6.10	0.00

Table 2: *Average SD of the two-split vector quantizer as a function of bitrate (using the proposed BS-LP analysis method with the weighted LSF distance measure).*

Bits used	Av. SD (in dB)	Outliers (%) 2 – 4 dB	>4 dB
26	0.77	0.12	0.00
25	0.85	0.19	0.00
24	0.89	0.26	0.00
23	0.97	0.53	0.00
22	1.02	0.79	0.00
21	1.11	1.57	0.00
20	1.18	2.44	0.00

Table 3: *Average SD of the three-split vector quantizer as a function of bitrate (using the conventional LP analysis method with the weighted LSF distance measure).*

Bits used	Av. SD (in dB)	Outliers (%) 2 – 4 dB	>4 dB
30	0.78	0.20	0.00
29	0.80	0.27	0.00
28	0.85	0.51	0.00
27	0.88	0.56	0.00
26	0.97	0.91	0.00
25	1.05	1.73	0.00
24	1.18	3.10	0.01
23	1.21	3.83	0.01
22	1.30	5.91	0.03

possess improved robustness to noise. The advantage of this method is that it is fully compatible with current speech coder implementations. The LP quantisation performance of the BS-LP coefficients, in comparison with the conventional LP coefficients, was investigated for various split vector quantisation schemes. For both two- and three-split vector quantisation, the proposed BS-LP analysis method offered a saving of about 2 bits/frame and 1 bit/frame over the conventional LP analysis method, respectively, in terms of the spectral distortion. In addition, the results demonstrated the improved noise-robustness of the BS-LP coefficients for low to medium SNR levels and for both white and street noise. Since the Bark frequency triangular filters used for the power spectrum smoothing are similar to those used in the computation of Mel frequency cepstral coefficients (or MFCCs), these BS-LP coefficients may exhibit better recognition performance and noise robustness in automatic speech recognition tasks. We will investigate the use of BS-LP coefficients in the network speech recognition context in a future paper.

Table 4: *Average SD of the three-split vector quantizer as a function of bitrate (using the proposed BS-LP analysis method with the weighted LSF distance measure).*

Bits used	Av. SD (in dB)	Outliers (%) 2 – 4 dB	>4 dB
30	0.69	0.09	0.00
29	0.74	0.14	0.00
28	0.81	0.35	0.00
27	0.83	0.38	0.00
26	0.90	0.54	0.00
25	0.97	1.28	0.00
24	1.03	1.38	0.00
23	1.12	2.04	0.00
22	1.25	4.53	0.02

6. References

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[2] B. S. Atal, V. Cuperman, and A. Gersho, *Speech and Audio Coding for Wireless and Network Applications*. Springer Science & Business Media, 2012, vol. 224.

[3] A. El-Jaroudi, "Linear predictive coding," *Encyclopedia of Telecommunications*, 2003.

[4] W. B. Kleijn and K. K. Paliwal, *Speech coding and synthesis*. New York, NY, USA: Elsevier Science Inc., 1995.

[5] S. M. Kay, "The effects of noise on the autoregressive spectral estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, no. 5, pp. 478–485, 1979.

[6] M. R. Sambur and N. S. Jayant, "Lpc analysis/synthesis from speech inputs containing quantizing noise or additive white noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 6, pp. 488–494, 1976.

[7] A. Trabelsi, F. Boyer, Y. Savaria, and M. Boukadoum, "Improving lpc analysis of speech in additive noise," in *Circuits and Systems, 2007. NEWCAS 2007. IEEE Northeast Workshop on*. IEEE, 2007, pp. 93–96.

[8] H. Fletcher, "Auditory patterns," *Reviews of modern physics*, vol. 12, no. 1, p. 47, 1940.

[9] B. Moore, *An Introduction to the Psychology of Hearing*. Academic Press, 1997.

[10] M. H. Hayes, *Statistical digital signal processing and modeling*. John Wiley & Sons, 2009.

[11] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of lpc parameters at 24 bits/frame," *Speech and Audio Processing, IEEE Transactions on*, vol. 1, no. 1, pp. 3–14, 1993.

[12] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *Communications, IEEE Transactions on*, vol. 28, no. 1, pp. 84–95, 1980.

[13] S. So and K. K. Paliwal, "Efficient product code vector quantisation using the switched split vector quantiser," *Digital Signal Processing*, vol. 17, no. 1, pp. 138–171, 2007.