ROBUST LINEAR PREDICTION ANALYSIS FOR LOW BIT-RATE SPEECH CODING

Kuldip K. Paliwal and Nanda P. Koestoer

School of Microelectronic Engineering,
Griffith University, Brisbane, Australia, 4111
k.paliwal@me.gu.edu.au & n.koestoer@me.gu.edu.au

ABSTRACT

Spectral analysis of speech signals in noisy environments is an aspect of signal processing that deserves more attention. This paper uses some of the recently proposed robust linear prediction (LP) analysis methods for estimating the power spectrum envelope of speech signals. These methods provide robustness in spectral estimation without sacrificing their accuracy. These methods, which are the moving average, moving maximum, and the average threshold method, are compared to the more commonly used methods of LP analysis, such as the widely used autocorrelation method and the Spectral Envelope Estimation Vocoder (SEEVOC) method. The LP power spectrum envelope estimates from these methods are found to be more efficient for speech coding applications as well as more robust to noise.

1. INTRODUCTION

Linear prediction (LP) analysis has been used in the past in a number of applications such as speech coding, speech recognition and speaker recognition. Its most successful application is perhaps in speech coding where it is used to estimate the parameters of an all-pole model representing the envelope of the signal power spectrum [1]. To achieve linear prediction analysis that is more robust in noise-affected signals, a number of linear prediction analysis methods have been proposed recently by one of the authors [2]. These methods have shown great promise for speech coding and recognition application. These methods provide robustness by enhancing the spectrum of the *true* signal and ignoring the spectral parts affected by noise. In this paper, we provide some results which indicate that these methods provide better and more accurate estimation of LP parameters. Application of these methods for speech coding is also investigated and results about the quantization performance of LP parameters are provided.

2. CONVENTIONAL LP ANALYSIS METHODS

Most popular method of LP analysis is perhaps the autocorrelation method. This method uses an all-pole (or autoregressive (AR)) model for estimating the power spectrum. For a signal that becomes noisy, the AR model would not be the correct model. The noisy signal would follow an ARMA model. For estimating the spectral envelope of noisy signal, either we should assume an ARMA model for the signal or we should clean the signal prior to applying the autocorrelation method.

The Spectral Envelope Estimation Vocoder (SEEVOC) method is a technique that has been proposed to improve the performance of the conventional LP analysis method. The SEEVOC method uses only those parts of spectrum which are less noisy. Thus it tries to clean the spectrum of noisy signal by ignoring the spectral portions which are affected more by noise. Its analysis deploys a methodology that ignores the low-level spectral peaks that may be a result of noise or side-lobe effects [3]. This method seems to perform well on speech signals having low fundamental frequency. But, for signals having high fundamental frequency, it does not perform well [4]. Accuracy of its spectrum envelope estimation also depends heavily on *a priori* knowledge of the average signal pitch (for non-periodic waveforms), which is a complication in real-world applications.

3. ROBUST LP ANALYSIS METHODS

A number of robust linear prediction analysis methods have been proposed recently by one of the authors [2]. These methods compute the LP parameters in two steps: In the first step, they manipulate the FFT-computed power spectrum with the aim of removing the effect of noise. In the second step, they apply the conventional autocorrelation method on the autocorrelation coefficients computed by taking the inverse FFT of the clean power spectrum. By improving the estimation process of the power spectrum envelope, the accuracy of its linear prediction would then increase and, with regards to signal compression and coding, will further remove unnecessary redundancies in the code.

An example of a speech frame affected by noise can be seen in Figure 1. It can be seen that as noise is introduced, the lower-level peaks of the power spectrum are affected most. Generally, noise effects the power spectrum

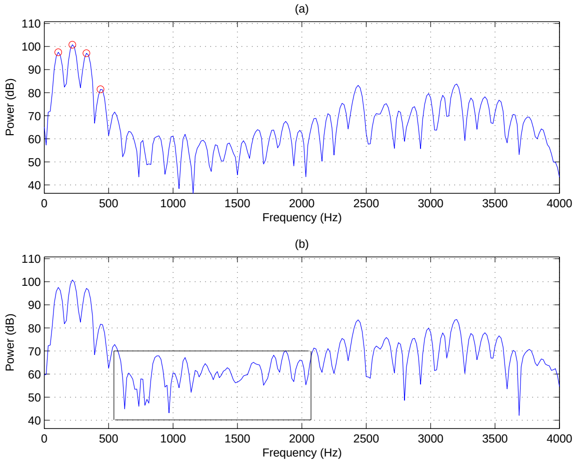

Figure 1: Power spectrum of speech for (a) clean signal (no noise) and (b) noisy signal (SNR=25 dB).

of speech signal in 2 areas: a) the space between the harmonic peaks (Figure 1a shows the first few harmonic peaks, marked with circles) and b) the non-formant regions of the spectrum (area inside the box in Figure 1b). Because of this, the LP spectrum of such a signal would be severely distorted, as LP analysis treats the high and low level peaks equally.

In order to overcome this problem, three new spectral envelope estimation methods are proposed; these are the average threshold (AT), moving average (MA) and the moving maximum (MM) method. These methods rely more on the harmonics peaks and ignore valleys between the harmonic peaks. Hence when noise is introduced, the estimated spectrum envelope would maintain the general shape of the power spectrum, whilst not being overly affected by the noise.

3.1. Moving Average Method

This method employs the moving average filter to smooth the FFT-computed power spectrum of the input signal. Using an M-size averaging window, the filter range can be defined as

$$w(i) = \frac{N - |i|}{N^2} \qquad (1)$$

for $-N \leq i \leq N$ and $N = \frac{M-1}{2}$.

3.2. Moving Maximum Method

This method searches for a maximum level from the FFT-computed power spectrum of the input signal over a defined range. The maximum point will then be used to represent a certain interval surrounding that frequency point.

Implementation of the moving maximum method is defined as follows:

- For each spectral point k in the frequency plane, the algorithm searches for a maximum value in the region of $[k - N, k + N]$.
- It then replaces the original value of that point with the resultant maximum value. The span of the moving maximum window would be (2N+1).

3.3. Average Threshold Method

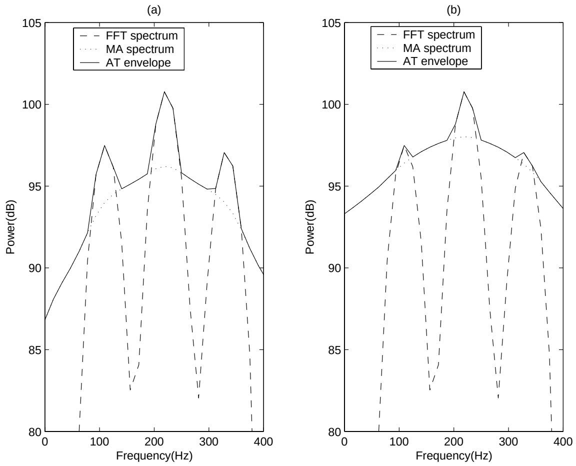

Figure 2: (a) Power spectrum envelope after first step of average threshold (AT) search method; (b) Power spectrum envelope after fourth AT repetition.

The average threshold method takes into account the benefit of both the moving average and moving maximum methods. It is based on a repetitive search of the FFT-computed power spectrum of its moving average spectrum, then taking its maximum in comparison to its original power spectrum envelope.

The methodology for the AT method is as follows:

- The moving average algorithm is applied on the FFT-computed power spectrum of the input signal.
- The resultant average spectrum is then combined with the original power spectrum. The maximum value between the two spectra at any given frequency locations is then used as the new average threshold spectrum.
- The steps above are then repeated a certain number of times to achieve an optimum result.

Figures 2a and 2b show how this method is performed over a number of repetitions.

4. SIMULATION RESULTS

4.1. Database

The *TIMIT* database is used for the majority of the simulations performed for this paper[1]. It consists of 462 train speakers and 168 test speakers with a male-female speaker ratio of 70:30. The database, which was originally sampled at 16 kHz with 16-bit resolution, has been re-sampled at 8 kHz with identical resolution. Spectrum envelope estimation is performed on power spectrum with FFT length of 512 frequency samples. LP analysis is performed with a 10^{th} order on 20 ms analysis frames. Bandwidth widening of 10 Hz is applied ($\gamma = 0.996$). The train-test vector ratio of approximately 8:1 is determined to be sufficient for quantization of LP parameters. The noise sample from the *Aurora* database[2] are used in our experiments to simulate real-world noise conditions. Each of the noise samples will be varied for 8 different SNR values (ranging from 35 dB, 30 dB, . . ., 5 dB, 0 dB).

4.2. Performance Evaluation Criterion

For determining the quality of an estimated power spectrum envelope, its *spectral distortion* (SD) is calculated over the power spectrum on a frequency plane as an objective measure.

$$SD = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(10log_{10}\frac{P_i}{\hat{P}_i})^2} \qquad (2)$$

where P_i and $\hat{P}_i$ are the true and estimated power spectrum respectively. As can be observed in the above equation, the reduction of SD dictates the spectrum envelope estimate's level of accuracy.

This distortion measure is normally performed upon the power spectrum generated from 20-30 ms length of speech, or equal to the frame length used for its linear prediction analysis. This measure will be used to determine the accuracy and robustness of the proposed methods in Section 4.3. The number of bits allocated for quantization determines the efficiency of quantization of the LP parameters. Generally the number of bits allocated for quantization is determined only when a desired level of spectral accuracy has been achieved. This will give an equal basis of comparison between the proposed methods and the more conventional linear prediction methods.

In order to measure the performance of the quantization process, the SD will be observed in two separate classifications, which is the average SD for the entire data, and the percentage of outlier frames. A frame is considered to be an outlier frame if its SD ≥ 2 dB. Outlier frames are divided into the SD ranges between 2-4 dB, and SD > 4 dB. A desired performance for the quantization of LP parameters is reached when its *spectral transparency* has been fulfilled [5], which is defined by the following criteria:

- average SD $\simeq 1$ dB,
- no outlier frame > 4 dB,
- number of frames with SD between 2-4 dB is less than 2 % of the number of total frames.

4.3. Robust Analysis

In this section, the robustness of the moving average (MA), moving maximum (MM), and average threshold (AT) methods of spectrum envelope estimation will be studied. The performance will be compared to the conventional spectrum envelope estimation design using the autocorrelation (AM) and SEEVOC methods.

Simulation to determine the robustness of the proposed spectral analysis methods is done by measuring spectral distortion between the power spectrum of the proposed method on clean signal and the spectrum of the same method on noisy signal. As the effect of noise upon the speech signal is increased, the spectrum envelope resulted from the proposed methods is expected to keep its general shape. As the aim of the development of the proposed methods is not to ensure complete isolation from noise, then the quality of the spectral analysis is sure to decrease as the effects of noise are increased. However the spectrum envelope generated by the proposed methods is expected to generally maintain its vigour even as the low-power peaks are masked by the power of the noise.

Figure 3 shows the robustness of the methods in comparison to each other. The SEEVOC method uses linear interpolation with a coarse pitch (CP) setting of 15 frequency samples for [$\frac{1}{2}CP$, $\frac{3}{2}CP$] range. The three proposed methods employ a window length of 21 frequency samples, and AT repetition of 3.

4.4. Split Vector Quantization of LP Parameters

Full-search VQ has a very high computational complexity and requires too much of memory space for the quantization codebook. Though the split VQ approach is suboptimal, it reduces computational complexity and memory requirements to manageable limits without affecting the VQ

[1] *This database was developed under the sponsorship of the United States Defence Advanced Research Projects Agency - Information Science and Technology Office (DARPA-ISTO), released in October 1990.*

[2] *This database was created at the Ericsson Eurolab Germany in January 2000. It consists of background noises that are commonly encountered such as sounds from people dining in a restaurant, cars and pedestrians on a street, trains, and motor vehicles.*

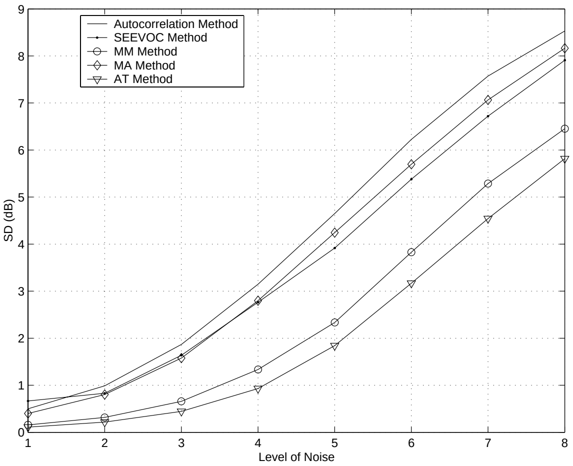

Figure 3: SD measurements for 30 ms speech vowel [e] introduced with Gaussian noise.

performance too much. Because of this, we use split VQ to investigate the quantization performance of the LP parameters. Since the line spectral frequency (LSF) representation of LP parameters is shown to be more effective for quantizing the LP information [6], we transform each LP vector to a LSF vector prior to its use in split vector quantization.

As the name suggests, the split VQ method divides an LSF vector into separate partitions of lower-order. The VQ codebooks are generated using the K-means algorithm for each part. We investigate the quantization performance of all the robust methods using split VQ and compare it with that of the autocorrelation method. We use split VQ with two parts (4 LSFs in the first part and 6 LSFs in the second part) and three parts (first part having 3 LSFs, second 3 LSFs and third 4 LSFs).

Tables 1 and 2 shows the quantization performance in terms of spectral distortion (SD) for each method using a split VQ with 3 and 2 partitions, respectively. In these experiments we use the same settings as defined in Section 4.3 and uses uniform bit allocation for individual parts. We can see from these tables that all the robust methods provide better quantization performance than the autocorrelation method. Also, the AT method is the best robust method.

Table 1: Quantization performance of different methods for 3 part split VQ.

Estimation method	*Number of bits/frame*	*Average SD (dB)*	*Outliers* (%)	
			2-4 dB	*>4 dB*
AM	26	1.28	6.06	0.03
	27	1.16	3.22	0.01
SEEVOC	26	1.15	2.88	0.02
	27	1.04	1.29	0.01
MA	26	1.20	3.68	0.02
	27	1.08	1.83	0.00
MM	26	1.13	2.48	0.02
	27	1.01	1.12	0.01
AT	26	1.10	1.87	0.02
	27	0.99	0.75	0.00

Table 2: Quantization performance of different methods for 2 part split VQ at 24 bits/frame.

Estimation method	*Average SD (dB)*	*Outliers* (%)	
		2-4 dB	*>4 dB*
AM	1.37	9.09	0.02
SEEVOC	1.22	3.83	0.01
MA	1.28	5.32	0.01
MM	1.19	3.24	0.01
AT	1.15	2.34	0.00

5. CONCLUSION

This paper has described more robust approaches to spectrum estimation. From the simulations results detailed in this paper, the AT method can be seen to provide the most robust spectral analysis method. In addition, it achieves best quantization performance. Experiments show that all the proposed methods offer improvements in terms of robustness and quantization performance over the autocorrelation method of LP analysis.

6. REFERENCES

[1] Makhoul, J., “Linear Prediction: A Tutorial Review”, in *Proc. IEEE*, vol. 63(4), pp. 561-580, April 1975.

[2] Paliwal, K. K., “Robust Linear Prediction Analysis Methods and Their Application to Speech Coding and Recognition”, paper under preparation.

[3] Paul, D. B., “The Spectral Envelope Estimation Vocoder”, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29, pp. 786-794, 1981.

[4] Zhang, W., Holmes, W. H., “Performance and Optimization of the SEEVOC Algorithm”, in *Proc. 5th Int. Conf. Spoken Language*, vol. 2, pp. 523-526, 1998.

[5] Kleijn, W. B., Paliwal, K. K., *Speech Coding and Synthesis*, Elsevier, Amsterdam, 1995.

[6] Itakura, F., “Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals”, *J. Acoust. Soc. Am.*, Vol. 57, p. S35, 1975.