Robust Spectrum Analysis for Applications in Signal Processing

Nanda P. Koestoer and Kuldip K. Paliwal

Abstract— **Research into areas of signal processing has continued to flourish with numerous applications developed for specific needs. However, further research into the more fundamental parts of these fields is still required. The fundamental core of signal processing that is used in such applications as signal compression and speech recognition is the linear prediction theory. The conventional linear prediction method has performed well under normal conditions, but when introduced to noisy conditions still leaves much to be desired. This has fueled the need to develop a method that improves the performance of the linear prediction method for applications in noisy environments. This paper will introduce new methods in estimating the power spectra of signals to be used in linear prediction that is both accurate and robust. These methods, which are the Moving Average (MA), Moving Maximum (MM), and the Average Threshold (AT) method, will be compared to the more commonly used methods of spectral envelope estimation, such as the widely used Linear Predictive Coding (LPC) and Spectral Envelope Estimation Vocoder (SEEVOC) methods. This paper will discuss the advantages these proposed methods have over the conventional methods, and also show that it can be further optimized by applying different sized implementation windows to reach specific purposes. The effects of noise will also be researched to obtain optimum robustness for signal recognition purposes. The simulation results will show that the proposed methods will perform better under noisy environments.**

I. Introduction

LINEAR prediction theory and algorithms surrounding this matter have matured to the point where they are now an integral part of many real-world adaptive systems. The linear prediction model provides a parametric description of an observed process. As an example in speech data, a certain set of linear prediction information may represent considerably different sets of data. Hence, the methodology of these processes involves describing large records of data by a uniform set of parameters containing its significant process information.

Linear prediction models treat problems of adaptive linear systems and explain it to sets of measurements or observations. Such observations in this case can be used to analyze and estimate the power spectrum envelope (PSE) of a given data set. This analysis is widely used in data compression areas and more specifically in speech recognition and verification.

The purpose of PSE estimation is to establish the spectral component of a random process from a finite set of observation. The spectral estimation is done by modeling a signal so it can be seen as sinusoidal waves located at certain frequencies, performing its function as a preliminary data analysis tool. Further implementations involved with the spectral estimation process include a broad range of signal processing application such as speech recognition and signal compression.

The methods that are being proposed here intend to produce a spectrum estimation method to be used in linear prediction coding that is robust when introduced with noise. The theory behind these methods are covered in Sections II-A and II-B. The methodology behind the proposed PSE estimation methods are explained in Section II-C. A direct comparison with the conventional linear prediction (LP) coding method proposed in [1] and the Spectral Envelope Estimation Vocoder (SEEVOC) method proposed in [2] and [3] can be seen in Section III.

II. Spectrum Estimation

A. Background

The estimation of spectral envelopes offers a concise representation of important signal properties which largely simplifies the control of synthesis models. In processing a digital signal, the quality of a synthetic or modeled decoded signal relies heavily on how well its spectral envelope is estimated. The requirements needed to be reached in estimating a model are the spectral envelope properties be fulfilled, and that a certain level of robustness be reached. The estimation should be as precise and as smooth as possible for a variety of signals in ideal or real-world noisy environments.

The main obstacle in modeling an ideal spectral envelope is, it is impossible to estimate the spectral envelope from a finite discrete set of values without introducing some constraints on the envelope. Usually these constraints are entirely fixed by the parameterization. Improvements suggested to the linear prediction methods are mostly application dependent.

The weakness of the LPC spectral envelope is, it tends to wrap the spectrum as tightly as possible and then descend down into the space between the partial structures when they are spaced far apart (eg., high pitched sounds). This means that between peaks it tends to go down and not maintain a smooth envelope. Under certain conditions, it may descend down to the level of the residual noise between harmonic partial gaps.

The SEEVOC method is a spectral envelope estimation technique that has been proposed to improve the performance of the conventional LPC. It is chosen here because it ignores the low-level spectral peaks that may be a result of residual noise or side-lobe effects. However, although this method performs well on low frequency data (range 50-400 Hz), it has a disadvantage of not offering a significant improvement on the overall frequency plane. This method's spectral estimation accuracy also depends highly on *a priori* knowledge of the average signal pitch, which is a definite complication in real-world applications.

B. Application in LP

All the proposed methods in Section II-C would manipulate the magnitude spectrum of the periodogram (P_{xx}) of a given signal. The periodogram method of spectrum estimation is the *classic* way of estimating the power spectrum which is highly vulnerable to noise. Despite certain disadvantages, the periodogram magnitude spectrum still provides a very good model to be used in designing the proposed methods.

This method is based on a direct approach through a short-time Fourier transform on a frame of data. It can be defined as follows,

$$P_{xx}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{(-j2\pi fn)}\right|^2 = \frac{1}{N}|X(f)|^2 \quad (1)$$

where $|X(f)|^2$ represents the Fourier transform of a signal $x(n)$.

The resulting spectral estimate would then be used to optimize the calculation of the predictor-coefficients that would be used in a linear prediction based system.

The PSE would generally be generated according to the *peaks* of the periodogram magnitude spectrum. This approach has been very widely used due to the computational efficiency of the Fast Fourier Transform (FFT) algorithm. As explained by Kay and Marple [4], a statistically consistent spectrum using this method can only be reached by separating the data sequence into smoothed segments. Leakage effects that occur due to data windowing can be minimized through the selection of nonuniform weighted windows.

The autocorrelation coefficients (γ_{xx}) can then be generated from each of the spectrum envelope estimate, as defined below,

$$\gamma_{xx}(\tau) = \int_{-\infty}^{\infty} \Gamma_{xx}(f) e^{j2\pi f\tau} d\tau \quad (2)$$

where $\Gamma_{xx}(f)$ represents the distribution of the power spectrum as a function of frequency.

Further explanation regarding the implementation of the autocorrelation values in determining its prediction coefficients using the linear prediction method can be attained in [5] and [6].

C. *Proposed Methods*

C.1 Moving Average (MA) Method

This method uses a moving average filter to smooth the periodogram spectrum of the input signal. Using an M-size MA window, the averaging range can be defined as follows,

$$w(i) = \frac{N-|i|}{N^2} \quad (3)$$

where $-(N-1) \le i \le (N-1)$ and $N = \lfloor \frac{M}{2} \rfloor + 1$.

C.2 Moving Maximum (MM) Method

This method searches for a maximum level from the periodogram spectrum of the input signal. The maximum point will then be used to represent a certain interval surrounding that frequency point.

The mathematical implementation of the MM method is 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 MM window would be (2N+1).

C.3 Average Threshold (AT) Method

The methodology for the AT method is as follows:

- The MA algorithm is applied on the periodogram spectrum of the input signal.
- The resultant average spectrum is then combined with the original power spectrum. The maximum value between the two spectrum at any given frequency location would then be used as the new AT spectrum.
- The steps above are then repeated a certain number of times to achieve an optimum result.

III. Results

As an objective measure, its Spectral Distortion (SD) was calculated over the power spectrum on a frequency plane. It is defined as follows,

$$SD = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(10log_{10}\frac{P_i}{\hat{P}_i})^2} \quad (4)$$

where P_i and $\hat{P}_i$ are the true and estimated power spectrum's respectfully.

The following results were simulated using an 8 kHz vocal tract [e] length of 240 samples. The following coefficients were used: Coarse pitch for SEEVOC method = 10, with range settings [0.5 1.5]; AT repetitions set to 5; MA windowing range = 24; MM window = 12 samples; LPC order = 10. The predictor-coefficients were generated from the autocorrelation values using the *Levinson-Durbin* algorithm. Linear interpolation on the magnitude spectrum was used in the SEEVOC method. This provides both mathematical simplicity and no significant disadvantage compared to other methods of interpolation according to [2].

Increasing the AT repetitions would minimize the spectral distortion. On the other hand the PSE estimate may become less accurate as the spectrum variation gradients between the peaks and its *valleys* are smoothed through repetitious AT processes. The same effect is also apparent when increasing the window range for the MA and MM methods. It was found that by using the window settings set above, a generally accurate estimate for most simulations were obtained.

Figure 1 compares the PSE estimation methods on a noise-added signal to its PSE estimate on a clean signal. Figure 2 shows the SD when the PSE estimation methods are applied to a uniform signal synthesized from a power spectrum of a clean signal. The measurements show that the AT and MM methods are almost identical in providing the largest improvement to the conventional LPC method.

The signal compression application for the linear prediction methods is done using a low bit-rate data compression technique referred to as the Code Excited Linear Prediction (CELP) coding method.

This method is a very commonly reported coding method for usage of signals below 8 kHz, appropriate for speech compression. The encoder can be separated into three main operations: short-term prediction, long-term prediction, and excitation determination. For this particular coder, the excitation is determined from a codebook of white Gaussian vectors. Moreover, each Gaussian excitation is weighted accordingly in order to produce minimum error of synthesized signal. A detailed explanation of the CELP encoder can be attained in [7].

Synthesized results using the CELP coder were simulated using a closed-loop long-term prediction, with frame size limited to 20 ms. The excitation vector and its weighting were determined from a 6-bit Gaussian codebook. The speech data was generated from a male speaker counting random numbers for 6 seconds. The linear prediction methods were used in generating the autocorrelation functions, and were furthermore used in generating its predictor-coefficients. Each simulation was applied on a speech signal with added noise. The synthesized result was

Figure 1. Spectral Distortion for the PSE Estimation Methods

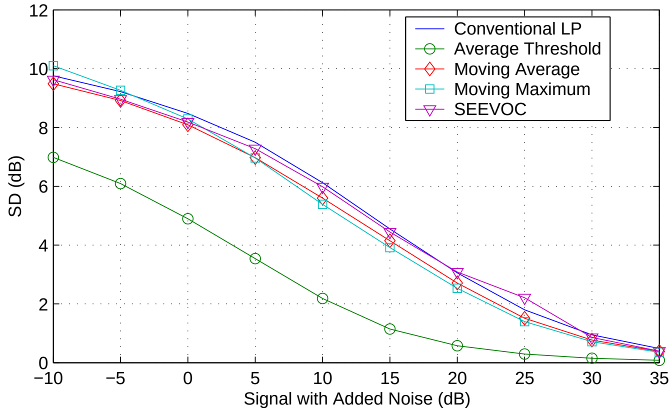
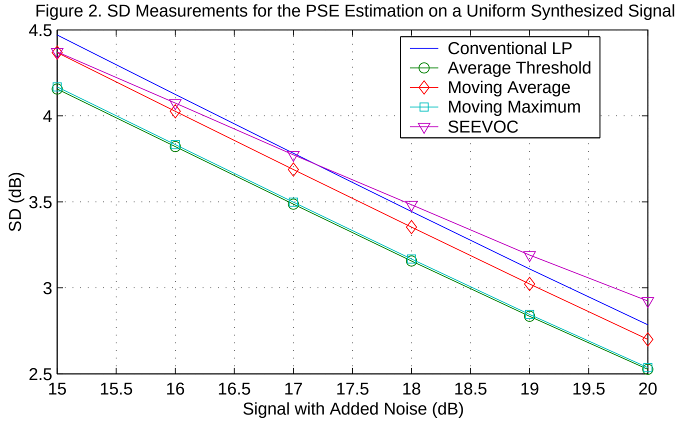


Figure 2. SD Measurements for the PSE Estimation on a Uniform Synthesized Signal



then compared to the original clean signal using the Signal-to-Noise Ratio (SNR) as a uniform measure. As can be seen in Figure 4, the AT method provides the most robust estimation in noisy environments.

IV. Conclusion

Various applications of signal processing have commonly used the conventional power spectrum estimation of linear prediction. However many areas are still open for further improvements, especially when a system is introduced into a noisy environment. The methods proposed in this paper will increase the robustness and accuracy of PSE estimation in these conditions, hence improving further applications of signal compression and speech recognition.

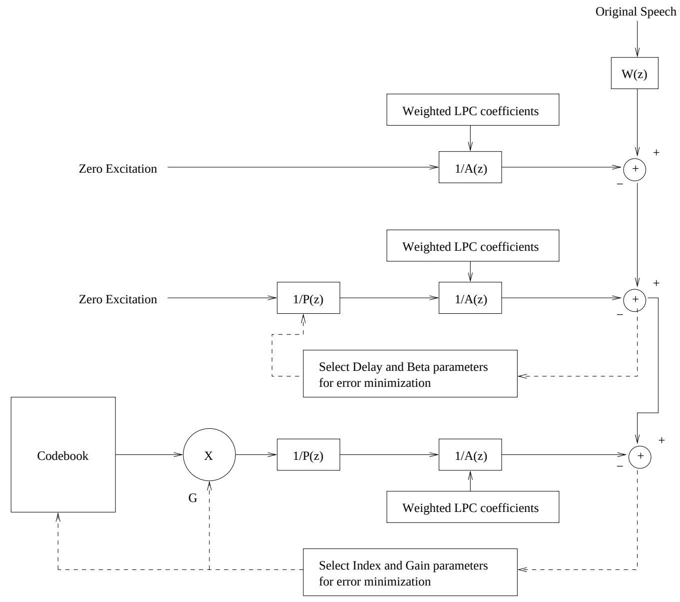
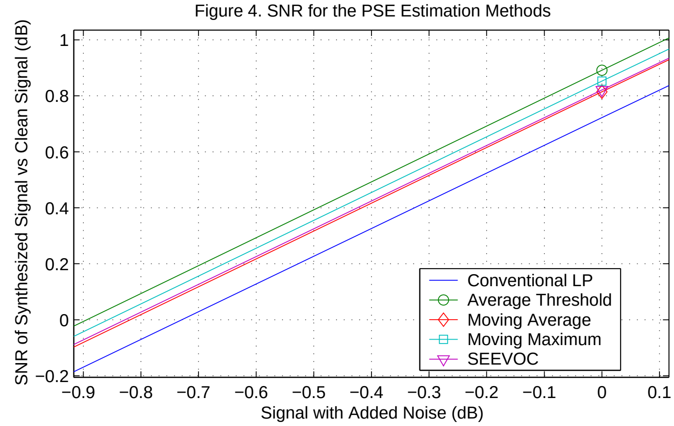


Figure 3. Block Diagram of the CELP Coder

Figure 4. SNR for the PSE Estimation Methods



References

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

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

[3] Zhang, W., Holmes, W. H., "Performance And Optimization Of The SEEVOC Algorithm", *Proc. 5th Int. Conf. Spoken Language*, vol. 2, pp. 523-526, 1998.

[4] Kay, S. M., Marple Jr., S. L., "Spectrum Analysis: A Modern Perspective", *Proc. IEEE*, vol. 69(11), pp. 1380-1419, November 1981.

[5] Strobach, P., *Linear Prediction Theory: A Mathematical Basis for Adaptive Systems*, Springer-Verlag, Berlin, 1990.

[6] Proakis, J.G., Manolakis, D.G., *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed, Prentice Hall, New Jersey, 1996.

[7] Schroeder, M., Atal, B., "Code Excited Linear Prediction: High Quality Speech at Very Low Bit Rates", *Proc. of ICASSP*, pp. 937-940, 1985.