CYCLIC AUTOCORRELATION-BASED LINEAR PREDICTION ANALYSIS OF SPEECH

K.K. Paliwal * *and Y. Sagisaka*

ATR Interpreting Telecommunications Res. Labs.
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

ABSTRACT

In this paper, a new approach for linear prediction (LP) analysis is proposed. This approach makes the assumption that the speech signal is cyclostationary and uses cyclic autocorrelation function for computing LP parameters. Since the cyclic autocorrelation function of a stationary random signal is zero, independent of its statistical description, this analysis is robust to additive noise, white or colored. It is applied to speech recognition. Preliminary results demonstrate its robustness to white additive noise.

1. INTRODUCTION

Linear prediction (LP) analysis is widely used in various speech processing applications for representing the short-time spectral envelope information of speech. This analysis assumes the speech signal to follow an autoregressive (AR) model. It performs reasonably well for clean speech signals. But, when these signals are corrupted by the addition of random noise, the AR model is no more valid and, as a consequence, its performance is poor for noisy signals [1].

In this paper, we propose a new approach for LP analysis. This approach makes the assumption that the speech signal is cyclostationary and uses cyclic autocorrelation function for computing LP parameters. Since the cyclic autocorrelation function of a stationary random signal is zero independent of its statistical description [2], this analysis is robust to additive noise, white or colored.

This cyclostationarity property has been exploited in the past in many communication applications [3, 4]. In this paper, we use it in conjunction with LP analysis to get an estimate of LP power spectrum which is robust to additive noise.

2. SOME DEFINITIONS

Autocorrelation function of a stationary random process $u(t)$ at time lag τ is defined as follows:

$$R_u(\tau) = E\{u(t)u(t+\tau)\}, \tag{1}$$

where $E\{.\}$ denotes expectation. Power spectrum of the random process $u(t)$ can be computed from its autocorrelation function, using the Wiener-Khinchin theorem, as

*On leave from School of Microelectronic Engineering, Griffith University, Brisbane, QLD 4111, Australia.

follows:

$$P_u(f) = E\{R_u(\tau)exp(-j2\pi f\tau)\}. \tag{2}$$

Consider another process $v(t)$ which is cyclostationary with time period T_0, or fundamental frequency $F_0 \ (= 1/T_0)$. We can define cyclic autocorrelation function of this process as follows [2]:

$$R_v(\tau, f) = E\{v(t)v(t+\tau)exp(-j2\pi ft)\}. \tag{3}$$

The cyclic autocorrelation function of a cyclostationary random process satisfies the following property:

$$R_v(\tau, f) \quad = \quad \begin{cases} \text{finite}, & \text{if } f = nF_0, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where n is a non-zero integer. The cyclic autocorrelation function of the stationary random process $u(t)$ which is not cyclostationary is zero for all values of f, except $f = 0$.

Note from Eqs. (2) and (3) that the cyclic autocorrelation function reduces to the conventional autocorrelation function when $f = 0$.

3. CYCLIC LP ANALYSIS METHOD

Consider an observed signal $x(t)$ obtained by corrupting the clean speech signal $s(t)$ by an additive noise signal $w(t)$; i.e.,

$$x(t) = s(t) + w(t). \tag{5}$$

Here, we are assuming the clean signal $s(t)$ to be cyclostationary. The noise signal $w(t)$ is assumed here to be stationary with any statistical distribution, white or colored.

Let $R_x(\tau, f)$, $R_s(\tau, f)$ and $R_w(\tau, f)$ be the cyclic autocorrelation functions of $x(t)$, $s(t)$ and $w(t)$, respectively. Eq. (5) in cyclic autocorrelation domain can be written as

$$R_x(\tau, f) = R_s(\tau, f) + R_w(\tau, f), \tag{6}$$

Since $w(t)$ is not cyclostationary, it means that $R_w(\tau, f) = 0$, for $f \neq 0$, and Eq. (6) becomes

$$R_x(\tau, f) = R_s(\tau, f) \qquad \text{for } f \neq 0. \tag{7}$$

This means that, independent of noise statistics, the cyclic autocorrelation function is insensitive to noise as long as the noise is not periodic.
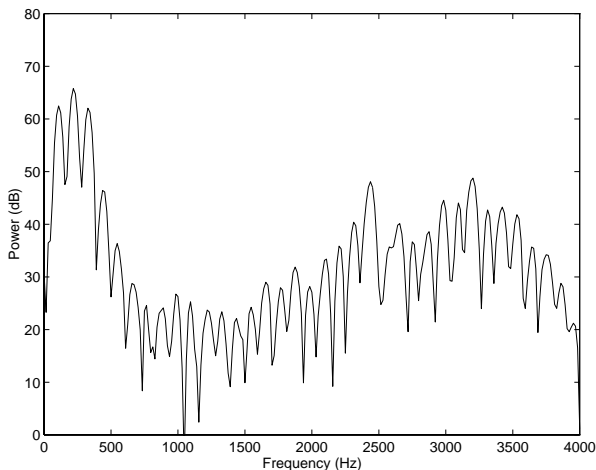
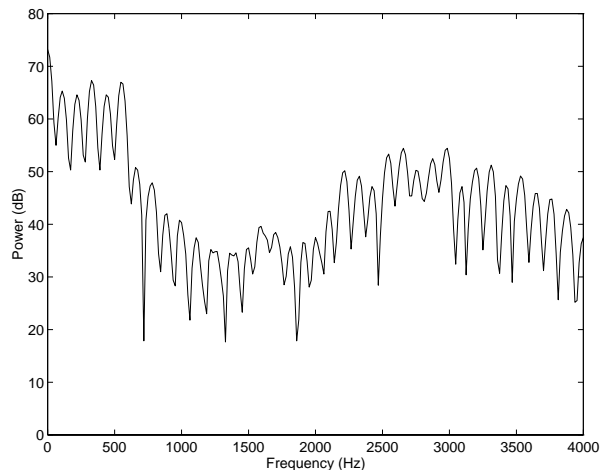Figure 1. Conventional power spectrum of vowel /i/.

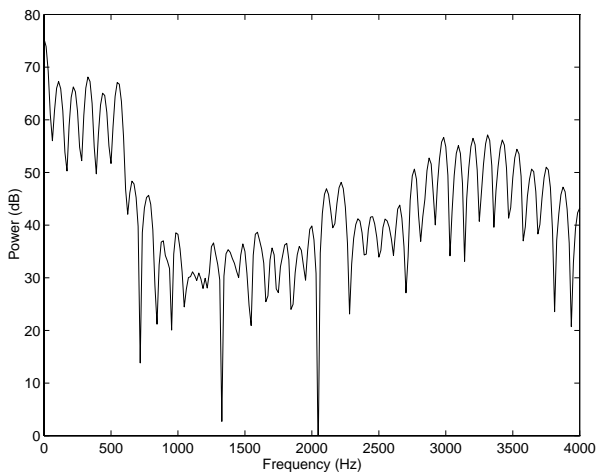Figure 2. Cyclic power spectrum of vowel /i/ for $m = 2$.

Figure 3. Cyclic power spectrum of vowel /i/ for $m = 4$.

Figure 4. Cyclic power spectrum of vowel /i/ for $m = 6$.

In practice, only N samples $\{x(n), n = 0, 1, ..., N - 1\}$ of the observed signal are available for analysis. The cyclic autocorrelation function can be computed from these N samples as follows:

$$R_x(m, f) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n)x(n+m) \exp(-j2\pi f nT), \quad (8)$$

where T is the sampling period. The signal $x(n)$ can be weighted by a tapered window function (such as the Hamming window function) prior to its use in Eq. (8).

In order to use the above-mentioned robustness property for developing an LP analysis, let us define the cyclic power spectrum,

$$P_x(m, f) = |R_x(m, f)|^2. \quad (9)$$

Since the noise corrupts the cyclic power spectrum at $f = 0$ (see Eq. (7)), we select a portion of the cyclic power spectrum which does not include a small region near $f = 0$
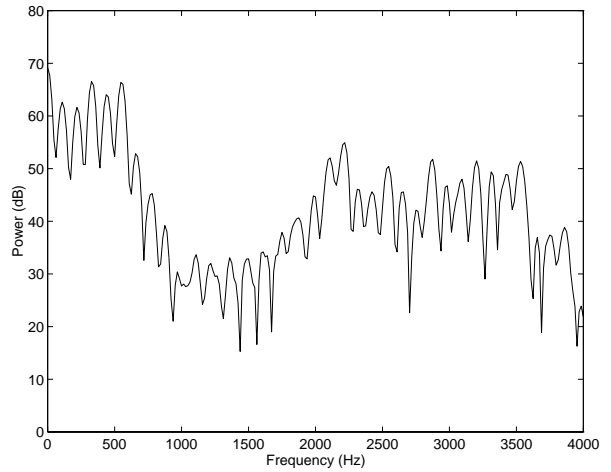
and perform a p-th order selective linear prediction analysis [5]. This results in p LP coefficients which are immune to noise.

From Eqs. (4) and (9), we can see that the cyclic power spectrum is finite at harmonic locations, i.e., at $f = nF_0$. For $f \neq nF_0$, it is zero. That is, regions in the cyclic power spectrum between harmonic peaks do not contain any meaningful information. Thus, it will be better if these regions are not used at all in the analysis. This will improve the robustness of the LP analysis method further. This can be done by using cyclic power spectrum only at harmonic frequencies for computing the LP parameters through a discrete all-pole modeling method [6].

Note that different values of m can result in different sets of LP coefficients.

4. SPECTRAL ANALYSIS RESULTS

In order to illustrate our results, we take a segment of 240 samples of vowel /i/ (sampling frequency = 8000 Hz). Fig. 1 shows the conventional power spectrum of this segment.

Figures 2, 3 and 4 show the cyclic power spectra of this segment for $m=2$, 4 and 6, respectively. Each of these cyclic power spectra has a harmonic structure with the same pitch frequency as seen in the conventional power spectrum. However, the shape of the spectral envelope undergoes a change in the cyclic power spectrum. For speech recognition applications, this change in spectral shape is acceptable as long as it is consistent and preserves separation between linguistic classes.
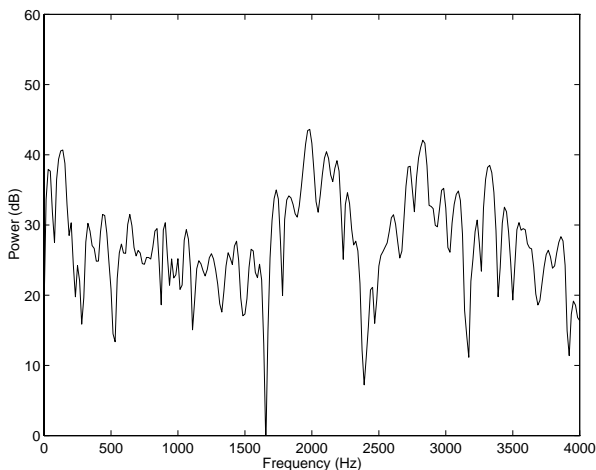

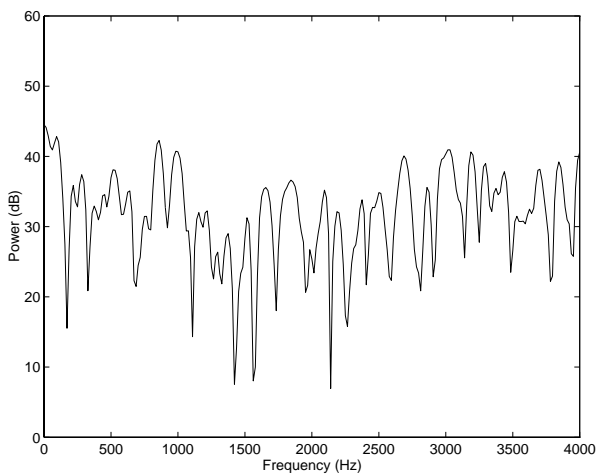**Figure 5. Conventional power spectrum of fricative /s/.**

Figure 6. Cyclic power spectrum of fricative /s/ for $m = 2$.

Fig. 5 shows the conventional power spectrum of a 240-sample long segment of an unvoiced sound /s/. Figures 6, 7 and 8 show the cyclic power spectra of this segment for $m=2$, 4 and 6, respectively. Since the unvoiced speech signals are not cyclostationary, the spectral envelope information is totally lost in the cyclic power spectrum; i.e., the cyclic power spectrum is approximately flat for $f \neq 0$.

Fig. 9 shows a 150–4000 Hz portion of the cyclic power spectrum of vowel /i/ and the resulting spectral envelope computed through 10-th order selective LP analysis.
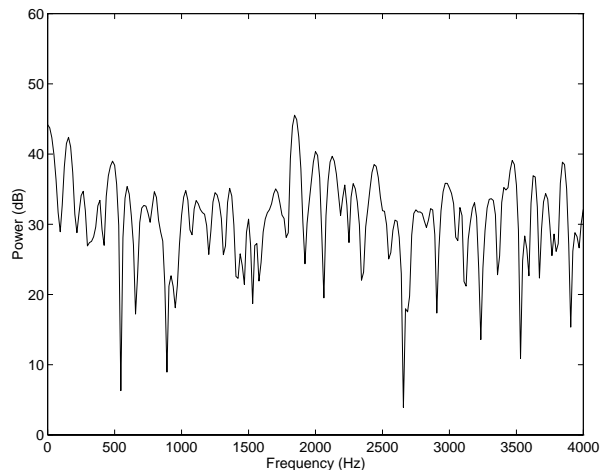

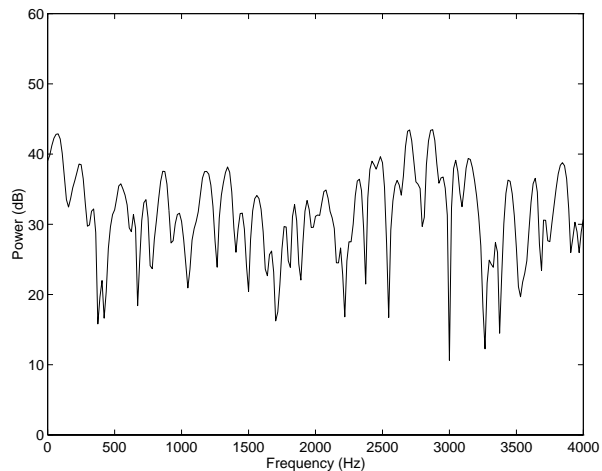**Figure 7. Cyclic power spectrum of fricative /s/ for $m = 4$.**

Figure 8. Cyclic power spectrum of fricative /s/ for $m = 6$.

5. SPEECH RECOGNITION RESULTS

We have used the cyclic LP analysis method for speech recognition at different signal-to-noise ratios (SNRs) and compared its performance with that of the conventional LP analysis method. We have used a very simple speech recognition task; namely, to classify steady-state vowel segments into 10 vowel classes. The speech data base used for this purpose is derived from 300 utterances which consist of 30 repetitions of 10 different /b/-vowel-/b/ syllables spoken by a single male speaker. These utterances are lowpass filtered to 4 kHz and digitized at 10 kHz sampling rate. The steady-state part of the vowel segment is manually located for each of the 300 utterances and a 20 ms segment excised from its center. A 10-th order LP analysis is performed for each such 20 ms segment. The first 15 repetitions are used for training and the remaining 15 repetitions are used for testing.

The maximum likelihood (ML) classifier is used here for vowel classification. Ten cepstral coefficients derived through LP analysis are used as recognition features. Pre-
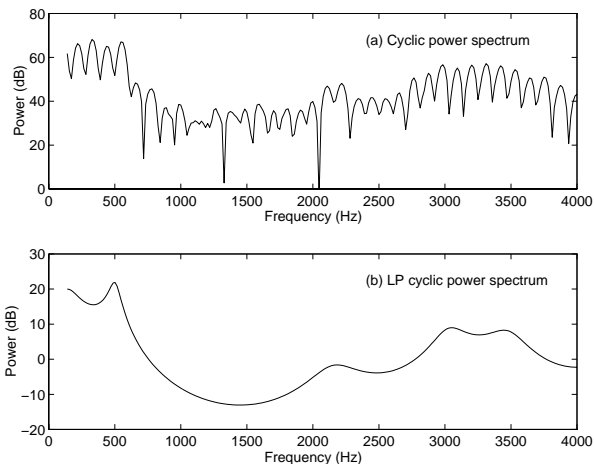
Figure 9. (a) Cyclic power spectrum and (b) LP spectral envelope for vowel /a/.

liminary results are listed in Table 1. We can observe from this table that cyclic LP analysis results in better speech recognition performance than the conventional LP analysis at lower SNRs.

Table 1. Speech recognition performance of the conventional and the cyclic LP analysis methods in presence of additive noise distortion.

SNR (dB)	Recognition accuracy (in %) with	
	LP method	Cyclic LP method
∞	94.7	91.3
30	87.3	87.3
25	80.0	80.6
20	56.7	76.7
15	42.0	66.7

It may be noted that cyclic LP analysis provides meaningful information for voiced speech sounds which are periodic in nature. For unvoiced speech, this analysis provides no meaningful information. Therefore, for a general speech recognition problem, where we have to deal with both voiced and unvoiced speech sounds, this analysis alone is not sufficient. For this, we must use cepstral coefficients not only from the cyclic LP analysis but from the conventional LP analysis as well.

6. CONCLUSIONS

In this paper, a new approach for linear prediction (LP) analysis has been proposed. This approach makes the assumption that the speech signal is cyclostationary and uses cyclic autocorrelation function for computing LP parameters. Since the cyclic autocorrelation function of a stationary random signal is zero, independent of its statistical description, this analysis is robust to additive noise, white or colored. It has been applied to a simple speech recognition task. Preliminary results demonstrate its robustness to white additive noise.

REFERENCES

[1] S.M. Kay, "The effects of noise on the autoregressive spectral estimator", IEEE Trans. Acoust. Speech and Signal Processing, Vol. ASSP-27, pp. 478-485, 1979.

[2] W.A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals", IEEE Signal Processing Magazine, Vol. 8, pp. 14-36, Apr. 1991.

[3] W.A. Gardner, "Signal detection: A unifying theoretical framework for geature detection", IEEE Trans. Commun., Vol. 36, pp. 897-906, Aug. 1988.

[4] G.K. Yeung and W.A. Gardner, "Search-efficient methods of detection of cyclostationary signals", IEEE Trans. Signal Processing, Vol. 44, No. 5, pp. 1214-1223, May 1996.

[5] J. Makhoul, "Linear prediction: A tutorial review", Proc. IEEE, Vol. 63, pp. 561-580, 1975.

[6] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling", IEEE Trans. Signal Processing, Vol. 39, No. 2, pp. 411-423, Feb. 1991.