PERFORMANCE OF LINEAR PREDICTION ANALYSIS ON SPEECH WITH ADDITIVE NOISE

B. Yegnanarayana and T.K. Raja

Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore-560012, India

ABSTRACT

We shall present in this paper the results of investigations on the performance of Linear Prediction (LP) analysis of speech when the signal is corrupted with additive white noise. In particular we shall discuss the effect of noise on various parameters derived from LP analysis. Theoretical justification for the observed results is given.

I. INTRODUCTION

Noise is the single most important factor that ultimately determines the performance efficiency of a speech signal processing system [1]. Speech produced under normal environment will have different signal to noise ratios (S/N) in different segments depending upon the level of signal in each segment as for example voiced segments will have higher S/N compared to unvoiced segments. Linear prediction analysis may be used for transmission of speech information even when speech is corrupted with noise because temporal variations of some spectral features are preserved and the high redundancy of language supplies the missing information. But for applications such as automatic speech and speaker recognition and speech aids to handicapped, the features extracted by LP analysis depend critically on the quality of speech signal. We shall discuss in this paper the effect of S/N of analysis segment on parameters obtained through LP analysis.

II. LINEAR PREDICTION ANALYSIS

LP analysis is performed on speech segments corrupted with additive noise. The noise signal is obtained by adding the samples of the output of a random noise generator to speech samples. The sampling is done at 10 KHz. LP coefficients obtained by solving the auto-correlation normal equations are used to derive the inverse filter response of the vocal tract. Inverse filter responses for a single resonator output and for voiced speech segments are shown in Figs. 1 to 3. The responses are plotted with S/N as a parameter. Figs. 4 and 5 show variation of LPCs with S/N for the two vowel segments.

III. DISCUSSION

It is interesting to observe that the inverse filter response predicts the resonance frequencies reasonably well even for low S/N. But the bandwidths of the resonance peaks are progressively increased as S/N is reduced. The explanation for such a behaviour can be found in spectral approximation in LP analysis [2]. The model spectrum approximates the envelope of signal spectrum with higher accuracy of approximation at peaks of the envelope. Addition of white noise merely adds noise spectrum to signal spectrum thus reducing the dynamic range of the overall spectrum without affecting the resonance peaks of the signal spectrum. For a given dynamic range the minimum resonance bandwidth is fixed as can be seen from the expression for the frequency response of a single resonator:

$$|H(jw)|^2 = \frac{1}{(w^2-w_0^2)^2 + 4\alpha^2 w^2}$$

where w_0 is the resonance frequency and α is the half power bandwidth. At resonance $w = w_0$,

$$|H(jw_0)| = \frac{1}{4\alpha^2 w_0^2} .$$

At large frequencies $w \gg w_0$, and

$$|H(jw)|^2 - \frac{1}{w^4} .$$

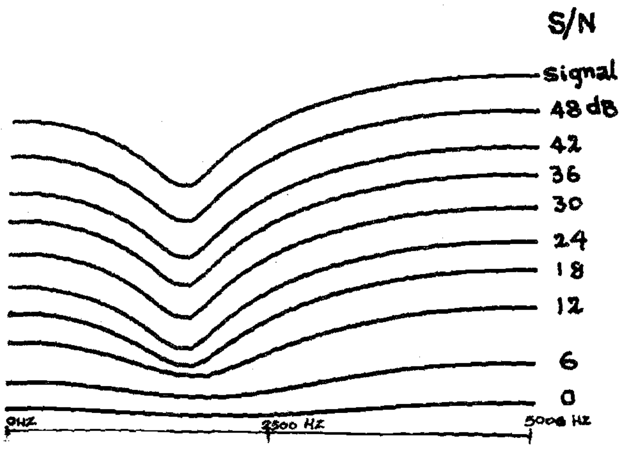

Fig.1: Inverse filter responses for a single resonator model (Resonance frequency = 1500 Hz and bandwidth=150 Hz)

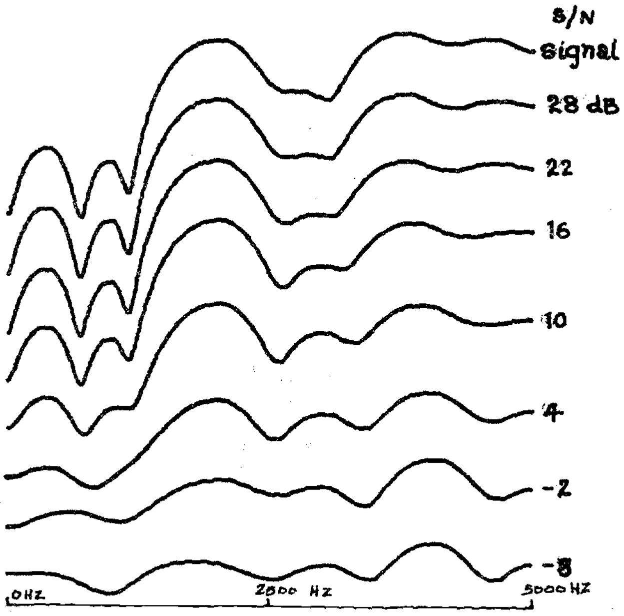

Fig. 2: Inverse filter responses for the vowel segment |a|.

Let w_1 be the highest frequency of interest. Then the dynamic range d is given by

$$d = \left(\frac{1}{4\alpha^2 w_0^2}\right) \Bigg/ \left(\frac{1}{w_1^4}\right) = \frac{w_1^4}{4\alpha^2 w_0^2}$$

Since w_0 is estimated correctly by LP analysis and w_1 is fixed, the dynamic range is inversely proportional to square of bandwidth. Therefore if d is reduced bandwidth is automatically increased as can be seen in Fig. 1.

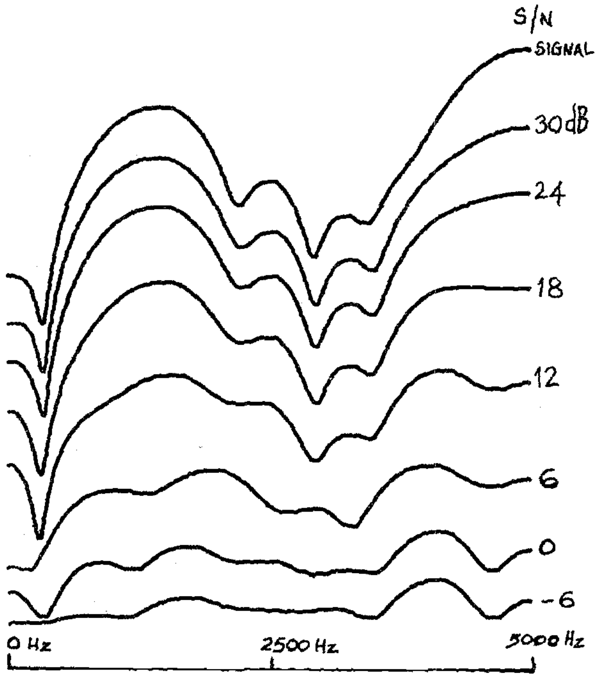

Fig. 3: Inverse filter response for the vowel segment |i|.

For speech signal higher formants are not generally affected as much as the first formant when S/N is reduced. This is because the first formant has much lower bandwidth and it requires large dynamic range for accurate estimation. In fact better estimates of higher formants are obtained by reducing the dynamic range which is usually realized by providing high frequency emphasis at preprocessing stage. In Fig.2 higher formants are clearly resolved when S/N is lower to about 10 dB.

Although formant frequencies are estimated correctly even under poor S/N conditions it is difficult to identify them from LP magnitude spectrum by peak picking methods especially when the peaks are close to each other or when the bandwidths are large. Derivative of LP phase spectrum can be effectively used for this purpose [3] as shown in Fig.6 for a two resonator model.

The above discussion indicates that LP vocoders should perform satisfactorily even for noisy speech as the temporal variations of formant frequencies are preserved. However when LPCs or related parameters such as reflection coefficients and area functions are used in speech recognition or speaker verification systems, additive noise increases the

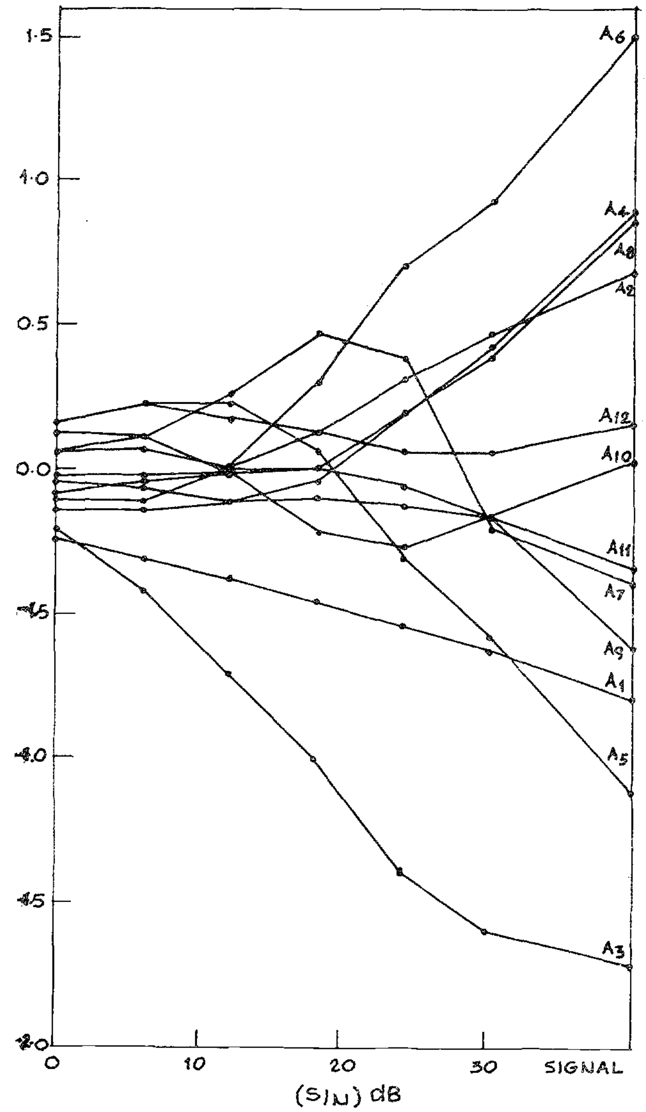

Fig. 4: LPCs for the vowel segment |i|

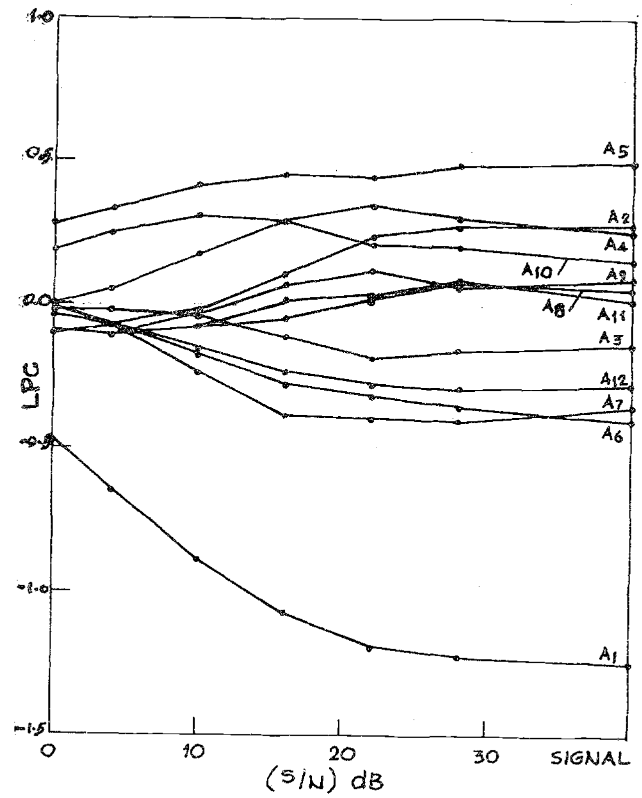

Fig. 5: LPCs for the vowel segment |a|.

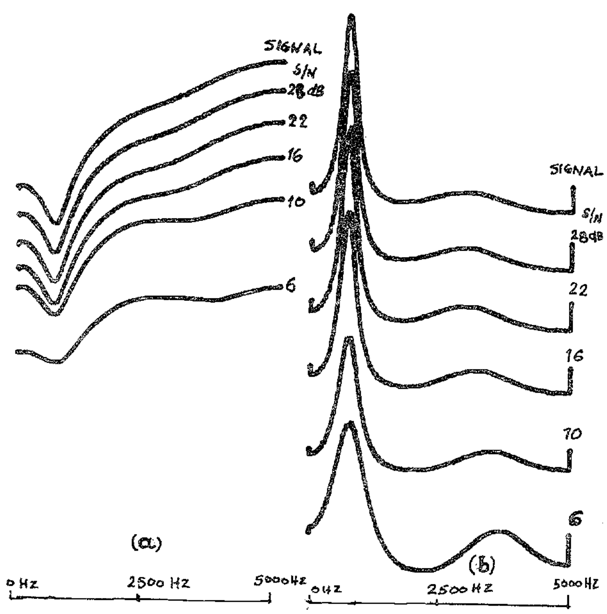

Fig.6: Formant extraction using LP phase spectrum

(a) Frequency response
(b) Differentiated phase spectrum

intra-variation significantly making it difficult to design classifiers for pattern recognition. Typical variations of LPCs with noise for two voiced segments are shown in Figs. 4 and 5. The changes in LPCs even at 30 dB S/N is significant and the ordering the coefficients is disturbed as noise level increases. One of the important applications of LP analysis is to obtain nonuniform acoustic tube model for vocal tract in terms of area functions. When such models are used as aids to handicapped the noise level becomes a critical factor.

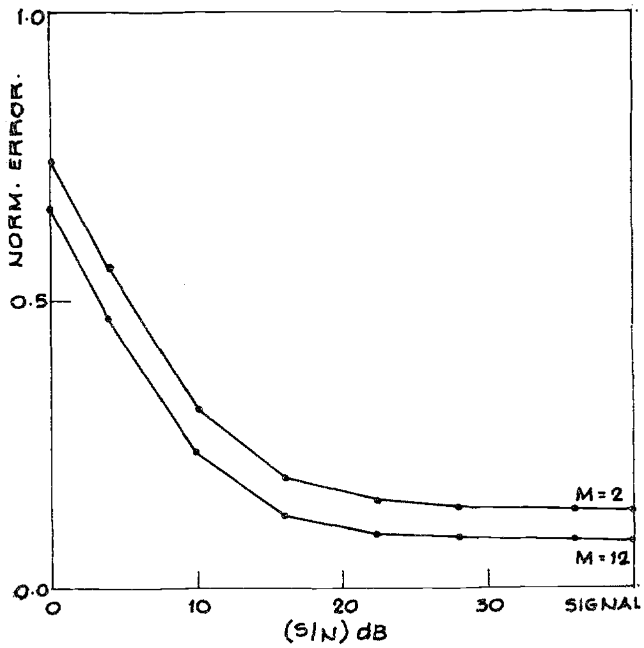

Fig.7: Normalized error for the vowel segment |a|

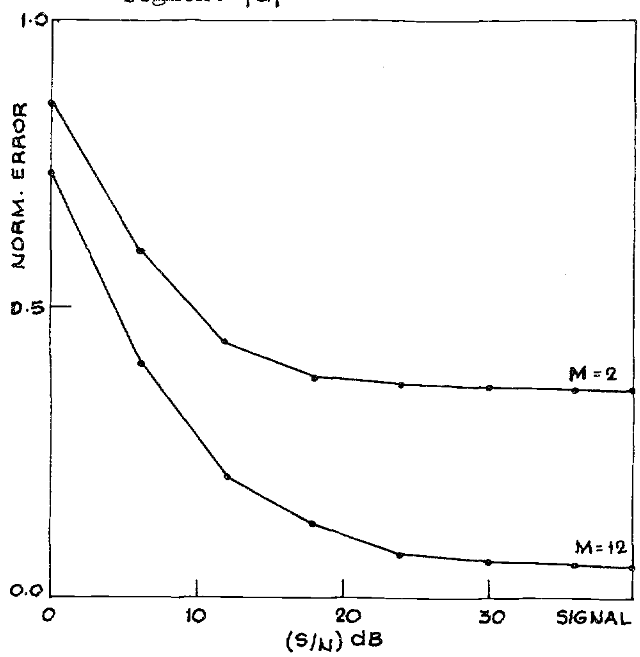

Fig. 8: Normalized error for the vowel segment |i|

Another useful parameter is the normalized error which is plotted for different S/N in Figs. 7 and 8 for two vowel segments. Although the general behaviour depicted is to be expected, it is interesting to note that the error increases steeply below about 20 dB S/N. The figures also show that normalized error increases as the dynamic range of the spectrum decreases.

IV. CONCLUSIONS

Although speech is intelligible even at low S/N , parameters extracted from speech vary significantly with S/N. We have shown that for vowel segments S/N below 20 dB reduces dynamic range of spectrum and increases error in estimating parameters. For low level speech segments such as unvoiced sounds the S/N will be very low making parametric extraction by LP analysis extremely difficult.

ACKNOWLEDGMENTS

The authors wish to thank Prof. B.S. Ramakrishna, Dr. V.V.S. Sarma and Mr. T.V. Ananthapadmanabha for many useful discussions.

REFERENCES

1. D.R. Reddy, 'Speech Recognition by Machine:A Review', Proc. IEEE, Vol.64, No.6, p.501, April 1976.

2. J. Makhoul, 'Linear Prediction: A Tutorial Review', Proc. IEEE, Vol.63, No. 4, p. 561, April 1975.

3. B. Yegnanarayana, 'Formant Extraction from Linear Prediction Phase Spectra', (submitted).