Signal Processing 4 (1982) 59–63
North-Holland Publishing Company

SHORT COMMUNICATION

ON THE PERFORMANCE OF BURG'S METHOD OF MAXIMUM ENTROPY SPECTRAL ANALYSIS WHEN APPLIED TO VOICED SPEECH

K.K. PALIWAL and P.V.S. RAO

Speech and Digital Systems Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Received 10 November 1980
Revised 5 March 1981

Abstract. Burg's method of maximum entropy spectral analysis is used to analyse voiced speech signal and its performance is compared with that of the autocorrelation and covariance methods of linear prediction using the following three criteria: (1) normalized total-squared linear prediction error, (2) error in estimating the power spectrum and (3) errors in estimating the first three formant frequencies and bandwidths. Results of pitch-synchronous and pitch-asynchronous analyses when applied to synthetic vowel signals are discussed.

Zusammenfassung. Das Verfahren von Burg zur Spektralanalyse mit maximaler Entropie wird auf die Analyse stimmhafter Sprachsignale angewendet. In einer vergleichenden Untersuchung dieses Verfahrens sowie der linearen Prädiktion nach der Autokorrelations- bzw. der Kovarianzmethode werden die folgenden drei Kriterien verwendet: (1) der relative quadratische Prädiktionsfehler, (2) der relative Fehler bei der Bestimmung des Leistungsspektrums, sowie (3) der relative Fehler bei der Bestimmung von Frequenz und Bandbreite der ersten drei Formanten. Die Analyse erfolgt grundperiodensynchron und -asynchron für synthetische Vokale.

Résumé. La méthode de Burg, d'analyse spectrale par le maximum d'entropie, est utilisée pour analyser des signaux de parole voisée et ses performances sont comparées avec celles des méthodes d'auto-corrélation et de covariance de la prédiction linéaire suivant les 3 critères: (1) énergie de l'erreur de prédiction normalisée, (2) erreur d'estimation du spectre, (3) erreur d'estimation des 3 premiers formants et de leur largeur de bande. Les résultats de l'analyse synchrone ou non avec la période fondamentale sont discutés dans le cas de voyelles synthétiques.

Keywords. Voiced speech, maximum entropy spectral analysis, Burg's method, linear prediction, autocorrelation method, covariance method, formants.

1. Introduction

The maximum entropy method of spectral analysis is known for making high resolution spectral estimates especially for short analysis-segments and has been found to be very useful for processing geophysical and sonar signals [1–3]. Van den Bos [4] has shown that maximum entropy spectral analysis is equivalent to least-squares fitting of an all-pole model to the available data. Since voiced speech is usually modeled as the output of an all-pole filter, the maximum entropy method of spectral analysis can be used to analyse the voiced speech signal. Burg [5] has given an efficient method for estimating the parameters of the all-pole model. (For details about this method, see references [2] and [6].)

In the present paper, we shall use Burg's method of maximum entropy spectral analysis to analyse voiced speech and compare its performance with that of the autocorrelation and covariance methods of linear prediction. (For detailed treat-

0165-1684/82/0000–0000/$02.75

ment of the autocorrelation and covariance methods of linear prediction, see references [7] and [8].)

2. Criteria for performance evaluation

For evaluating the performance of the analysis methods, the following three criteria are used:

(1) normalized total-squared linear prediction (LP) error,

(2) error in estimating the power spectrum, and

(3) errors in estimating the first three formant frequencies and bandwidths.

These three criteria indicate the effectiveness of the analysis methods in representing the speech signal in time and frequency domains and thus provide useful guidelines about the suitability of these methods in various speech processing applications such as speech analysis-synthesis, automatic speech recognition and speaker recognition.

The normalized total-squared LP error (E_p) is defined here as the total output energy of the linear-prediction-error filter divided by the total input energy and is given by

$$E_p = \frac{\sum_{n=M+1}^{N} \left\{ x(n) + \sum_{k=1}^{M} \hat{a}_k x(n-k) \right\}^2}{\sum_{n=M+1}^{N} \{x(n)\}^2}$$

where $\hat{a}_k$'s are the coefficients of the estimated Mth order linear-prediction-error filter ($\hat{A}(z) = 1 + \sum_{k=1}^{M} \hat{a}_k z^{-k}$) and $x(n)$ is the nth sample of the analysis segment of duration N.

The error (E_s) in estimating the power spectrum is defined as

$$E_s = 2T \int_0^{1/2T} [|\hat{P}(f) - P(f)|/P(f)] \, df$$

where T is the sampling period and $\hat{P}(f)$ and $P(f)$ are the estimated and true power spectra, respectively.

The error (F_{ie}) in estimating the ith formant frequency is defined as

$$F_{ie} = \hat{F}_i - F_i$$

where $\hat{F}_i$ and F_i are the estimated and true frequencies, respectively, of the ith formant. Similarly, the error (BW_{ie}) in estimating the bandwidth of the ith formant is defined as

$$\mathrm{BW}_{ie} = \hat{\mathrm{B}}\mathrm{W}_i - \mathrm{BW}_i$$

where $\hat{\mathrm{B}}\mathrm{W}_i$ and BW_i are the estimated and true bandwidths, respectively, of the ith formant.

3. Results

The performance of Burg's method is compared with that of the autocorrelation and covariance methods for both pitch-synchronous and pitch-asynchronous analyses and synthetic vowel signals are used for analysis. The reason for using synthetic signals is that the true values of spectral parameters are known *a priori* for such signals. This helps in making a more objective performance evaluation of the analysis methods. The synthetic vowel signal is generated here by driving an all-pole filter

$$H(z) = G \Big/ \left(1 + \sum_{k=1}^{M} a_k z^{-k} \right)$$

by a periodic impulse train. Here, the gain G and filter coefficients $\{a_k\}$ are known *a priori*. The order of the linear predictor used in the analysis of the signal is taken to be the same as that of the all-pole filter used in the synthesis of the signal. When the autocorrelation method of linear prediction is used for analysis, the speech samples are weighted by a Hamming window function [9]; but no such windowing is carried out for the covariance and Burg's methods.

In order to illustrate the results, we take, as an example, a synthetic signal of vowel /o/ generated at 10 kHz sampling rate with 8 msec pitch period.

A 10-pole filter is used here for synthesis.[1] The true values of the various parameters are:

$G = 0.1354$, $a_1 = -1.53527$,

$a_2 = 0.97789$, $a_3 = -1.48396$,

$a_4 = 1.78023$, $a_5 = -0.71704$,

$a_6 = 0.73514$, $a_7 = -0.76348$,

$a_8 = -0.12135$, $a_9 = 0.15552$,

$a_{10} = 0.178143$, $F_1 = 403$ Hz,

$F_2 = 834$ Hz, $F_3 = 2645$ Hz,

$BW_1 = 53$ Hz, $BW_2 = 97$ Hz,

$BW_3 = 188$ Hz.

This synthetic signal is analysed in a pitch-synchronous manner with predictor order M equal to 10. The speech sample with the maximum value in a pitch period is taken as the first sample of the analysis segment and the duration of the analysis segment is taken to be 0.8 times the pitch period (i.e., $N = 64$ samples). The errors E_p, E_s, F_{1e}, F_{2e}, F_{3e}, BW_{1e}, BW_{2e} and BW_{3e} are computed and listed in Table 1. It can be seen from this table that the values of these errors for Burg's method are more than those for the autocorrelation and covariance methods, thus indicating the inferiority of this method in comparison to the other two methods. Also, the formant bandwidths are seriously underestimated by Burg's method. This underestimation of bandwidth is more for narrower bandwidths (72.3% for the first formant bandwidth, 65.5% for the second formant bandwidth and 42.9% for the third formant bandwidth). These observations are also illustrated in Fig. 1 where the true spectrum of the synthetic vowel /o/ and its pitch-synchronous estimates are shown. The difference between the spectrum estimated using Burg's method and the true spectrum can be obviously seen here.

For pitch-asynchronous analysis of the synthetic vowel /o/, the duration of the analysis-segment is taken to be 2.5 times the pitch period (i.e., $N =$ 200 samples). Four segments of the signal beginning at $n = 0$, 20, 40 and 60 (where $n = 0$ is the location of the maximum of the speech signal within a pitch period) are chosen for analysis. The errors E_p, E_s, F_{1e}, F_{2e}, F_{3e}, BW_{1e}, BW_{2e} and BW_{3e} are computed (with $M = 10$) for each of these segments. The average values of these errors are shown in Table 1. Using these errors as the criteria, it can be seen from this table that the performance

[1] In order to make the covariance matrix non-singular (which is necessary when the covariance method of linear prediction is used for analysis), some small amount of noise (uniformly distributed between ±0.005) is added to the driving process.

Table 1

Normalized total-squared LP error (E_p), error in estimationg the power spectrum (E_s) and errors in estimating the first three formant frequencies (F_{1e}, F_{2e}, F_{3e}) and bandwidths (BW_{1e}, BW_{2e}, BW_{3e}) for the synthetic vowel /o/

Error	Pitch-synchronous analysis			Pitch-asynchronous analysis		
	Burg's method	Auto. method	Cova. method	Burg's method	Auto. method	Cova. method
E_p	0.00312	0.00149	0.00002	0.02146	0.02146	0.02133
E_s	0.81334	0.62440	0.31946	0.24354	0.25145	0.14764
F_{1e} (in Hz)	51.9	−6.2	−0.6	−12.4	−14.1	−15.2
F_{2e} (in Hz)	−8.2	0.2	−0.8	8.6	11.9	12.2
F_{3e} (in Hz)	−104.0	17.1	2.6	−4.0	0.6	−0.4
BW_{1e} (in Hz)	−38.3	−7.0	1.9	7.1	9.3	3.8
BW_{2e} (in Hz)	−63.5	−1.8	−0.9	11.6	15.0	7.2
BW_{3e} (in Hz)	−80.6	13.7	4.2	8.2	5.6	−1.6

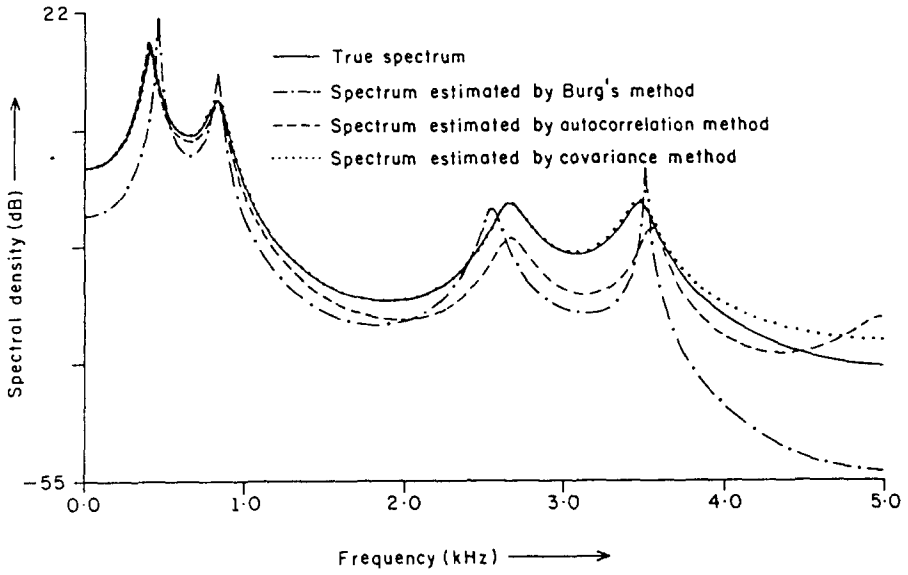

Fig. 1. True spectrum of the synthetic vowel /o/ and its pitch-synchronous estimates.

of Burg's method is comparable to that of the autocorrelation and covariance methods of linear prediction.

A number of other synthetic vowels and voiced consonants (/j/, /r/, /l/ and /w/) are also analysed and similar results are obtained for both pitch-synchronous and pitch-asynchronous analyses. Burg's method is also tried out on a number of segments of vowels and voiced consonants from real speech. For these segments, the instant of glottal closure (which is taken here to be the location of the maximum of the speech signal within a pitch period) is used for synchronization. The performance of Burg's method for real speech is found to be similar to its performance for synthetic speech. However, no objective evaluation of the method could be done here for real speech because of the nonavailability of true values of the spectral parameters.

4. Discussion and conclusion

It has been shown in the preceding section that the performance of Burg's method is inferior to that of the autocorrelation and covariance methods for pitch-synchronous analysis and comparable to that of the other two methods for pitch-asynthronous analysis. This can be explained as follows. In Burg's method, the same prediction error filter ($A(z) = 1 + \sum_{k=1}^{M} a_k z^{-k}$) is run over the analysis segment in both the forward and backward directions and the average of the total-squared forward and backward prediction errors is minimized with respect to a_M [2, 6]. This can be done only when the forward prediction error filter (which is obtained by minimizing the total-squared forward prediction error) works equally well when reversed and run backward in time [5]. This means that Burg's method can be expected to perform well only when the backward and forward output powers of the forward prediction error filter are nearly equal. For pitch-asynchronous analysis, the backward and forward output powers are of the same order (their ratio for the synthetic vowel /o/ is 1.01). This explains the satisfactory performance of Burg's method for pitch-asynchronous analysis. For pitch-synchronous analysis, the duration of the analysis segment is comparatively less and the first M samples of the analysis segment have larger magnitude than the last M samples. This makes the backward output power to be much larger than

the forward output power (the ratio of these two powers is 2.1 for the synthetic vowel /o/). This explains the poor performance of Burg's method for pitch-synchronous analysis.

However, Burg's method has the distinct advantage that it always guarantees the stability of the estimated all-pole filter, even with finite wordlength computations. The autocorrelation method of linear prediction guarantees the stability of the all-pole filter only with floating-point computations; the covariance method of linear prediction does not guarantee stability even with floating-point computations. Computationally, Burg's method is approximately five times more expensive than the autocorrelation and covariance methods of linear prediction (the number of multiplications for the Burg's, autocorrelation and covariance methods are approximately $5MN$, $MN+M^2$ and $MN+\frac{1}{6}M^3+\frac{3}{2}M^2$, respectively [10]).

Thus, if we consider the performance of the method, its computational cost and the stability of the estimated all-pole filter to decide the suitability of the method for the analysis of voiced speech, we can conclude that Burg's method is not suitable for pitch-synchronous analysis; it may be used for pitch-asynchronous analysis (specially when only finite wordlength arithmetic is available for computations).

References

[1] H.R. Radoski, P.F. Fougere and E.J. Zawalick, "A comparison of power spectral estimates and applications of maximum entropy method", *J. Geophys. Res.*, Vol. 80, No. 4, Feb. 1975, pp. 619–625.

[2] T.J. Ulrych and T.N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition", *Rev. Geophysics and Space Phys.*, Vol. 13, No. 1, Feb. 1975, pp. 183–200.

[3] T. Dyson and S. Rao, "Some detection and resolution properties of maximum entropy spectral analysis", *Signal Processing*, Vol. 2, No. 3, July 1980, pp. 261–270.

[4] A. van den Bos, "Alternative interpretation of maximum entropy spectral analysis", *IEEE Trans. Information Theory*, Vol. IT-17, No. 4, July 1971, pp. 493–494.

[5] J.P. Burg, "A new analysis technique for time series data", paper presented at Advanced Study Institute on Signal Processing, NATO, Enschede, Netherlands, Aug. 12–23, 1968.

[6] N. Anderson, "On the calculation of filter coefficients for maximum entropy analysis", *Geophysics*, Vol. 39, No. 1, Feb. 1974, pp. 69–72.

[7] J. Makhoul, "Linear prediction: A tutorial review", *Proc. IEEE*, Vol. 63, No. 4, April 1975, pp. 561–580.

[8] J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.

[9] K.K. Paliwal and P.V.S. Rao, "Windowing in linear prediction analysis of voiced speech", *J. Acoust. Soc. Amer.*, Vol. 66, Nov. 1979, pp. S63(A).

[10] J. Makhoul, "Stable and efficient lattice methods for linear prediction", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-25, No. 5, Oct. 1977, pp. 423–428.