Signal Processing 3 (1981) 181–185
North-Holland Publishing Company

SHORT COMMUNICATION

A MODIFIED AUTOCORRELATION METHOD OF LINEAR PREDICTION FOR PITCH-SYNCHRONOUS ANALYSIS OF 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 11 July 1980
Revised 20 October 1980 and 15 December 1980

Abstract. A modified autocorrelation method of linear prediction is proposed for pitch-synchronous analysis of voiced speech. The method needs one full period of speech data for analysis and assumes periodic extension of the data. This method guarantees the stability of the estimated all-pole filter and is shown to perform better than the covariance and autocorrelation methods of linear prediction.

Zusammenfassung. Für die grundperiodensynchrone Analyse stimmhafter Sprachsignale wird eine modifizierte Version der Autokorrelationsmethode der linearen Prädiktion vorgeschlagen. Das Verfahren benötigt eine volle Grundperiode des Sprachsignals für die Analyse; hierbei wird angenommen, daß sich außerhalb dieser Grundperiode das Signal periodisch fortsetzt. Das Verfahren garantiert die Stabilität des ermittelten Prädiktorfilters; es wird gezeigt, daß es bessere Ergebnisse liefert als die Kovarianzmethode der linearen Prädiktion.

Résumé. Pour l'analyse synchronisée à la fondamentale de la parole voisée, on propose une méthode d'autocorrélation modifiée de la prédiction linéaire. La méthode nécessite une période complète des données pour l'analyse et est basée sur l'hypothèse d'une extention périodique des donées. Cette méthode garantie la stabilité du filtre tout-pôle éstimé et il est montré qu'elle est meilleure que les méthodes de covariance et d'autocorrélation de la prédiction linéaire.

Keywords. Speech, pitch-synchronous analysis, linear prediction, autocorrelation method, covariance method, spectrum, formants.

1. Motivation

For pitch-synchronous analysis of voiced speech (where the analysis-segment duration is less than or equal to one pitch period), the autocorrelation method as well as the covariance method of linear prediction are unacceptable because of the following reasons. The performance of the autocorrelation method is not good [3], though it guarantees the stability of the estimated all-pole filter. The covariance method performs well, but it does not always lead to a stable all-pole filter [2]. So there is a need for a method for pitch-synchronous analysis of voiced speech which can perform as well as or better than the covariance method and can guarantee the stability of the estimated all-pole filter. In the present paper, we have proposed one such method.

2. Method

In the autocorrelation method of linear prediction, it is assumed that the signal is defined for all time such that it is identically zero outside a portion of the signal N samples long, where N is some positive integer [5, 6]. This is accomplished by weighting the speech signal by a finite window

0165-1684/81/0000-0000/$02.50

of length N. This windowing causes unwanted spectral distortion which is more for smaller values of N. Thus if the signal can be defined for all time without windowing, such spectral distortion will not occur.

The method proposed in this paper is a modification over the autocorrelation method and can be used for pitch-synchronous analysis of voiced speech signal. Here the analysis-segment duration N is exactly equal to one pitch period (i.e., the analysis segment consists of all the samples between two consecutive pitch pulses) and instead of assuming zero extension of the signal beyond analysis-segment interval, a periodic extension of the signal is assumed. Thus, the signal that is known over the duration $0 \leq n \leq N-1$ is now known over the duration $-\infty < n < \infty$ such that

$$x(n) = x(kN+n) \tag{1}$$

where k is some integer. Thus since the signal is not windowed prior to its spectral analysis, the autocorrelation method of linear prediction can be used safely for the pitch-synchronous analysis of voiced speech signal. Here the linear predictor coefficients $\{a_k\}$ which satisfy the linear prediction equation

$$\tilde{x}(n) = -\sum_{k=1}^{M} a_k x(n-k) \tag{2}$$

(where $\tilde{x}(n)$ is the value of nth sample predicted from the past M samples) can be computed by solving the following set of equations:

$$\sum_{k=1}^{M} a_k R(|i-k|) = -R(i), \quad i = 1, 2, \ldots, M \tag{3}$$

where

$$R(i) = \lim_{L\to\infty} \frac{1}{(2L+1)} \sum_{n=-L}^{L} x(n)x(n+i)$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+i). \tag{4}$$

Eq. (3) can be solved for the linear predictor coefficients $\{a_k\}$ by using an efficient recursive procedure given elsewhere [5, 6].

3. Performance of the method

In this section, we compare experimentally the performance of the modified autocorrelation method with that of the covariance and autocorrelation methods of linear prediction[1]. For this, synthetic vowel signals are used for analysis and accuracy in estimating the power spectrum and formant frequencies is used as criterion. 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 speech sample with the maximum value in a pitch period is taken as the first sample of the analysis segment. When using the autocorrelation method of linear prediction for the analysis of a speech segment, the speech signal is weighted by a Hamming window function prior to its analysis; no such windowing is carried out for the modified autocorrelation and covariance methods [6, 7].

(a) *Accuracy in estimating the power spectrum*

For comparing the three methods using accuracy in estimating the power spectrum as criterion, the synthetic vowel signal is generated by exciting an all-pole filter

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

by a periodic train of impulses. Here, the gain factor G and the filter coefficients $\{a_k\}$ are known a priori. The pitch-synchronous power spectrum estimates $\hat{P}(f)$ are obtained from this signal by using the three methods and then these three estimates are compared with one another using the true power spectrum

$$P(f) = |H\{\exp(2\pi \mathrm{j} f T)\}|^2$$

as the reference spectrum (T is sampling period).

[1] The computational cost is approximately proportional to $MN + M^2$ for the autocorrelation and modified autocorrelation methods and to $MN + \frac{1}{6}M^3 + \frac{3}{2}M^2$ for the covariance method [8].

Comparison is made by visual inspection as well as by computing the spectral bias B which is defined as

$$B = 2T \int_0^{1/2T} [|\hat{P}(f) - P(f)|/P(f)]\, df. \qquad (5)$$

As an example, we show in Fig. 1 the true spectrum of a synthetic vowel signal and its pitch-synchronous estimates (predictor order $M = 10$ and analysis segment duration $N = 80$ samples). It can be seen from this figure that the spectrum estimated by the modified autocorrelation method is closest to the true spectrum. Also, the value of spectral bias B is 0.115 for the modified autocorrelation method, 0.294 for the covariance method and 0.957 for the autocorrelation method. Thus the modified autocorrelation method leads to least spectral bias. Similar results are obtained for other synthetic vowels. Thus the modified autocorrelation method estimates the power spectrum more accurately than the covariance and autocorrelation methods.

(b) *Accuracy in estimating the formant frequencies*

For comparing the performance of the three methods using accuracy in estimating the formant frequencies as the criterion, a digital formant synthesizer [4] is used to synthesize the signal for nine different vowels. The values of formant parameters for these synthesized vowels are taken from reference [3]. The synthetic vowel signals are then analysed and the first three formant frequencies are estimated by computing the zeroes of the polynomial $[1 + \sum_{k=1}^{M} \hat{a}_k z^{-k}]$ where $\{\hat{a}_k\}$ are the estimated linear predictor coefficients. Table 1 gives the errors in estimating the first three formant frequencies of these nine synthesized vowels. It can be seen from this table that the estimation errors made by the modified autocorrelation method are smaller in magnitude than those made by the covariance and autocorrelation methods.

The modified autocorrelation method is also tried out on a number of vowel segments from natural speech. For these segments, the sample with the maximum value in a pitch period (which is

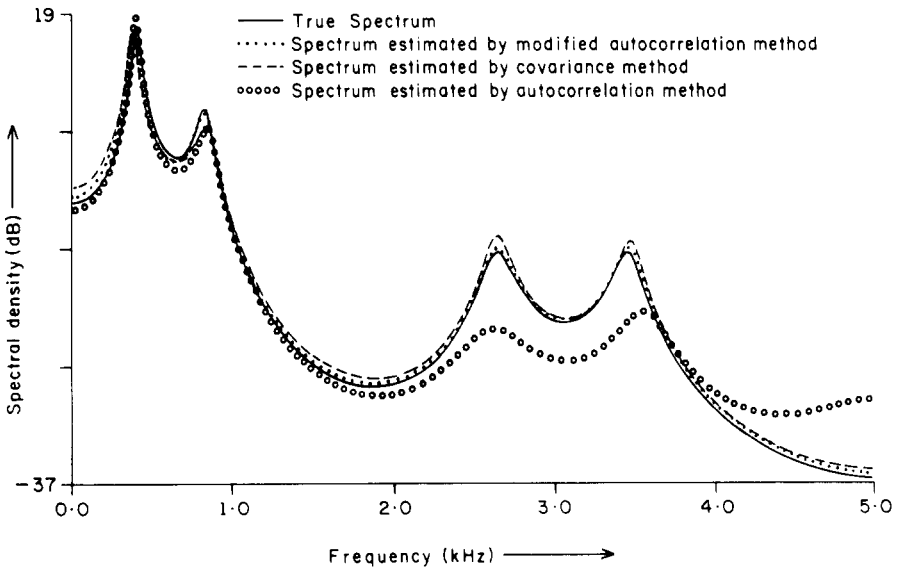

Fig. 1. True spectrum of the synthetic vowel /o/ and its pitch-synchronous estimates. (For synthetic vowel, $G = 0.1354$, $a_1 = -1.53527$, $a_2 = 0.97789$, $a_3 = -1.48396$, $a_4 = 1.78023$, $a_5 = -0.71704$, $a_6 = 0.73514$, $a_7 = -0.76348$, $a_8 = -0.12135$, $a_9 = 0.15552$, $a_{10} = 0.178143$, sampling frequency = 10 kHz, pitch period = 8 msec).

Table 1
Errors (EF 1, EF 2 and EF 3) in estimating the first three formant frequencies in Hertz for nine synthetic vowels.

Vowel		Modi. Auto. Method	Cova. Method	Auto. Method
/i/ (heed)	EF1	−16.0	−26.4	21.7
	EF2	−9.7	−23.1	−20.8
	EF3	−4.9	−11.8	−6.4
/I/ (hid)	EF1	−11.2	−13.9	17.2
	EF2	4.5	11.3	−1.2
	EF3	−2.6	−15.9	4.0
/ɛ/ (head)	EF1	−16.5	−29.6	6.3
	EF2	8.9	13.7	4.8
	EF3	6.8	10.5	7.2
/æ/ (had)	EF1	−15.7	−29.2	−0.9
	EF2	8.0	23.1	7.5
	EP3	−6.2	−18.1	9.3
/ʌ/ (hud)	EF1	−13.1	−22.0	0.8
	EF2	0.9	3.9	39.4
	EF3	3.1	9.7	−28.8
/a/ (hod)	EF1	13.6	20.3	−17.8
	EF2	13.1	18.5	−29.8
	EF3	2.0	3.3	21.1
/ɔ/ (hawed)	EF1	6.6	17.8	−22.7
	EF2	3.0	40.3	0.7
	EF3	−5.7	−23.2	75.7
/u/ (hood)	EF1	0.1	7.9	1.5
	EF2	−10.1	−22.4	25.7
	EF3	3.1	8.4	−34.0
/ɜ/ (herd)	EF1	8.4	13.0	1.2
	EF2	12.5	16.2	55.5
	EF3	4.4	−3.0	124.4

used for synchronisation) is manually detected. (No attempt is made here to give an automatic method for the detection of the maximum in a pitch period; this problem is related to pitch estimation and has been studied by others [1, 9].) The performance of the method is found to be good for natural speech. However, no objective evaluation of the method could be done for natural speech because of the unavailability of true values of the spectral parameters.

4. Conclusion

A modified autocorrelation method of linear prediction is proposed for pitch-synchronous analysis of voiced speech. It is shown that this method estimates the power spectrum and formant frequencies more accurately than the covariance and autocorrelation methods of linear prediction. Computationally, this method is as efficient as the autocorrelation method of linear prediction and more efficient than the covariance method of linear prediction. Furthermore, this method, like the autocorrelation method of linear prediction, guarantees the stability of the estimated all-pole filter. Thus, the proposed method combines in itself the advantages of both the autocorrelation and the covariance methods of linear prediction and hence is recommended for pitch-synchronous analysis of voiced speech.

Acknowledgement

The authors would like to thank the referees for their helpful suggestions, which added to the usefulness of this paper.

References

[1] T.V. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval", *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-27, No. 4, Aug. 1979, pp. 309–319.

[2] B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave", *J. Acoust. Soc. Amer.*, Vol. 50, No. 2, Aug. 1971, pp. 637–655.

[3] S. Chandra and W.C. Lin, "Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis", *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-22, No. 6, Dec. 1974, pp. 403–415.

[4] J.L. Flanagan, C.H. Coker, L.R. Rabiner, R.W. Schafer and N. Umeda, "Synthetic voices for computers", *IEEE Spectrum*, Vol. 7, No. 10, Oct. 1970, pp. 22–45.

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

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

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

[8] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978, Ch. 8, pp. 417–418.

[9] H.W. Strube, "Determination of the instant of glottal closure from the speech wave", *J. Acoust. Soc. Amer.*, Vol. 56, No. 5, Nov. 1974, pp. 1625–1629.