Speech Communication 3 (1984) 221–231
North-Holland

PERFORMANCE OF THE WEIGHTED BURG METHODS OF AR SPECTRAL ESTIMATION FOR PITCH-SYNCHRONOUS ANALYSIS OF VOICED SPEECH

K.K. PALIWAL*
Division of Telecommunications, University of Trondheim, Trondheim-NTH, Norway

Received 6 January 1984
Revised 11 September 1984

Abstract. Recently three different modifications over the Burg method of autoregressive (AR) spectral estimation are proposed by Swingler [5], Kaveh and Lippert [6], and Scott and Nikias [7], where the linear prediction error is weighted prior to its minimization. In the present paper, we study these weighted Burg methods for pitch-synchronous analysis of short segments (duration less than one pitch period) of non-nasalized voiced speech and make their comparative performance evaluation. Errors in estimating the power spectrum, formant frequencies and formant bandwidths are used as criteria for performance evaluation. It is shown that the weighted Burg method of Kaveh and Lippert results in the best performance. These methods are also compared with the autocorrelation and covariance methods and the results are discussed.

Zusammenfassung. Kürzlich wurden drei Varianten der Spektralanalyse nach Burg, welche auf dem autoregressiven Modell beruht, von Swingler [5], Kaveh und Lippert [6] und Scott und Nikias [7] vorgestellt. Diese neuen Methoden ponderieren den Prädiktionsfehler vor dessen Minimisierung. Der Beitrag beschäftigt sich mit einer vergleichenden Studie der ponderierenden Methoden im Hinblick auf ihre Anwendung auf die synchrone Bestimmung des Grundtons stimmhafter, nicht nasalierter Segmente kurzer Dauer des Sprachsignals. Die Abweichungen in der Bestimmung des Leistungsspektrums, der Formantenfrequenzen und Bandbreiten werden als Kriterien in einer vergleichenden Leistungsbewertung herangezogen und es wird aufgezeigt, dass die Methode nach Lippert und Kaveh am leistungsfähigsten ist. Die Resultate eines Vergleichs der ponderierenden Methoden mit der Autokorrelationsanalyse und mit der Kovarianzanalyse werden ebenfalls diskutiert.

Résumé. Trois modifications à la méthode d'analyse spectrale de Burg basée sur le modèle autorégressif ont été proposées récemment par Swinger [5], Kaveh et Lippert [6], ainsi que par Scott et Nikias [7]. Ces méthodes modifiées pondèrent l'erreur de prédiction avant sa minimisation. L'objet de cet article est une étude comparée des performances de ces méthodes pondérées en vue de leur application à court terme synchronisée à la fondamentale de brefs segments voisés non nasalisés (durée de l'intervalle d'analyse plus courte qu'une période). Les erreurs dans l'estimation du spectre de puissance, des fréquences formantiques et des largeurs de bande sont utilisées comme critère de performance. Nous montrons que la méthode de Burg pondérée proposée par Kaveh et Lippert est la plus performante. Les résultats d'une comparaison de ces méthodes avec les méthodes d'autocorrélation et de covariance sont également discutés.

Keywords. Spectral estimation, Burg method, pitch-synchronous analysis, speech.

1. Introduction

In an earlier paper [1], we studied the Burg method of autoregressive (AR) spectral estimation for non-nasalized voiced speech and compared its performance with that of the autocorrelation and covariance methods. We found that for pitch-asynchronous analysis of speech segments (of duration more than two times the pitch period), the performance of the Burg method was comparable to that of the autocorrelation and covariance methods. For pitch-synchronous analysis of speech segments (of duration less than one pitch period), we showed that the Burg method did not perform as well as the other two methods. In particular, the formant frequencies

* Present address: Computer Systems and Communications Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay-400005, India.

were estimated with relatively large errors and the formant bandwidths were grossly underestimated. Similar results have been reported by Gray and Wong [2].

In speech processing applications (such as speech analysis-synthesis and speech recognition), it is sometimes desirable to analyse short speech segments (shorter than 10 ms) for tracking the dynamic behaviour of the speech signal during transients. These short segments can not be analysed in a pitch–asynchronous manner because the arbitrary placement of speech segments with respect to the pitch pulses causes too much of error in parameter estimation [3]. Thus, a pitch-synchronous analysis of short segments is required to follow the fast transitions in speech occurring when the articulators move rapidly from one phoneme position to another. If this pitch-synchronous analysis is performed over a portion of a pitch period where the glottis is closed (i.e., the force-free or undriven portions of the speech signal), the estimated AR parameters will represent only the vocal tract system and will have no interaction from the glottal source. (See [4] and references given therein for other advantages of the pitch-synchronous analysis over the pitch-asynchronous analysis.) Since the Burg method does not perform well for pitch-synchronous analysis of short speech segments, it becomes necessary to modify it.

Fortunately, three modifications over the Burg method are recently reported by Swingler [5], Kaveh and Lippert [6], and Scott and Nikias [7]. These modifications have been proposed to overcome a long-standing problem associated with the Burg method when used for analysing the sinusoid signals. The Burg method is known to result in relatively high and phase-dependent bias in estimating the sinusoid frequencies, specially for short analysis segments [8]. These modifications have been found to reduce the phase-dependent bias in estimating the frequencies of sinusoid signals and their comparative performance results are reported elsewhere [9].

These modifications over the Burg method use some form of weighting of instantaneous linear prediction errors prior to their minimization. For example, Swingler [5] has used a Hamming window function to weight the instantaneous prediction errors. Kaveh and Lippert [6] have analytically derived a window function (called optimum tapered window function) which minimizes the average bias in estimating the sinusoid frequency. Scott and Nikias [7] have used the energies of the past and future samples to weight the present instantaneous prediction error in their minimization procedure. In the present paper, these modifications will be referred to as the Hamming-windowed Burg method (BURGH), the optimum-tapered Burg method (BURGO) and the energy-weighted Burg method (BURGE). The original Burg method will be referred to as the rectangular-windowed Burg method (BURGR).

Since the problem of relatively high frequency estimation bias associated with the Burg method for short analysis segments is common to both the sinusoid and voiced speech signals, it is only natural to inquire how these modifications improve the performance of the Burg method when applied to short segments of the voiced speech signals. So, the aim of the present paper is to study the weighted Burg methods (BURGR, BURGH, BURGO and BURGE) for pitch-synchronous analysis of non-nasalized voiced speech and make their comparative performance evaluation. Since the autocorrelation (AUTO) and covariance (COVA) methods are popular in speech processing literature [10,11], we also include them in the present performance evaluation study.

2. Criteria for performance evaluation

In order to evaluate the performance of the different AR spectral estimation methods, we employ here the following three criteria:

1) error in estimating the power spectrum,

2) normalized errors in estimating the first four formant frequencies, and

3) normalized errors in estimating the first four formant bandwidths.

Error (E_s) in estimating the power spectrum is measured in terms of the cross-correlation distance measure which is defined as follows [12,13]:

$$E_s = 1 - \frac{\int_0^{F_s/2} P(f)\hat{P}(f)\mathrm{d}f}{\left(\int_0^{F_s/2} \{P(f)\}^2\mathrm{d}f\right)^{1/2}\left(\int_0^{F_s/2} \{\hat{P}(f)\}^2\mathrm{d}f\right)^{1/2}},$$

where F_s is the sampling frequency. $P(f)$ and $\hat{P}(f)$ are the actual and the estimated power spectra, respectively, and f is the frequency variable. This type of distance measure has an important property that it remains unchanged even if either the estimated spectrum or the actual spectrum both are multiplied by a constant [12,13]. Thus, we need not normalize the estimated and actual power spectra for computing the spectrum estimation error E_s. The cross-correlation distance measure has been successfully used earlier in speech recognition [13,15] and speaker recognition [12,16] applications.

Normalized error, F_{ie}, in estimating the i-th formant frequency is defined as

$$F_{ie} = (\hat{F}_i - F_i)/F_i,$$

where F_i *and* $\hat{F}_i$ are the actual and the estimated frequencies, respectively, of the i-th formant.

Similarly, normalized error, BW_{ie}, in estimating the i-th formant bandwith is defined as

$$BW_{ie} = (\hat{BW}_i - BW_i)/BW_i,$$

where BW_i and $\hat{BW}_i$ are the actual and the estimated bandwidths, respectively, of the i-th formant.

The bandwidth estimation errors are used in the present evaluation as secondary criteria; i.e., these are invoked only when a particular spectral estimation method is found to give reasonable estimates of formant frequencies. If the formant frequency estimates are having relatively large errors, it does not matter whether bandwidths of those formants are estimated correctly. In order to say whether the formant frquency estimates are reasonable, we use difference limens (also called as the just noticable differences [17]) for formant frequencies obtained from human perception experiments as thresholds. If the error (F_{ie}) in estimating the i-th formant frequency is less than the difference limen for this frequency, we consider the i-th formant frequency estimate to be reasonable. A similar reasoning is applied when the formant bandwidth errors are used as performance evaluation criteria. In the present paper, difference limens are taken to be 5% for the formant frequencies and 40% for the formant bandwidths [17, pp. 279–281].

3. Performance evaluation and results

In this section, we conduct a comparative performance evaluation of the different AR spectral estimation methods (BURGR, BURGH, BURGO, BURGE, AUTO and COVA) for pitch-synchronous analysis of non-nasalized voiced speech. For this, these methods are studied on a number of synthetic as well as real speech signals representing different non-nasalized voiced speech sounds. However, for illustrating our results we employ here three typical examples; the first for the synthetic speech signal and the second and third for the real speech signals. It might be noted that study of synthetic speech signals offers two main advantages: firstly, the actual values of the AR model parameters are known a priori which helps in making objective evaluation of analysis methods and, secondly, the location of the pitch pulse is known a priori which is needed for pitch-synchronous analysis. Though the real speech signals do not offer these advantages, we can still use them, as discussed later, for performance evaluation.

Example 1 (Synthetic speech)

We select here the same example as used in our earlier papers [1,18]. In this example, the synthetic speech signal is generated at sampling rate of 10 kHz from an AR (all-pole) system

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^{M} a_k z^{-k}},$$

where M is the order of the AR system (equal to 10 in this example), z the z-transform variable and $\{a_k\}$ are the AR parameters (or linear predictor coefficients). A periodic train of impulses (i.e., unit samples separated from each other by a sequence of zero samples of duration equal to one pitch period) is used as an excitation signal to this AR system. The pitch period is equal to 8 ms. The AR parameters $\{a_k\}$ used for synthesis

are determined from a real speech signal of vowel /o/. These are listed in [1,18]. The actual values of the first four formant frequencies are 403, 834, 2645 and 3472 Hz, and bandwidths are 53, 97, 188 and 170 Hz.

The synthetic speech signal is analysed in a pitch-synchronous manner. For this, as mentioned earlier, the beginning sample of the speech segment should be aligned to the instant of glottal closure and the length of the speech segment should be less than or equal to the duration of glottal closure. Since the instant of glottal closure is the point of maximum excitation in voiced speech production [19,20], the speech sample with maximum value in a pitch period is taken as the first sample of the analysis segment. Mathews et al. [21], Rosenberg [22], and Matausek and Batalov [23] have reported the duty cycles (ratios of the open-glottal time to the pitch period) computed from the speech signal to be less than 0.5. Photographic evidence also indicates values of 0.33 to 0.5 to be typical of the duty cycle [21]. From these experimental evidences, the duration of glottal closure can be considered to be about 50% of the pitch period. In the present analysis, the duration of the speech segment is taken to be 45% of the pitch period.

This speech segment is analysed by the six different AR spectral estimation methods (BURGR, BURGH, BURGO, BURGE, AUTO and COVA), using a 10-th order AR model for parameter estimation. The speech signal is weighted by a Hamming window function prior to its analysis by the autocorrelation (AUTO) method, while no such windowing is done for the covariance (COVA) method [10,11,24]. The resulting power spectra are shown for different methods in Fig. 1 along with the actual power spectrum[1]. Error in estimating the power spectrum and normalized error (in %) in estimating the formant frequencies and bandwidth are listed in Table 1 for these methods. The formant frequencies and bandwidths are computed by solving the denominator of the estimated AR system function for its roots. In this table, all formant frequency estimation errors and bandwidths estimation errors are marked with asterisk (*) if they exceed the corresponding difference limens (which are 5% and 40%, respectively).

We can make the following observations from Table 1 and Fig. 1:

[1] The power spectra are computed by evaluating the estimated AR system transfer function $\hat{H}(z) = 1/\hat{A}(z)$ on the unit circle.

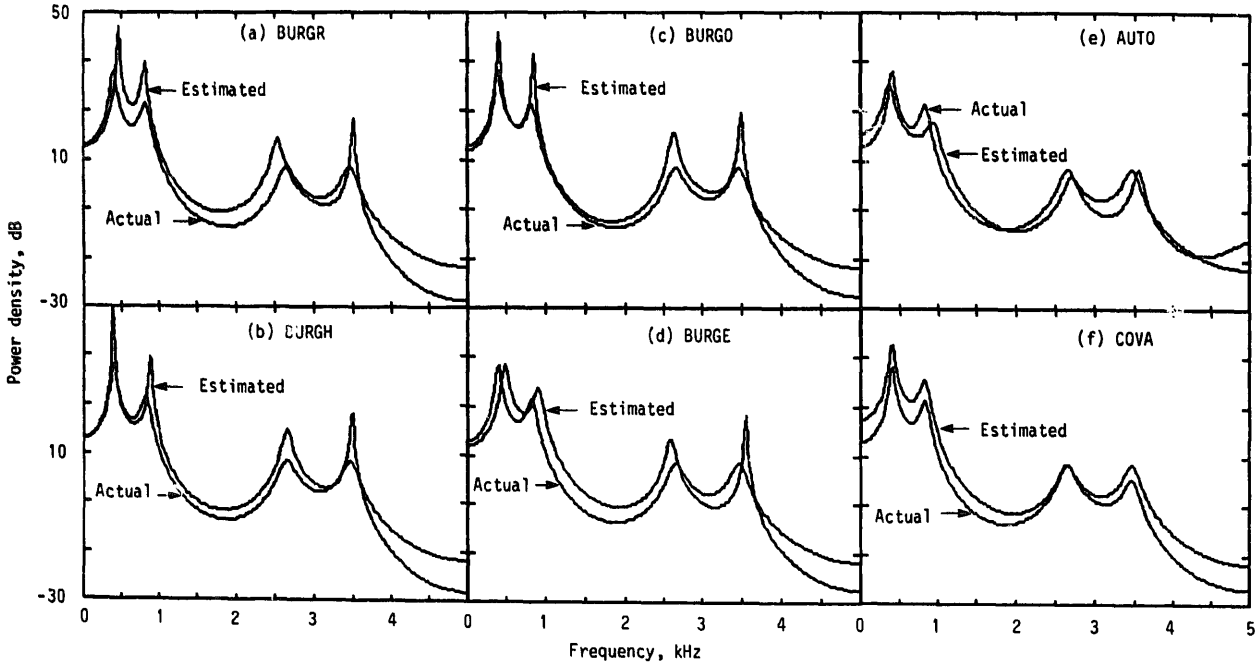

Fig. 1. Pitch-synchronous spectral estimates of the synthetic speech signal of vowel /o/ along with its original power spectrum. (a) BURGR method, (b) BURGH method, (c) BURGO method, (d) BURGE method, (e) AUTO method and (f) COVA method.

Table 1
Error in estimating the power spectrum and normalized errors (in %) in estimating the formant frequencies and bandwidths for pitch-synchronous analysis of the synthetic speech signal of vowel /o/. Pitch period is 8 ms.

Error	BURGR	BURGH	BURGO	BURGE	AUTO	COVA
E_s	0.79586	0.38621	0.23076	0.68475	0.24154	0.00004
F_{1e}	16.6*	−2.4	3.6	21.6*	−9.0*	0
F_{2e}	−2.0	6.6*	2.3	9.7*	12.2*	0
F_{3e}	−4.1	0.4	−0.6	−2.2	2.1	0
F_{4e}	1.1	0.5	0.7	2.3	2.4	0
BW_{1e}	−65.5*	−85.3*	−88.6*	12.9	77.7*	0
BW_{2e}	−59.2*	−80.5*	−85.9*	−5.8	56.4*	0
BW_{3e}	−43.4*	−47.1*	−53.7*	−48.8*	−12.2	0
BW_{4e}	−89.3*	−87.0*	−84.3*	−88.8*	−40.1*	0

Note: Errors marked with asterisk (*) are unacceptable by the difference limen criterion.

1) The original Burg method (BURGR) performs very poorly in comparison to the covariance method (COVA). Its performance is also inferior to that of the autocorrelation method (AUTO). Similar results are reported in our earlier paper [1] where speech segments of duration 0.8 times the pitch period were used for pitch-synchronous analysis.

2) Now we study how the modified methods (i.e., BURGH, BURGO and BURGE) affect the pitch-synchronous analysis performance of the Burg (BURGR) method. In terms of spectral estimation error (E_s), the BURGH and BURGO methods are better than the BURGR method, but the BURGE method is worse than the BURGR method. In terms of formant frequency estimation errors, the BURGR method estimates F_1 wrongly. The BURGH method corrects this estimation error, but estimates F_2 wrongly. The BURGO method estimates both F_1 and F_2 correctly. The BURGE method, on the contrary, estimates both of these formant frequencies wrongly. In terms of formant bandwidth estimation errors, the BURGR, BURGH and BURGO methods grossly underestimate the formant bandwidths. The BURGE method estimates BW_1 and BW_2 correctly, but it is of no use because it estimates the corresponding formant frequencies (i.e., F_1 and F_2) wrongly. Thus, we can say that the BURGH and BURGO methods provide an improvement in performance over the BURGR method (the BURGO method resulting in more improvement than the BURGH method). The BURGE method leads to more inferior performance than the BURGR method. It might be noted that like the BURGR and BURGH methods the BURGO method also has the problem of formant bandwidth underestimation. This problem is studied in Section 4.

3) Now we compare the performance of the BURGO method (which is the best among the weighted Burg methods) with that of the AUTO and COVA methods. In terms of spectral estimation error (E_s), the BURGO method is comparable to the AUTO method, but inferior to the COVA method. In terms of formant frequency estimation errors, the BURGO method performs better than the AUTO method. Though the formant frequency estimation errors by the BURGO method are acceptable according to the difference limen criterion, these errors are larger in magnitude than those due to the COVA method. In terms of formant bandwidth estimation errors. the BURGO method does not perform as well as the AUTO and COVA methods. As mentioned earlier, it has the problem of formant bandwidth underestimation (which is shared by other Burg methods). This problem will be discussed later in Section 4.

So far we have studied the performance of the AR spectral estimation methods for the noise-free synthetic speech signal. Now we study these methods for the synthetic signal distorted by the additive white Gaussian noise. For illustrating our results, we use here the noisy synthetic speech signal at signal-to-noise ratio (SNR) of 10 dB. For

this, we generate fifty different realizations of white Gaussian noise and compute the normalized bias and standard deviation in estimating the formant frequencies and bandwidths. The normalized bias and standard deviation (SD) in estimating the i-th formant frequency is defined as follows:

$$\text{Bias } (F_i) = (\langle \hat{F}_i \rangle - F_i)/F_i$$

and

$$SD(F_i) = [\langle (\hat{F}_i - \langle \hat{F}_i \rangle)^2 \rangle]^{1/2}/F_i,$$

where $\langle\ \rangle$ denotes ensemble averaging operation. The normalized bias and standard deviation in estimating the formant bandwidths can be defined in a similar manner.

The values of these biases and standard deviations for the formant frequencies and bandwidths are listed in Table 2 for the six AR spectral estimation methods. It can be seen from this table that in terms of biases, the results remain the same as those obtained for the noise-free case discussed earlier. But, in terms of standard deviations the BURGO method gives the best performance among the six methods.

Examples 2 and 3 (Real speech)

Now we study the performance of different AR spectral estimation methods for pitch-synchronous analysis of real speech. As mentioned earlier, we come across two problems for the real speech signals; firstly, the position of the maximum of the speech signal within a pitch period is not known a priori, and, secondly, the actual values of AR parameters are not known a priori which makes the objective evaluation of the spectral estimation method difficult. The first problem is solved by visually examining the speech waveform. The pitch period is manually isolated

Table 2
Normalized bias and standard deviation (SD), in %, in estimating the formant frequencies and bandwidths for pitch-synchronous analysis of the noisy synthetic speech signal of vowel /o/. SNR = 30 db. 50 different realizations of noise are used. Pitch period is 8 ms.

		BURGR	BURGH	BURGO	BURGE	AUTO	COVA
F_1	Bias	16.9*	−2.2	3.7	22.1*	−9.5*	0.1
	SD	0.5	0.2	0.2	2.3	0.6	0.5
F_2	Bias	−1.5	7.1*	2.8	10.4*	11.4*	0.5
	SD	0.9	0.5	0.5	3.4	1.0	0.7
F_3	Bias	−4.2	−0.3	−0.8	−2.5	2.0	−0.2
	SD	1.1	0.8	0.8	0.8	1.0	0.8
F_4	Bias	1.1	0.2	0.3	2.2	2.7	0.0
	SD	0.4	0.5	0.5	1.2	0.6	0.5
BW_1	Bias	−51.1*	−77.8*	−80.1*	15.6	67.0*	8.6
	SD	9.4	5.5	5.1	32.6	13.5	11.5
BW_2	Bias	−44.0*	−72.3*	−76.3*	−5.9	55.1*	6.7
	SD	11.7	6.9	6.2	33.4	12.2	11.9
BW_3	Bias	−54.3*	−54.9*	−61.2*	−46.9*	16.6	3.8
	SD	10.6	12.6	10.3	17.5	22.7	20.1
BW_4	Bias	−62.0*	−58.7*	−59.7*	−64.7*	−11.4	10.4
	SD	17.1	19.5	17.7	17.2	27.5	28.7

Note: Biases marked with asterisk (*) are unacceptable by the difference limen criterion.

from the speech signal and the position of maximum within this pitch period is determined. For solving the second problem, we note from the results of Example 1 that the covariance (COVA) method results in almost error-free estimation of formant frequencies and bandwidths. Hence, we can use the parameter values estimated by this method as reference values for making objective evaluation of other spectral estimation methods.

In Example 2, we perform pitch-synchronous analysis of the real speech signal of vowel /a/. Here the reference formant frequencies are 716, 1152, 2634 and 3604 Hz, and bandwidths are 85, 90, 93 and 179 Hz. The pitch period for this example is 8.1 ms and speech segment of duration 0.45 times the pitch period is used here for pitch-synchronous analysis. Figure 2 shows the power spectra estimated by the BURGR, BURGH, BURGO and BURGE methods along with the reference spectrum. The normalized errors in estimating the formant frequencies and bandwidths are listed in Table 3 for the four weighted Burg methods and the AUTO method. It can be seen from Fig. 2 and Table 3 that the BURGR method results in relatively large error in estimating F_1. The BURGH method corrects this estimation error and the BURGO method further improves the estimation. On the contrary, the BURGE method causes larger errors in estimating the formant frequencies than the BURGR method. The BURGR, BURGH and BURGO methods underestimate the formant bandwidths here also. We can also see from Table 3 that the BURGO method gives more correct estimates of the formant frequencies than the AUTO method. These results are similar to those obtained in Example 1 for synthetic speech.

In Example 3, we study the real speech signal of vowel /æ/. For this example, we have the reference formant frequencies 673, 1773, 2583 and 3585 Hz, and bandwidths 52, 71, 226 and 295 Hz.

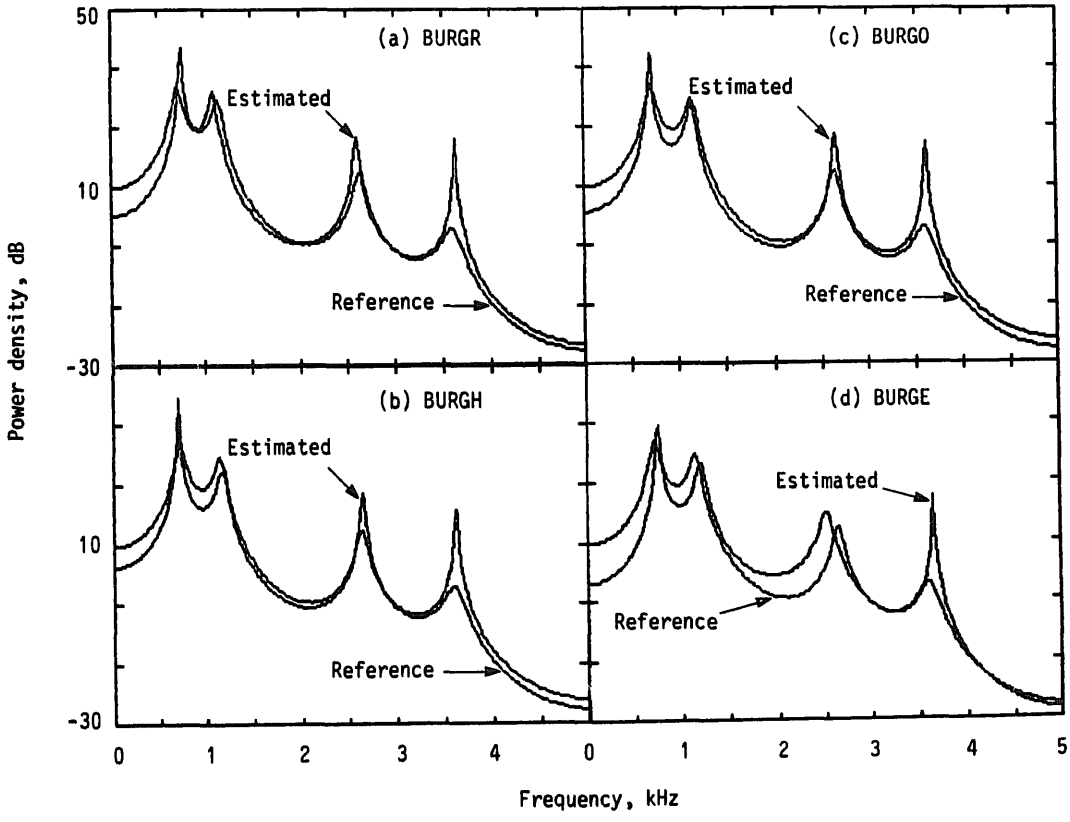

Fig. 2. Pitch-synchronous spectral estimates of the real speech signal of vowel /a/ along with its reference power spectrum. (a) BURGR method, (b) BURGH method, (c) BURGO method and (d) BURGE method.

Here the pitch period is 7.9 ms and speech segment of duration 0.45 times the pitch period is used for analysis. Results of pitch-synchronous analysis for this example are shown in Fig. 3 and Table 4. Here also we observe similar results as obtained in Examples 1 and 2. That is, the BURGH and BURGO methods perform better than the BURGR method, but the BURGE method does not perform as well as the BURGR method. Also the BURGO method results in better formant frequency estimates than the AUTO method. But this method (like other Burg methods) causes underestimation of the formant bandwidths.

It might be noted that the results described so far use speech segments of duration 0.45 times the pitch period for pitch-synchronous analysis. We have also studied the pitch-synchronous analysis performance of these methods for different analysis segment durations (equal to 0.4, 0.5 and 0.6 times the pitch period) and obtained similar results.[2]

Table 3
Normalized errors (in %) in estimating the formant frequencies and bandwidths for pitch-synchronous analysis of the real speech signal of vowel /a/. Pitch period is 8.1 ms.

Error	BURGR	BURGH	BURGO	BURGE	AUTO
F_{1e}	5.8*	−1.6	−0.7	5.4*	−5.6*
F_{2e}	−4.2	1.6	−0.5	4.7	0.4
F_{3e}	−1.3	0.3	0.4	−4.7	−0.1
F_{4e}	0.8	0.5	0.4	1.3	0.9
BW_{1e}	−73.9*	−80.7*	−83.7*	−68.3*	17.7
BW_{2e}	−28.6	−4.2	−22.3	−11.1	−175.4*
BW_{3e}	−57.0*	−64.5*	−64.3*	20.4	10.4
BW_{4e}	−89.7*	−87.7*	−88.6*	−90.1*	−40.1*

Note: Errors marked with asterisk (*) are unacceptable by the difference limen criterion.

[2] We have also studied these six AR spectrum estimation methods for pitch-asynchronous analysis (using speech segments of durations more than two times the pitch period). All these methods are found to result in comparable performance except for the BURGE method which gives more inferior performance than the other methods.

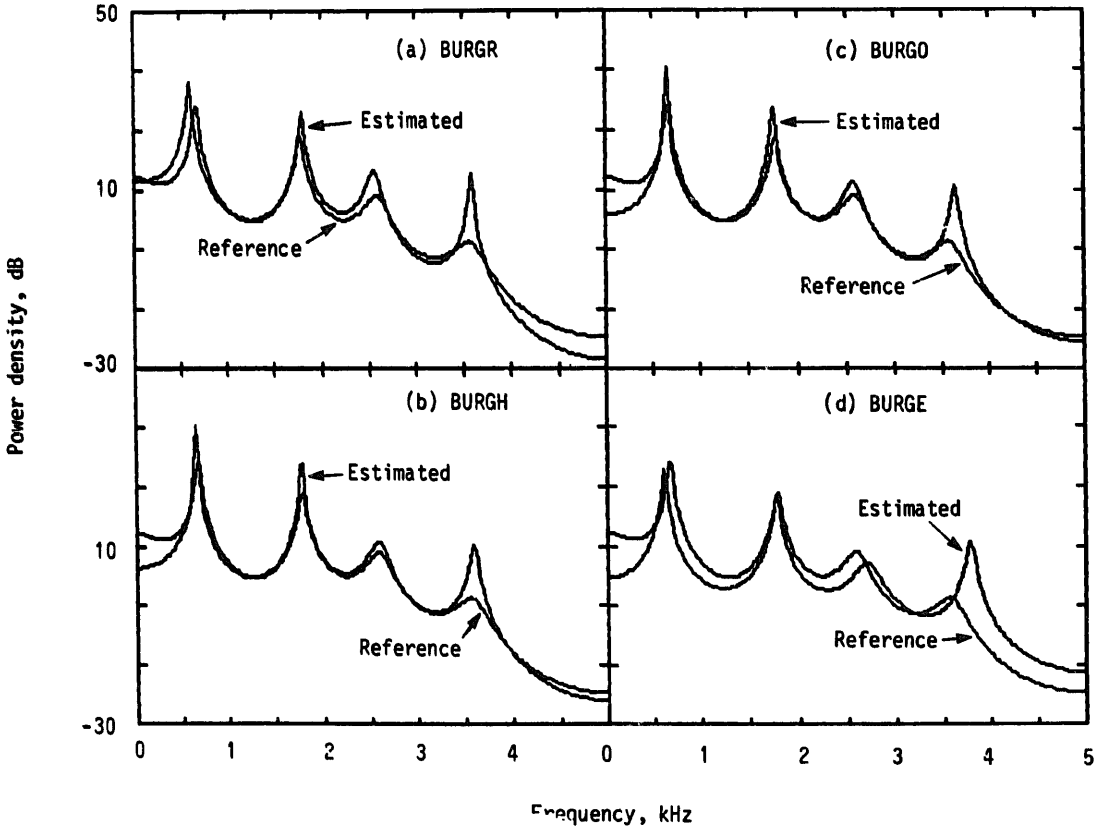

Fig. 3. Pitch-synchronous spectral estimates of the real speech signal of vowel /æ/ along with its reference power spectrum. (a) BURGR method, (b) BURGH method, (c) BURGO method and (d) BURGE method.

Table 4
Normalized errors (in %) in estimating the formant frequencies and bandwidths for pitch-synchronous analysis of the real speech signal of vowel /æ/. Pitch period is 7.9 ms.

Error	BURGR	BURGH	BURGO	BURGE	AUTO
F_{1e}	−11.2*	−3.1	−2.3	−9.6*	1.1
F_{2e}	1.2	−0.2	−1.4	−0.6	0.5
F_{3e}	−1.2	−0.1	−0.5	4.5	−*
F_{4e}	0.1	0.3	1.5	5.6*	1.6
BW_{1e}	−37.2	−80.3*	−80.5*	−19.1	96.8*
BW_{2e}	−41.1*	−58.5*	−56.8*	−11.0	221.3*
BW_{3e}	−46.5*	−21.5	−28.9	8.8	−*
BW_{4e}	−87.7*	−77.0*	−79.4*	−75.3*	15.3

Note: Errors marked with asterisk (*) are unacceptable by the difference limen criterion. The third formant is not resolved by the AUTO method.

4. The BURGO method with formant bandwidth expansion

In the previous section, we have seen that the BURGO method improves the performance of the BURGR method for pitch-synchronous analysis of voiced speech, but it still has the drawback that it underestimates the formant bandwidths. In order to get an accurate estimate of the power spectrum, it is therefore necessary to expand the formant bandwidths.

A simple method of expanding the formant bandwidths is to evaluate the estimated AR model transfer function $H(z)$ at a circle of radius greater than one. This can be achieved by simply multiplying the estimated linear predictor coefficients, a_k, $k = 1, 2, \ldots, M$, by r^{-k}, where $r > 1$ for formant bandwidth expansion. We show in Fig. 4 the power spectrum estimated by the BURGO method after formant bandwidth expansion for the synthetic signal of Example 1. Here, r is equal to 1.02 which corresponds to an increase in the bandwidths of all the four formants by 63 Hz at the sampling frequency of 10 kHz. By comparing this figure with Fig. 1(c), we see that the estimated power spectrum matches the actual power spectrum better after bandwidth expansion. The value of spectrum estimation error (E_s) after the formant bandwidth expansion is 0.08037 which is much smaller than 0.23076, the value of E_s before bandwidth expansion. When we compare the power spectra estimated by the BURGO method after formant bandwidth expansion (Fig. 4), the AUTO method (Fig. 1(e)) and the COVA method (fig. 1(f)), we find that the BURGO method performs much better than the AUTO method (the value of E_s for the AUTO method is 0.24154), but it still performs worse than the COVA method (the value of E_s for COVA method is 0.00004). The power spectra estimated

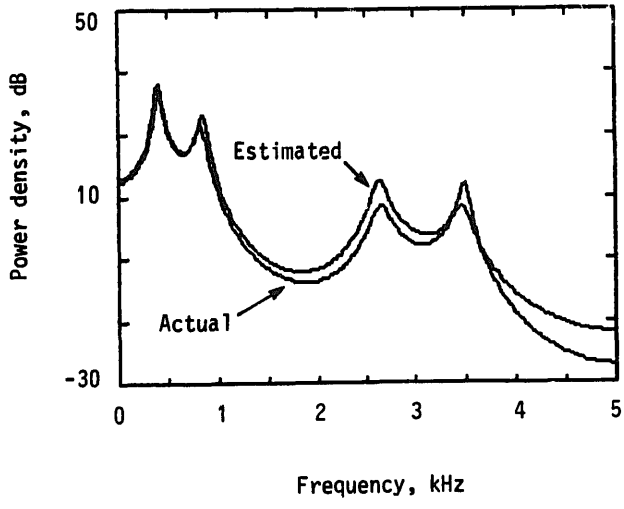

Fig. 4. Pitch-synchronous spectral estimate of the synthetic speech signal of vowel /o/ along with its original power spectrum. The BURGO method with formant bandwidth expansion is used for spectrum estimation.

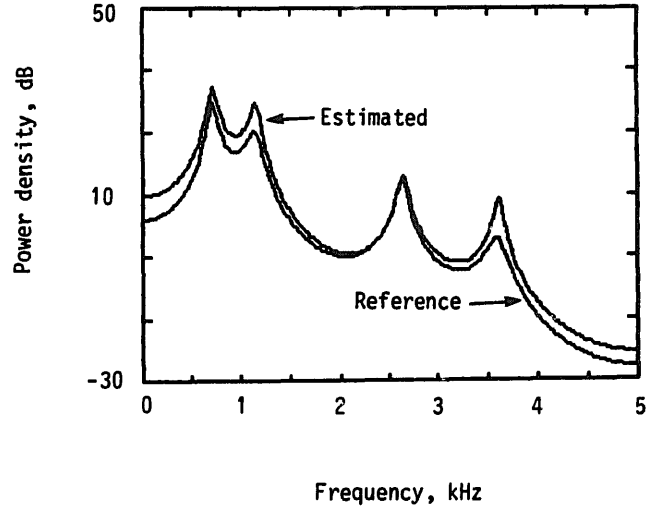

Fig. 5. Pitch-synchronous spectral estimate of the real speech signal of vowel /a/ along with its reference power spectrum. The BURGO method with formant bandwidth expansion is used for spectrum estimation.

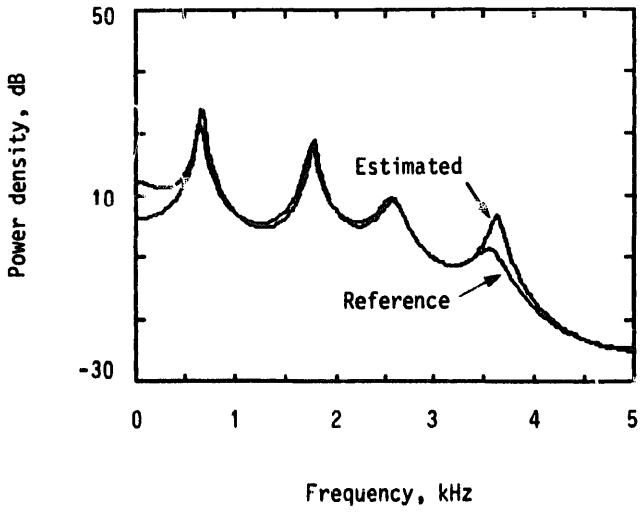

Fig. 6. Pitch-synchronous spectral estimate of the real speech signal of vowel /æ/ along with its reference power spectrum. The BURGO method with formant bandwidth expansion is used for spectrum estimation.

by using the BURGO method (after formant bandwidth expansion) are shown in Figures 5 and 6 for the real speech signals of Examples 2 and 3, respectively. These figures also reveal similar results.

The method, described above, for formant bandwidth expansion is computationally very efficient and can be applied to the PARCOR speech synthesizer by merely adding a multiplier to the lattice filter [25]. But, it has the drawback that it increases the bandwidths of all the four formants by the same amount. It might be desirable sometimes to expand the bandwidths differently for different formants (say, for example, expansion proportional to the estimated value of the formant bandwidth). In that case this method can not be used and one has to resort to another method which involves the following steps: solve the denominator polynomial of $\hat{H}(z)$ for its roots, compute formant bandwidths, expand them explicitly by the desired amount and reconstruct the AR system transfer function from the resulting poles. This method is computationally expensive and has not been studied further in the present paper. However, it can be used if desired.

5. Conclusion

We have studied the performance of four different weighted Burg methods for pitch-synchronous analysis of voiced speech (using speech segment of duration less than one pitch period and analysing the signal over closed glottal portion within the pitch period). We have shown that the optimum tapered Burg (BURGO) method of Kaveh and Lippert gives the best performance among these methods. However, this method still has the problem of underestimating the formant bandwidth. A simple method of formant bandwidth expansion is used to solve partially this problem. We have shown that this helps in getting a spectral estimate which matches better with the original power spectrum.

We have also compared the performance of the BURGO method with that of the autocorrelation and covariance methods. It is shown that the BURGO method (with formant bandwidth expansion) results in better performance than the autocorrelation method, but its performance is poorer to that of the covariance method. All of these six methods are also studied for the speech signal corrupted by additive white Gaussian noise. The BURGO method is found to result in least variance (even less than the covariance method). Since the covariance method does not guarantee the stability of the estimated AR system while the BURGO method does it even for fixed-point computations, it can be considered to be a suitable alternative pitch-synchronous analysis method, specially for those speech processing applications where stability of estimated AR system is an important factor (for example, in speech analysis-synthesis applications).

References

[1] K.K. Paliwal and P.V.S. Rao, "On the performance of Burg's method of maximum entropy spectral analysis when applied to voiced speech", *Signal Processing*, Vol. 4, No. 1, Jan. 1982, pp. 59–63.

[2] A.H. Gray, Jr. and D.Y. Wong, "The Burg algorithm for LPC speech analysis-synthesis", *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-28, No. 6, Dec. 1980, pp. 609–615.

[3] L.R. Rabiner, B.S. Atal and M.R. Sambur, "LPC pre-

diction error—Analysis of its variation with the position of the analysis frame", *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-25, No. 5, Oct. 1977, pp. 434–442.

[4] A.K. Krishnamurthy, "Two channel (speech and EGG) analysis for formant tracking and glottal inverse filtering", *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, 1984, pp. 36.6.1–36.6.4.

[5] D.N. Swingler, "A modified Burg algorithm for maximum entropy spectral analysis", *Proc. IEEE*, Vol. 67, No. 9, Sep. 1979, pp. 1368–1369.

[6] M. Kaveh and G.A. Lippert, "An optimum tapered Burg method for linear prediction and spectral analysis", *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-31, No. 2, Apr. 1983, pp. 438–444.

[7] P.D. Scott and C.L. Nikias, "Energy-weighted linear predictive spectral estimation: A new method combining robustness and high resolution", *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-30, No. 2, Apr. 1982, pp. 287–293.

[8] W.Y. Chen and G.R. Stegen, "Experiments with maximum entropy power spectra of sinusoids", *J. Geophys. Res.*, Vol. 79, No. 20, July 1974, pp. 3019–3022.

[9] K.K. Paliwal, "Frequency errors in AR spectral estimation of sinusoids: A comparative performance evaluation of different modifications over the Burg method", to be presented at the IEEE Int. Conf. Computers, Systems and Signal Processing, Banglore, India, Dec. 1984.

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

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

[12] B.S. Atal, "Automatic recognition of speakers from their voices", *Proc. IEEE*, Vol. 64, No. 4, Apr. 1976, pp. 460–475.

[13] H.F. Silverman and N.R. Dixon, "A comparison of several speech-spectra classification methods", *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-24, No. 4, Aug. 1976, pp. 289–295.

[14] K.K. Paliwal and P.V.S. Rao, "Evaluation of various linear prediction parametric representations in vowel recognition", *Signal Processing*, Vol. 4, No. 4, July 1982, pp. 323–327.

[15] K.K. Paliwal, "Effect of preemphasis on vowel recognition performance", *Speech Communication*, Vol. 3, No. 1, Apr. 1984, pp. 101–106.

[16] K.K. Paliwal, "Effectiveness of different vowel sounds in automatic speaker identification", *J. of Phonetics*, Vol. 12, 1984, pp. 17–21.

[17] J.L. Flanagan, *Speech Analysis Synthesis and Perception*, Springer-Verlag, Berlin, 1972.

[18] K.K. Paliwal and P.V.S. Rao, "A modified autocorrelation method of linear prediction for pitch-synchronous analysis of voiced speech", *Signal Processing*, Vol. 3, No. 2, Apr. 1981, pp. 181–185.

[19] J. Van den Berg, "Myoelastic-aerodynamic theory of speech production", *J. Speech Hearing Res.*, Vol. 1, 1958, pp. 227–244.

[20] R.L. Millar, "Nature of the vocal cord wave", *J. Acoust. Soc. Am.*, Vol. 31, June 1959, pp. 667–677.

[21] M.V. Mathews, J.E. Millar and E.E. David, Jr., "Pitch synchronous analysis of voiced sounds", *J. Acoust. Soc. Am.*, Vol. 33, No. 2, Feb. 1961, pp. 179–186.

[22] A.E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels", *J. Acoust. Soc. Am.*, Vol. 49, No. 2, 1971, pp. 583–590.

[23] M.R. Matausek and V.S. Batalov, "A new approach to the determination of glottal waveform", *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-28, No. 6, Dec. 1980, pp. 616–622.

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

[25] F. Itakura, S. Saito, T. Koike, H. Sawabe and M. Nishikawa, "An audio response unit based on partial autocorrelation", *IEEE Trans. Commun. Technol.*, Vol. COM-20, Aug. 1972, pp. 792–797.