LPC Prediction Error—Analysis of Its Variation with the Position of the Analysis Frame

LAWRENCE R. RABINER, FELLOW, IEEE, BISHNU S. ATAL, MEMBER, IEEE, AND MARVIN R. SAMBUR

Abstract—The LPC prediction error provides one measure of the success of linear prediction analysis in modeling a speech signal. Although a great deal is known about the properties of the prediction error, relatively little has been published about its variation as a function of the position of the analysis frame. In this paper it is shown that a fairly substantial variation in the prediction error is obtained within a single frame (i.e., 10 ms), independent of the analysis method (i.e., the covariance, autocorrelation, or lattice method). The implication of this result is that standard methods of LPC analysis may be inadequate for some applications. This is because the error signal is generally uniformly sampled at a low rate (on the order of 100 Hz), and this can lead to aliased results because of the variation of the error signal within the frame. For applications such as word recognition with frame-to-frame distance calculations using the prediction error, the errors due to uniform sampling can accrue. For speech synthesis applications, the effect of uniform sampling of the error signal is a small, but noticeable roughness in the synthetic speech. Various techniques for reducing the intraframe variation of the prediction error are discussed.

I. Introduction

Although the class of linear prediction analysis methods [1]-[7] are generally well understood, there still remain several interesting problems concerning the ways in which linear prediction analyses are implemented. One of these open questions concerns the properties of the normalized LPC prediction-error signal. In particular, there have been few investigations into the variation of the prediction error as a function of the position of the analysis frame within a single stationary speech segment. This paper presents results on a fairly intensive investigation into this question.

II. LPC Error Signal

To set the problem in its proper perspective it is worthwhile reviewing the LPC analysis model. Fig. 1 shows a block diagram of an LPC analysis system. The input speech signal is denoted as $s(n)$. A pth order linear predictor operates on $s(n)$ to produce the estimate $\hat{s}(n)$ defined as

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{1}$$

where the a_k are chosen to minimize the mean-squared error between the actual $s(n)$ and the predicted value $\hat{s}(n)$. We define this difference as

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k). \tag{2}$$

Manuscript received November 16, 1976; revised March 22, 1977, and May 9, 1977.

The authors are with Bell Laboratories, Murray Hill, NJ 07974.

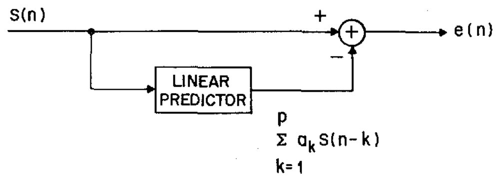

Fig. 1. Block diagram of LPC analysis system.

The a_k are thus chosen to minimize the mean-squared error E, defined as

$$E = \sum_{n=n_0}^{n_1} e^2(n) \tag{3}$$

where the limits in (3) are appropriately chosen. We shall denote the sequence of samples of $e(n)$ from $n = n_0$ to $n = n_1$ as the analysis frame. It can readily be shown that an equivalent expression for the mean-squared error is

$$E = -\sum_{i=0}^{p} \sum_{j=0}^{p} a_i c_{ij} a_j \tag{4}$$

where $a_0 = -1$, and c_{ij} is the autocorrelation of the signal, defined as

$$c_{ij} = \sum_{n=n_0}^{n_1} s(n-i)\, s(n-j). \tag{5}$$

Since the predictor coefficients which minimize E satisfy

$$\sum_{j=1}^{p} a_j c_{ij} = c_{0i} \tag{6}$$

(4) is further simplified to the form

$$E = -\sum_{i=0}^{p} a_i c_{0i}. \tag{7}$$

The normalized LPC error E_N is defined by normalizing E by the signal energy, i.e.,

$$E_N = \frac{\sum_{n=n_0}^{n_1} e^2(n)}{\sum_{n=n_0}^{n_1} s^2(n)} \tag{8}$$

or

$$E_N = 1 - \sum_{i=1}^{p} a_i \frac{c_{0i}}{c_{00}}. \tag{9}$$

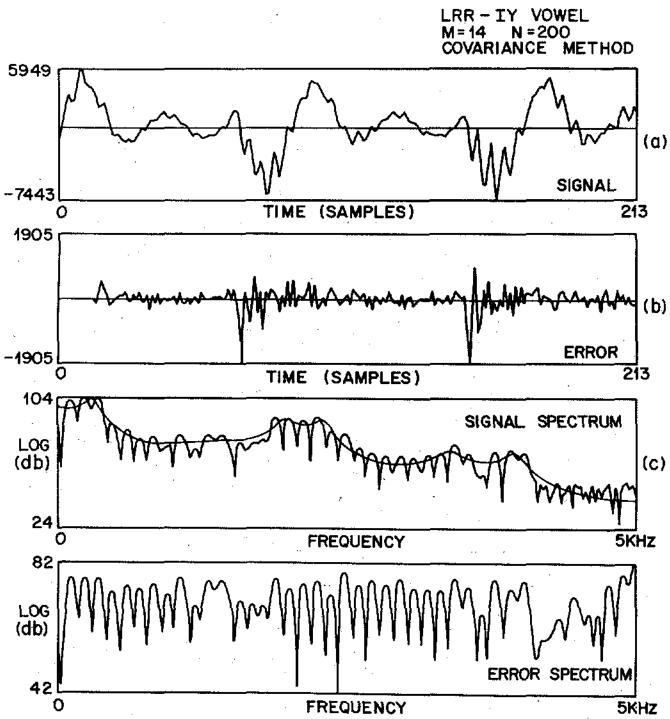

Fig. 2. Typical signals and spectra for LPC covariance method for a male speaker.

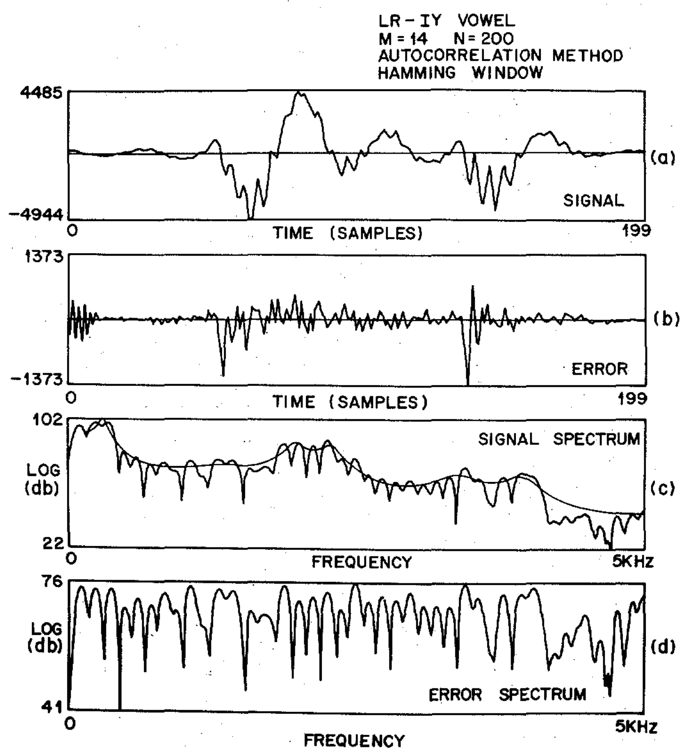

Fig. 3. Typical signals and spectra for LPC autocorrelation method for a male speaker.

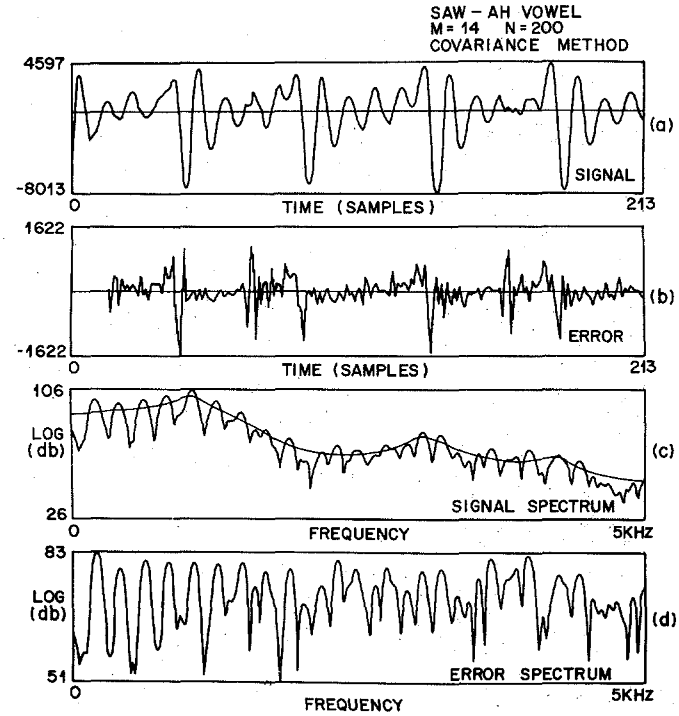

Fig. 4. Typical signals and spectra for LPC covariance method for a female speaker.

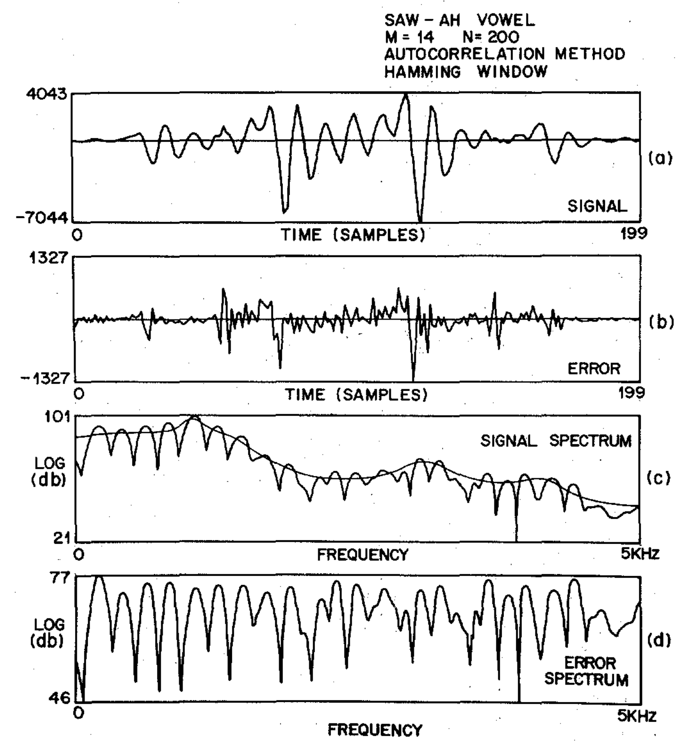

Fig. 5. Typical signals and spectra for LPC autocorrelation method for a female speaker.

Figs. 2–5 illustrate some typical analysis frames for two of the well-known methods of LPC analysis; i.e., the covariance method and the autocorrelation method.[1] Part (a) of each figure shows the signal samples used in the analysis;[2] part (b) shows the resulting error signal $(e(n))$; part (c) shows the signal spectrum (as obtained using an FFT analysis) as well as

[1]Results are not presented for the lattice method [3]–[7] since they are almost identical to those obtained for the covariance method.

[2]These figures show all samples used in the analysis. For the covariance method the first p samples are not used in computing the prediction residual.

the LPC model spectrum (the solid curve obtained directly from LPC analysis); and part (d) shows the spectrum of the error signal. The analysis-frame size for all four examples is 200 samples (20 ms at a 10-kHz rate). Fig. 2 shows the results of the covariance method with a p of 14.[3] The first (p-1) samples of $e(n)$ are not shown on the plot since they are not defined in the covariance method. The signal $e(n)$ distinctly shows the property that the error signal gets quite peaked at the beginning of each pitch period. Furthermore, the peaks in the error signal and the speech do not occur at the same instant of time. The spectral envelope of $e(n)$ is reasonably white, a property that is also evident in the quality of the spectral match of the LPC model to the signal spectrum.

Fig. 3 shows similar results for the autocorrelation method. The autocorrelation method used a 20-ms Hamming window ($w_H(n)$) to give the signal

$$s_w(n) = s(n) \cdot w_H(n) \tag{10}$$

for all values of n since $w_H(n)$ is 0 for $n < n_0$ and for $n > n_1$. Thus in Fig. 3 the signal shown is $s_w(n)$ and the error is

$$e_w(n) = s_w(n) - \sum_{k=1}^{p} a_k s_w(n-k). \tag{11}$$

Since $s_w(n)$ is 0 outside the range $n_0 \leqslant n \leqslant n_1$, a fairly substantial component of $e(n)$ occurs for the first p samples since the method is trying to predict the windowed signal from zero-valued samples. Again $e_w(n)$ is quite peaked at the beginning of each pitch period. However, the position of the Hamming window relative to the beginning of the pitch has a marked influence on the amplitudes of the peaks in the error signal.

Figs. 4 and 5 show similar results for a female speaker for the covariance method (Fig. 4) and the autocorrelation method (Fig. 5). The 20-ms frame size, for this speaker, was large enough to contain more than 4 pitch periods. The substantial variation is seen in the error signal $e(n)$ [part (b) of each figure] at the beginning of each pitch period.

The normalized LPC error signal (or the prediction error) itself is important for a variety of applications of LPC analysis systems [8]-[10]. Thus it is necessary to understand its properties. Previous work has been carried out in studying the variation of the prediction error as a function of the analysis frame size, and the number of analysis parameters [1], [4], [5], [11]; however, little has been published on the variation of the prediction error as a function of the analysis-frame position. Kang [8] has studied this problem for an LPC vocoder implementation and, using techniques similar to those discussed in this paper, selected points of minimum prediction error for transmission.

In this paper we show that substantial variation in the prediction error can occur within an analysis frame for all three LPC methods. Results are presented that illustrate the effects of windows (for the autocorrelation method) and frame size on the amount of variability within a frame. An analysis of the mechanisms that cause this intraframe variation of the prediction error is given, and techniques for reducing the magnitude of the variations are discussed.

[3]In this and subsequent figures the quantity p is called M.

III. Variation of the Prediction Error

The results given in the previous section (Figs. 2-5) show that $e(n)$, the error signal, is generally quite peaked at the beginning of each pitch period and that these peaks are not coincident with the speech-signal peaks. Furthermore, for the autocorrelation method, $e(n)$ is also quite variable for the first p samples (although the effect of the window is to attenuate this component of the error signal). These results suggest that the prediction error E_N may also be quite variable, depending on the position of the analysis frame with respect to the signal.

To investigate this effect, an LPC analysis system was implemented in which each new frame was displaced by a single sample from the previous frame. Thus the prediction (E_N) could be computed at the signal sampling rate (i.e., 10 kHz) and any variation within a frame could be readily seen. Four voiced utterances were studied in this investigation.[4] These included steady-state regions of the vowels i (beet) and a (father) spoken by a male (LRR) and a female (SAW) speaker.

Some typical results of the analysis are presented in Figs. 6-9. Fig. 6 shows a series of plots of the signal energy [Fig. 6(a)]

$$E_s = \sum_{n=n_0}^{n_1} s^2(n) \tag{12}$$

and the prediction error E_N for a 14-pole ($M = 14$) analysis, with a 20-ms ($N = 200$) frame size, for the covariance method [Fig. 6(b)], the autocorrelation method using a Hamming window [Fig. 6(c)], and the autocorrelation method using a rectangular window [Fig. 6(d)].[5] The average pitch period for this speaker was 84 samples (8.4 ms); thus about $2\frac{1}{2}$ pitch periods were contained within the 20-ms frame. For the covariance analysis [Fig. 6(b)], the prediction error shows a substantial variation with the position of the analysis frame. This effect is essentially due to the large peaks in the error signal $e(n)$ at the beginning of each pitch period. Thus whenever the analysis frame is positioned to encompass three sets of these peaks, the prediction error is much larger than when only two sets of these peaks are included in the analysis interval. Thus, as seen in Fig. 6(b), the prediction error shows a fairly large discrete jump in level as each new error peak is included in the analysis frame, followed by a gradual tapering off and flattening of the normalized error. The exact detailed behavior of the prediction error between discrete jumps depends on details of the signal and the analysis method.

Fig. 6(c) and (d) shows somewhat different behavior of the prediction error for the autocorrelation analysis method using a Hamming window [Fig. 6(c)] and a rectangular window [Fig. 6(d)]. As seen in these figures, the prediction error shows a substantial amount of high-frequency variation, as well as a small amount of low-frequency and pitch-synchronous variation. The high-frequency variation is due primarily to the error signal for the first p samples in which the signal is

[4]For unvoiced speech the variations in prediction error with the position of the analysis frame are primarily statistical in nature and are therefore not discussed here.

[5]Results for the lattice method are identical (when plotted) to those of the covariance method for all results to be presented in this paper. Thus such plots are not included herein.

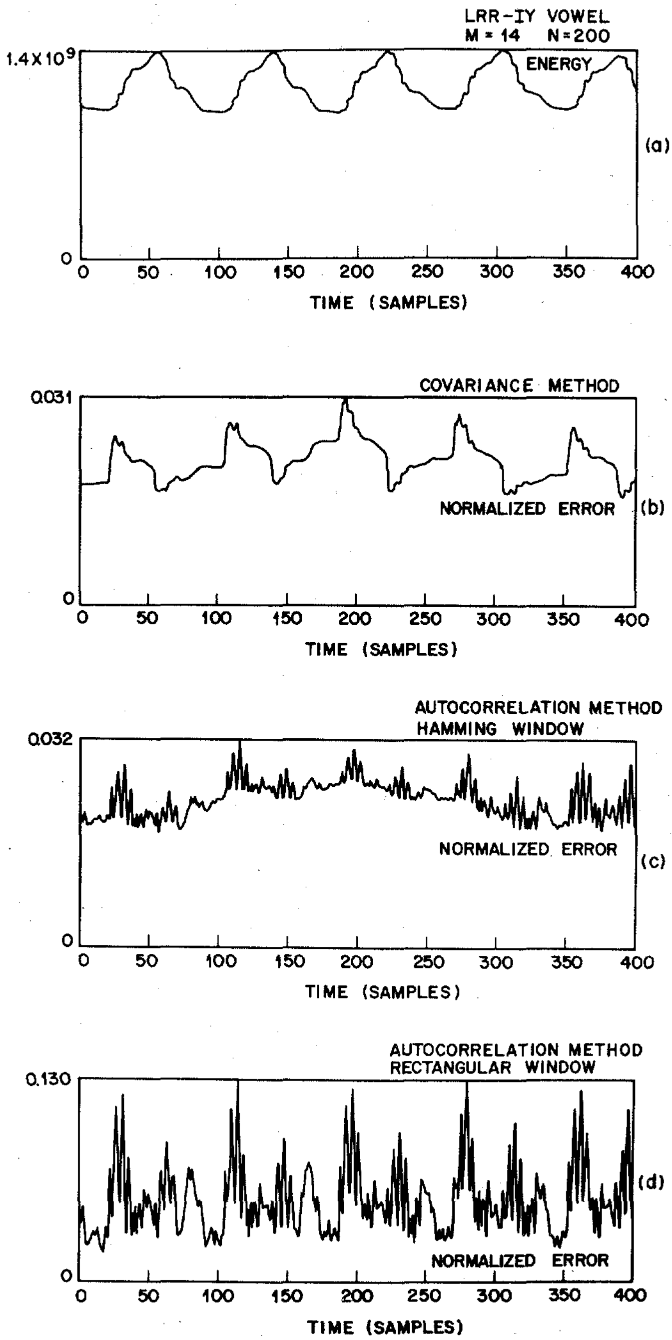

Fig. 6. Prediction-error sequences for 200 samples of speech for three LPC systems.

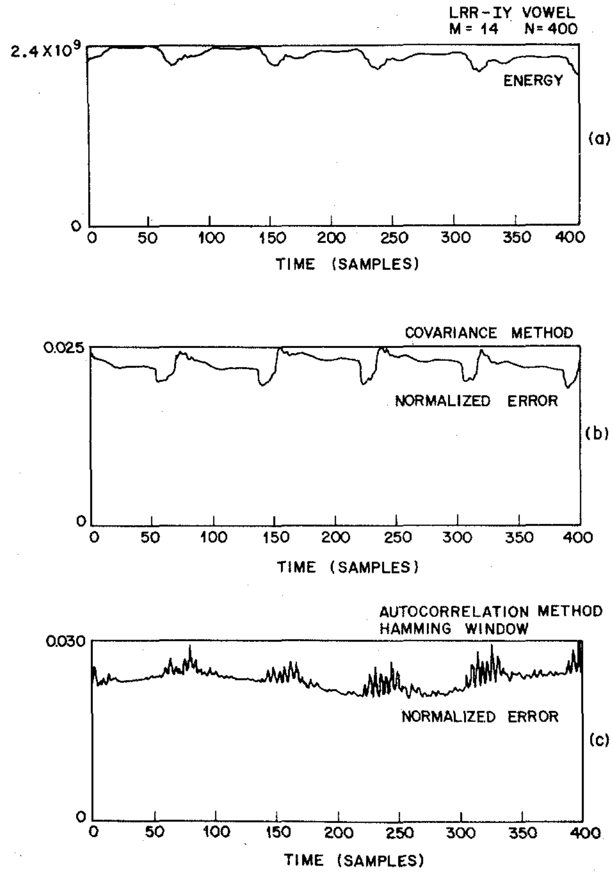

Fig. 7. Prediction-error sequences for $N = 400$ sample frames.

not linearly predictable. The magnitude of this variation is considerably smaller for the analysis using the Hamming window (30-percent variation) than for the analysis with the rectangular window (about 500-percent variation) due to the tapering of the Hamming window at the ends of the analysis window. Another component of the high-frequency variation of the prediction error is related to the position of the analysis frame with respect to pitch pulses, as discussed previously for the covariance method. However, this component of the error is much less a factor for the autocorrelation analysis than for the covariance method; especially in the case when a Hamming window is used, since new pitch pulses that enter the analysis frame are tapered by the window.

Fig. 7 illustrates the effect of using a longer analysis frame (40 ms) on the prediction error for both the covariance method [Fig. 7(b)] and the autocorrelation method using a Hamming window [Fig. 7(c)]. Similar effects to those obtained in Fig. 6 are obtained. However, the magnitude of the variation in the prediction error is smaller. For the covariance method [Fig. 7(b)] this is true since the frame always contains at least four complete pitch periods, and most of the time five pitch periods are contained within the frame. Thus the prediction error is smaller only when the analysis frame is positioned so that only four error peaks are included within the frame. For the autocorrelation method [Fig. 7(c)] the contribution of the first p (14) error terms to the prediction error is a smaller total percentage of the overall error for a 40-ms frame than for a 20-ms frame. Thus again the variation in the high-frequency component of the error is reduced.

Fig. 8 illustrates the results obtained for two cases of a pitch-synchronous analysis using the covariance method. Fig. 8(a) and (b) shows the signal energy and prediction error using a single period (8.4 ms); whereas, Fig. 8(c) and (d) shows the results for an analysis frame which is two periods (16.9-ms) long. It is seen in these figures that even though the prediction error shows long flat regions (i.e., little or no variation with the analysis-frame position) for both pitch-synchronous analyses, even slight changes in the pitch period lead to sharp discontinuities in the prediction error. These results strongly suggest that the sensitivity of the prediction error to small errors in choosing the pitch period for pitch-synchronous

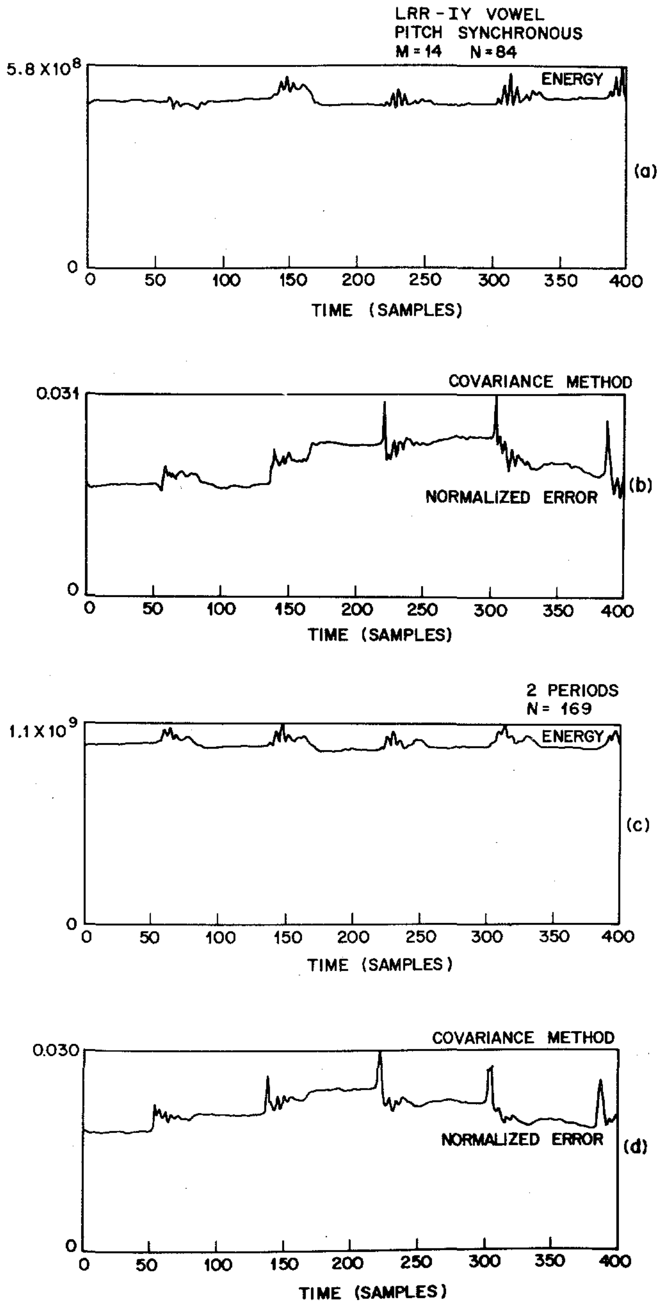

Fig. 8. Pitch-synchronous analysis results for single- and double-period analysis.

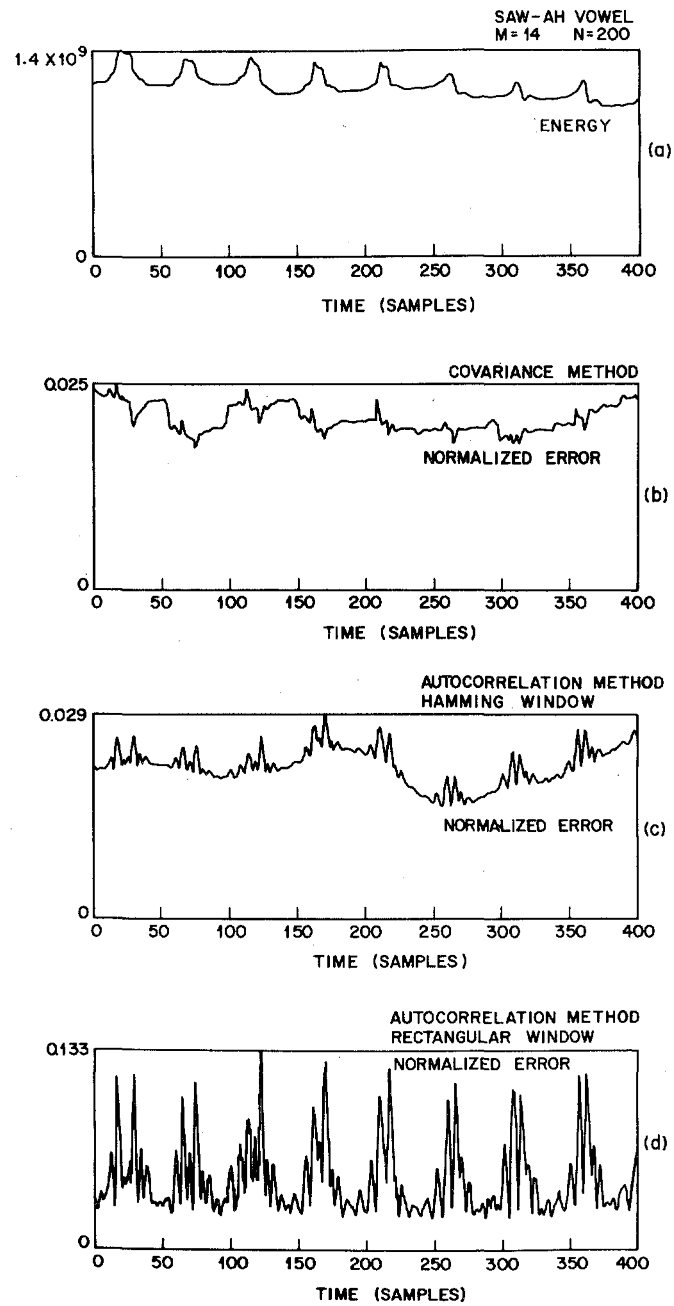

Fig. 9. Prediction-error sequences for speech from a female speaker.

analysis makes this type of analysis unattractive in terms of reducing or eliminating the variability of the prediction error with the analysis-frame position.

Finally, Fig. 9 shows the results obtained on the female speaker for a 20-ms analysis frame for the covariance method [Fig. 9(b)], a Hamming window autocorrelation method [Fig. 9(c)], and a rectangular window autocorrelation method [Fig. 9(d)]. Since the pitch period was much shorter for this speaker than for the previous speaker, and because the error signal itself was much less peaked at the beginning of each pitch period, the prediction error for the covariance analysis [Fig. 9(b)] showed less variation than in the previous examples. However, the results for the autocorrelation analysis [Fig. 9(c) and (d)] were similar since the dominant effect was still the high-frequency error for the first p analysis samples due to the poor prediction.

In summary, the results presented in this section have shown the following.

1) For the covariance method, the prediction error varies with the position of the analysis frame due to the presence, or absence, of the error peaks associated with the beginning of each pitch period. The magnitude of this variation depends on the analysis-frame size (i.e., the number of whole pitch periods nominally contained within the frame) and the degree of peaking of the error signal at the beginning of each pitch period.

2) For the autocorrelation method, the major component of the variation of the prediction error is due to the error in predicting the signal at the beginning of the analysis frame; i.e., the first p samples of the error. Thus a Hamming window serves to substantially reduce the variation in the error due to this effect. An additional component of the variation of the

prediction error is the number of whole pitch periods contained within the analysis frame.

It is important to be able to place some perspective on the importance and implications of these results. The prediction error is a measure of the success with which a frame of speech can be linearly predicted. Thus variations of from 30 to 500 percent in the prediction error due to slight shifts of the analysis-frame position means that the prediction error must be carefully interpreted and carefully used in any application on which it is based. One such application is the computation of distance proposed by Itakura [10] based on a ratio of prediction errors. In this application the quantity $D(a, \hat{a})$, defined as

$$D(a,\hat{a}) = \ln\left(\frac{aVa^t}{\hat{a}V\hat{a}^t}\right) = \ln\left(\frac{E_1}{E_2}\right) \tag{13}$$

is computed in which V is a correlation matrix and the terms $E_2 = \hat{a}V\hat{a}^t$ and $E_1 = aVa^t$ are prediction residuals or normalized LPC errors based on using the LPC sets $\hat{a}$ and a. What we have shown is that, depending on the analysis method and the frame position, variations of E_2 on the order of 30–500 percent can be obtained. The implications of this variation in defining LPC distance are discussed in Section V.

A second case in which variation of the prediction error with analysis-frame position may be important is the LPC vocoder. It can be argued that analysis frames with the smallest prediction error may lead to the best estimates of pole center frequencies and bandwidths since the peaks in the error signal due to the pitch pulses essentially make the analysis more noisy and subject to numerical errors. To test this conjecture LPC synthesis results were obtained using frames with both the highest and lowest prediction errors and the resultant synthetic utterances were compared. We discuss these comparisons in Section VI.

IV. Methods of Reducing the Variability of the Prediction Error

In order to reduce the variability in the LPC prediction error with the position of the analysis frame, two distinct preprocessing methods were investigated. These were as follows.

1) Allpass filtering of $s(n)$ to reduce the peakedness of $e(n)$ at the beginning of each pitch period; i.e., to spread out the error pulse in $e(n)$.

2) Preemphasizing $s(n)$ by a first-order network to reduce the effects of the high-frequency error in $e(n)$ at the beginning of the frame by making the magnitude of $e(n)$ larger throughout the frame.

Results of using these two techniques are presented in Figs. 10 and 11. Fig. 10 shows the effect on the prediction error of using a 24th-order allpass filter[6] [12] to spread out (i.e., phase disperse) the signal for the covariance method [Fig. 10(b)] and for the autocorrelation method using a Hamming window [Fig. 10(c)]. By contrasting these results with those previously shown in Fig. 6, the following effects are noted.

1) For the covariance method the allpass filter effectively spreads out the sharp changes in the prediction error signal. A pitch-synchronous variation in the prediction error representing low-frequency (gradual) changes is now the dominant effect of varying the position of the analysis frame. In addition, a small noiselike component rides on this low-frequency signal. The noise is due primarily to the detailed shape of the error signal from the linear prediction analysis.

2) For the autocorrelation method the effect of the allpass filtering is essentially negligible because the dominant error terms were shown to be related to the analysis method rather than details of the error signal itself. Thus a substantial high-frequency variation of the prediction error remains after the allpass filter is applied.

Fig. 11 shows the results obtained for the autocorrelation method when a first-order preemphasis network of the form

$$H(z) = 1 - \alpha z^{-1} \tag{14}$$

with $\alpha = 0.95$ is applied to the speech signal prior to the LPC analysis. Since the major effect of the preemphasis network is to reduce the spectral variations in the input speech signal, the prediction error increases in value quite significantly over the values obtained without the preemphasizer. The effect of this increased normalized error is to essentially swamp out the high-frequency variation in the prediction error due to the first p samples since the error signal is uniformly higher throughout the analysis interval. Thus, as seen in Fig. 11(b), the prediction error shows considerably smaller variation with respect to the position of the analysis frame when the equalizer is used than when it is not used.

We did not study the effect of preemphasis for the covariance method. The results for the covariance method cannot be significantly different from those of the autocorrelation method. Therefore the reasons previously stated for the reduction in variability of the prediction error for the autocorrelation method apply equally well to the covariance method; in both cases, the large prediction error resulting from the preemphasis of the speech signal will effectively swamp the smaller frame-dependent variations of the prediction error.

The conclusion of this section is that signal conditioning techniques provide effective methods of reducing the variation in the LPC prediction error with the position of the analysis frame. It should be noted that the signal conditioning techniques discussed in this section were independent of the signal. Signal-dependent methods for reducing the variability of the prediction error, such as adjusting the position and size of the frame based on the pitch period, can also be applied, but are much more sensitive to accurate determination of pitch.

V. Effects of Prediction-Error Variation on an LPC Distance Metric

Itakura [10] has proposed a measure of similarity between speech frames with measured LPC sets of a and $\hat{a}$ as

$$D(a,\hat{a}) = \ln\left(\frac{aVa^t}{\hat{a}V\hat{a}^t}\right) = \ln\left(\frac{E_1}{E_2}\right)$$

i.e., the distance between frames is related to the ratio of prediction residuals or normalized LPC errors. (Other LPC

[6] For this implementation the signal was processed by a cascade of three of the eight-order allpass filters of [9]. The detailed nature of the allpass filter is not critical. It is important that the effective duration of the impulse response of the allpass filter be large enough to spread out the excitation over the pitch period, but not so large so as to smear the temporal variations of the predictor coefficients.

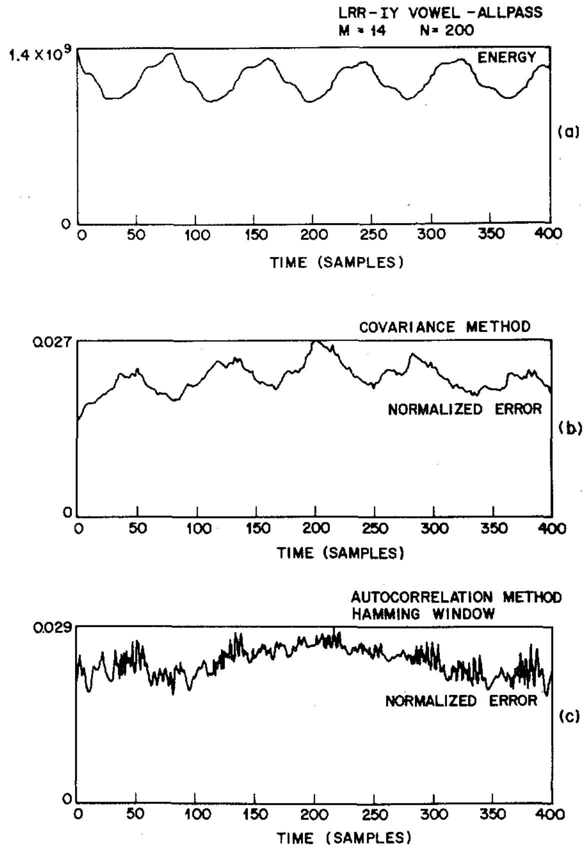

Fig. 10. The effects of allpass filtering on the prediction-error sequence.

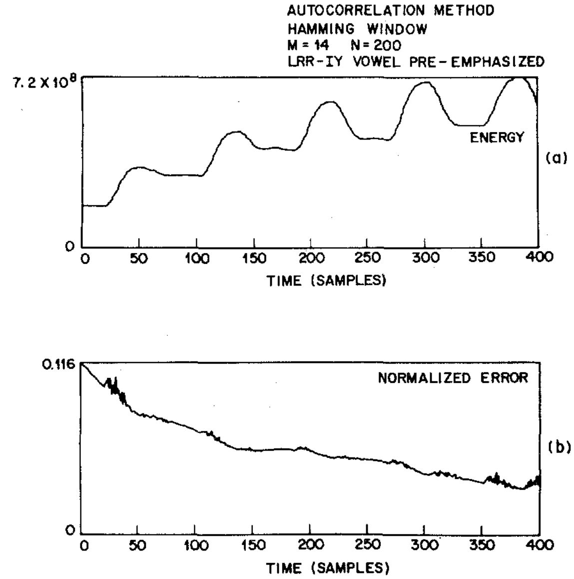

Fig. 11. The effects of preemphasis on the prediction-error sequence.

distance metrics have been proposed by Gray and Markel [13].) A fundamental implication of the previously given distance measure is that if a and $\hat{a}$ are basically from the same frame, then $D(a, \hat{a})$ should be essentially 0, or $E_1 \approx E_2$. Conversely, if a and $\hat{a}$ are from dissimilar frames, then $D(a, \hat{a})$ should be large. We have already shown that if a and $\hat{a}$ represent LPC sets obtained from positioning the analysis frame in slightly different positions, then the prediction error can vary by 30 percent or more by itself. This result, by itself, does not mean that $D(a, \hat{a})$ will vary by 30 percent or more, since E_1 and E_2 are not both minimum prediction errors, but instead E_1 is the error obtained when parameter set a is used with the correlation matrix V obtained from parameter set $\hat{a}$. However, this result suggests that significant variations in $D(a, \hat{a})$ might also exist because of variations in the prediction-error term E_2 due to the position of the analysis frame.

To test this hypothesis an LPC analysis was carried out on a sample-by-sample basis for a 1.75-s utterance. The autocorrelation method was used with a 20-ms frame size using a Hamming window. The analysis rate was nominally set at 100 frames/s. Thus, within each 10-ms frame, 100 LPC analyses were performed, thereby giving the prediction error at a 10 000-Hz rate. The positions at which the peak prediction error and the minimum prediction error occurred were obtained. From the LPC sets corresponding to these two positions ($a_{\max}$ and $a_{\min}$) the following quantities were computed:

$$D_1(a_{\min}, a_{\max}) = \frac{a_{\min} V_{\max} a_{\min}^t}{a_{\max} V_{\max} a_{\max}^t} \tag{15}$$

and

$$D_2(a_{\max}, a_{\min}) = \frac{a_{\max} V_{\min} a_{\max}^t}{a_{\min} V_{\min} a_{\min}^t}. \tag{16}$$

Since both sets of a's were from ostensibly the same analysis frame, one would theoretically expect D_1 and D_2 to be close to 1.0 for all frames. A total of 175 calculations of D_1 and D_2 were made. Fig. 12(a) shows the distribution of the occurrences of values of D_1, D_2 and the combined set of D_1 and D_2. Of these 175 calculations, D_1 exceeded a threshold of 1.2 a total of 97 times, and D_2 exceeded this threshold a total of 85 times; i.e., more than 50 percent of the frames were classified as dissimilar to shifted versions of the frames using a 20-percent variation threshold. Additionally, in many cases, the distances exceeded the 20-percent threshold by a considerable margin.

The previous experiment was repeated on the same sentence after it had been preemphasized using the network discussed in the previous section. The results for this case are shown in Fig. 12(b). For the 175 frames tested, D_1 exceeded the threshold of 1.2 a total of 34 times, and D_2 exceeded this threshold a total of 31 times. Of these 67 cases, the majority occurred in regions where large signal changes were occurring; i.e., in speech transitions, etc., where such behavior is entirely anticipated. These results indicate that signal preconditioning is a useful technique when a distance measure of the type previously described is to be used to compare LPC parameter sets.[7]

The implication of the results presented in this section is that in the worst case when one compares LPC sets obtained from positioning the analysis frame so as to give the maximum and minimum prediction errors, then fairly large distances between such frames can be obtained. Since such frames are physically the same this result suggests that spuriously large

[7]Results are not given for the covariance method, since the LPC distance of (13) is almost always used with the autocorrelation method.

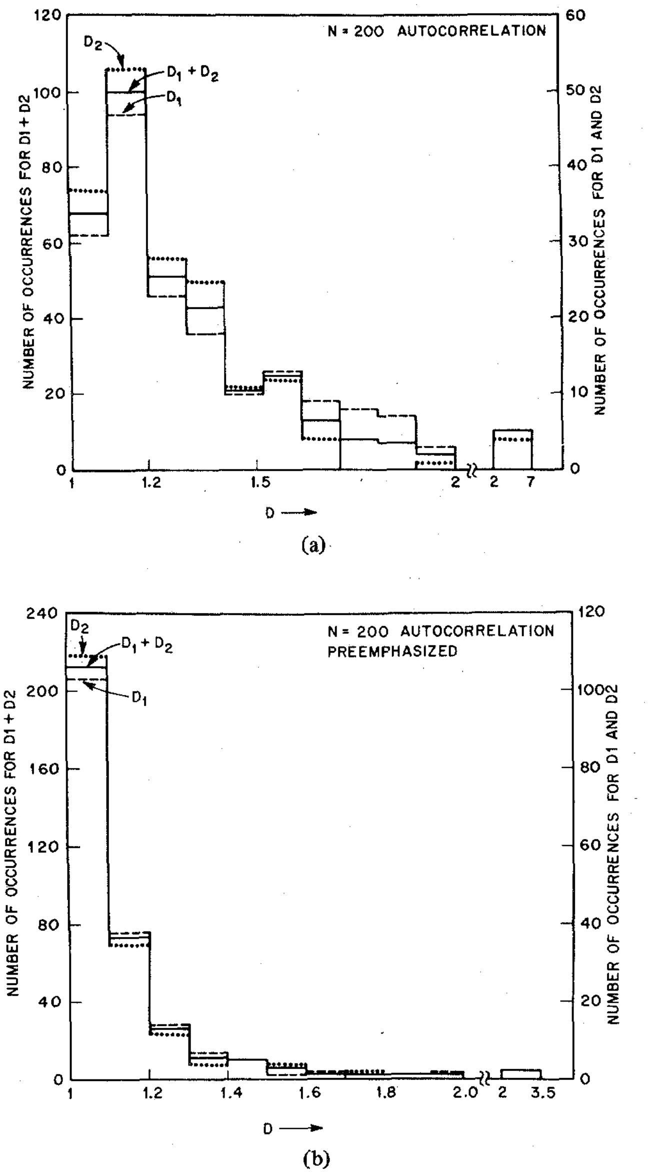

Fig. 12. The distributions of occurrences of D_1, D_2, and $D_1 + D_2$ for original speech, and preemphasized speech.

distances between similar frames can occur by random sampling due to the effects previously described. However, when averaging the distance measure over a large number of contingous frames, as would be the case for word recognition, etc., one would expect such statistical variations in distances to average out leaving a fairly reliable result. In addition, the use of a first-order preemphasis network can help to alleviate some of the statistical variations in distance.

VI. LPC Synthesis Effects of Prediction-Error Variations

Another application where the effects of frame position could be of importance is in the LPC vocoder. By examining the spectra of the individual frames, it was found that the differences in spectra between frame positions corresponding to maximum and minimum prediction errors primarily were in the bandwidths of the formants. The formant bandwidths corresponding to maximum prediction error were generally from 10 to 50 percent larger than the bandwidths corresponding to minimum prediction error. This effect was more pronounced for shorter analysis frames. The question now remained as to whether such changes in formant bandwidths could be perceived by a listener, and if so, what type of degradation would be heard.

To answer this question a standard LPC vocoder was implemented. The excitation parameters (i.e., pitch and voiced-unvoiced) were obtained from the semiautomatic procedure described in [14]. This effectively circumvents the interactions between problems in pitch detection and those that we wish to investigate here. Two sets of LPC coefficients were obtained for each frame, one set corresponding to the frame position with the maximum prediction error, the second set with the minimum prediction error. LPC synthesis was then performed using both sets of LPC parameters, and the two synthetic versions of each utterance were compared. A small but consistent difference in quality was obtained in informal listening tests. The major difference in quality was that the synthesis from the LPC coefficients with the maximum prediction error was somewhat more nasal-like than the synthesis from the coefficients with the minimum prediction error. This result is consistent with the observation that the LPC coefficients associated with the maximum prediction error tended to produce formants with broader bandwidths than those associated with the minimum prediction error.[8]

The informal listening experiments were done with LPC implementations of both the covariance and autocorrelation methods of analysis. The quality differences that were obtained were more pronounced for shorter analysis intervals than for longer analysis intervals. Additionally, the preprocessing functions discussed in Section IV were used on the utterances and the informal listening tests were repeated. For the covariance method, the use of the allpass filter reduced the differences in quality so that they were almost inaudible. For the autocorrelation method, the use of the preemphasis network had a similar effect. Thus in terms of synthesis it would seem reasonable to preprocess the signal prior to LPC analysis to eliminate the variations due to placement of the analysis interval within the frame.

VII. Summary

In this paper we have shown that, for LPC analysis, by suitable placement of the analysis frame, the LPC prediction error can vary significantly. An explanation for the effect in terms of the peakedness of the error signal at the beginning of each pitch period, and the error at the beginning of the analysis frame (for the autocorrelation method) was discussed. In addition two preprocessing methods for reducing the variation of the prediction error were introduced including allpass filtering and preemphasis of the speech. Finally, the physical implications of this prediction-error variation for distance calculations using the Itakura distance measure, and for speech synthesis based on LPC parameters were discussed.

[8]Kang [8] observed a noticeable flutter in the speech obtained from LPC analyses at a uniform rate.

REFERENCES

[1] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, pp. 637-655, 1971.

[2] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Japan*, vol. 53-A, pp. 36-43, 1970.

[3] ——, "Digital filtering techniques for speech analysis and synthesis," in *Proc. 7th Int. Congr. Acoustics* (Budapest, 1971), Paper 25C1.

[4] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. Berlin: Springer, 1976.

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

[6] J. Burg, "A new analysis technique for time series data," NATO Advanced Study Institute on Signal Processing, Enschede, The Netherlands, 1968.

[7] J. Makhoul, "New lattice methods for linear prediction," in *Proc. 1976 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1976, pp. 462-465.

[8] G. S. Kang, "Application of linear prediction encoding to a narrowband voice digitizer," Naval Res. Lab. Rep. 7774, Oct. 1974.

[9] G. S. Kang and D. C. Coulter, "600-bit-per-second voice digitizer (linear predictive formant vocoder)," Naval Res. Lab. Rep. 8043, Nov. 1976.

[10] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67-72, 1975.

[11] S. Chandra and W. C. Lin, "Experimental comparison between stationary and nonstationary formulations of linear prediction applied to speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 403-415, 1974.

[12] L. R. Rabiner and R. E. Crochiere, "On the design of all-pass signals with peak amplitude constraints," *Bell Syst. Tech. J.*, vol. 55, pp. 395-407, Apr. 1976.

[13] A. H. Gray and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 380-391, Oct. 1976.

[14] C. A. McGonegal, L. R. Rabiner, and A. E. Rosenberg, "A semi-automatic pitch detector (SAPD)," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 570-574, Dec. 1975.

A Necessary and Sufficient Condition for Quantization Errors to be Uniform and White

ANEKAL B. SRIPAD, STUDENT MEMBER, IEEE, AND DONALD L. SNYDER, MEMBER, IEEE

Abstract—**In this paper, a necessary and sufficient condition is given to model the output of a quantizer as an infinite-precision input and an additive, uniform, white noise. The statistical properties of the quantization error are studied, and a detailed analysis for Gaussian distributed inputs is given.**

I. INTRODUCTION

THE IMPLEMENTATION of filters with digital devices having finite word-length introduces unavoidable quantization errors. These effects have been widely studied [6]-[8]. There are three main sources of quantization error that can arise: input quantization, coefficient quantization, and quantization in arithmetic operations. Once a model to represent input quantization error has been developed, models to represent the other two types of error can easily be obtained [6]-[8]. Hence, our attention in this paper will be on input quantization, which occurs in the analog-to-digital conversion process.

A quantizer can be viewed as a nonlinear mapping from the domain of continuous-amplitude inputs onto one of a countable number of possible output levels. The analysis of errors introduced with this mapping can be approached either using nonlinear deterministic methods [9] or using stochastic methods [1]-[8]. The later approach is the one we adopt in this paper.

With the stochastic method, the output of the quantizer is modeled as an infinite-precision input and an additive noise. The additive noise is a random variable whose distribution is nonzero only over an interval equal to the quantization step size. Widrow [1] showed that under the condition that the input random variable has a certain band-limited characteristic function, the quantization noise is uniformly distributed; this is frequently referred to as the "Quantization Theorem" [1]-[5].[1] The band-limitedness assumption on the input random variable is a sufficient condition that is not satisfied universally. In this paper, a weaker sufficient condition which is also necessary for the Quantization Theorem is given. This expands the class of input distributions for which the uniform noise model to represent quantization errors can be used with confidence and is discussed in Section III.

Manuscript received September 17, 1976; revised February 14, 1977. This work was supported in part by the National Science Foundation under Grant ENG74-07800 and in part by the National Institutes of Health, Division of Research Resources, under Grant RR00396.

A. B. Sripad is with the Department of Systems Science and Mathematics, Washington University, St. Louis, MO 63130.

D. L. Snyder is with the Department of Electrical Engineering and the Biomedical Computer Laboratory, Washington University, St. Louis, MO 63130.

[1] More precisely, Widrow established that if the input random variable has a band-limited characteristic function, then the input distribution can be recovered from the quantized-output distribution, and the quantization noise density is uniform [1], [2]. Because of its dominant importance in applications, we are interested in the latter part of this result, and refer to it as the "Quantization Theorem" for brevity.