LEVEL II

(1)

AD A108350

Bolt Beranek and Newman Inc.
Report No. 2800

A.I. Report No. 14
April 1974

Quantization Properties of Transmission Parameters in Linear Predictive Systems

John Makhoul
R. Viswanathan

This research was supported by: Advanced Research Projects Agency under ARPA Order No. 1967; Contract No. DAHC-71-C-0088.

060 100

$$g = \log \frac{1+k}{1-k}$$

81 12 08 160

①

BBN Report No. 2800
A.I. Report No. 14

QUANTIZATION PROPERTIES OF TRANSMISSION PARAMETERS IN LINEAR PREDICTIVE SYSTEMS

John Makhoul
R. Viswanathan

April 1974

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Defense Supply Service - Washington under Contract No. DAHC-71-C-0088. Order No. 1967.

ABSTRACT

Several alternate sets of parameters that represent the linear predictor are investigated as transmission parameters for linear predictive speech compression systems. Although each of these sets provides equivalent information about the linear predictor, their properties under quantization are different. The results of a comparative study of the various parameter sets are reported. Specifically it is concluded that the reflection coefficients are the best set for use as transmission parameters. A more detailed investigation of the quantization properties of the reflection coefficients is then carried out using a spectral sensitivity measure. A method of optimally quantizing the reflection coefficients is also derived. Using this method it is demonstrated that logarithms of the ratios of the familiar area functions possess approximately optimal quantization properties. Also, a solution to the problem of bit allocation among the various parameters is presented, based on the sensitivity measure.

The use of another spectral sensitivity measure renders logarithms of the ratios of normalized errors associated with linear predictors of successive orders as the optimal quantization parameters. Informal listening tests indicate that the use of log area ratios for quantization leads to better synthesis than the use of log error ratios.

ACKNOWLEDGMENT

The authors wish to thank Brenda Aighes for her thoughtfulness and endurance during the preparation of this report.

TABLE OF CONTENTS

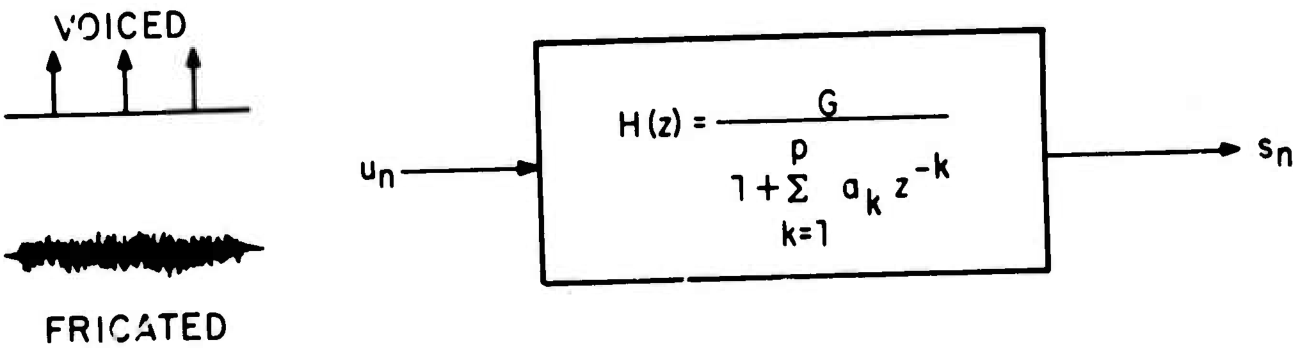
I. INTRODUCTION

In recent years the method of linear prediction has been quite successfully used in speech compression systems [1] - [5]. In this method, speech is modeled by an all-pole filter $H(z)$ as shown in Fig. 1. The input to the filter is either a sequence of pulses separated by the pitch period for voiced sounds, or white noise for fricated (or unvoiced) sounds. The parameters a_k, $1 \le k \le p$, are known as the predictor coefficients, and G is the filter gain. For a particular speech segment the filter parameters are obtained by passing the speech signal through the inverse filter $A(z)$ (as in Fig. 2) and then minimizing the total-squared prediction error

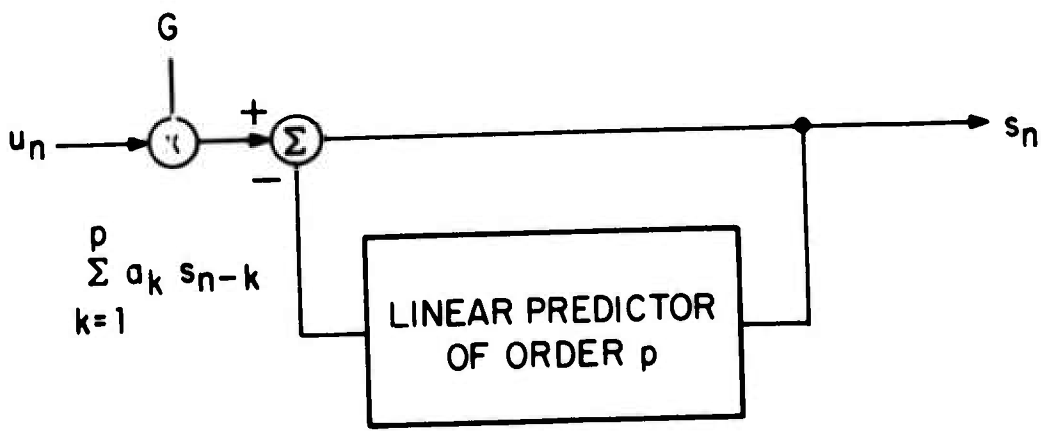
$$E = \sum_n e_n^2 = \sum_n \left(s_n + \sum_{k=1}^{p} a_k \, s_{n-k}\right)^2 \qquad (1)$$

with respect to a_k. If the signal s_n is assumed to be zero for $n<0$ and $n>N$ (e.g. by multiplying it by a finite window), the error minimization results in the set of equations

$$\sum_{k=1}^{p} a_k \, R_{|i-k|} = -R_i \; , \quad 1 \le i \le p \; , \qquad (2)$$



(a) FREQUENCY - DOMAIN MODEL



(b) TIME - DOMAIN MODEL

Fig. 1. Discrete model of speech production as employed in linear prediction.

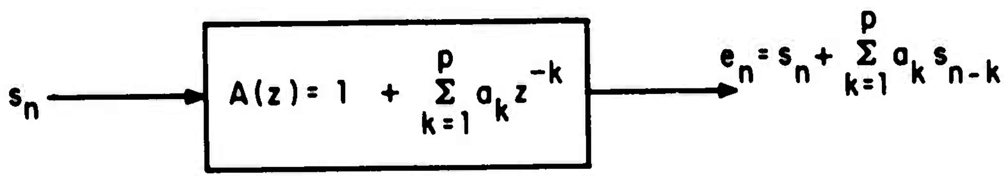


Fig. 2. The error sequence e_n as the output of an inverse filter A(z).

where

$$R_i = \sum_{n=0}^{N-|i|} s_n \, s_{n+|i|} \tag{3}$$

is the autocorrelation function of the signal s_n. The set of equations (2) can be recursively solved for the predictor coefficients a_k as follows:

$$E_0 = R_0 , \tag{4-a}$$

$$k_i = -(R_i + \sum_{j=1}^{i-1} a_j^{(i-1)} R_{i-j})/E_{i-1} , \tag{4-b}$$

$$a_i^{(i)} = k_i ,$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i \, a_{i-j}^{(i-1)} , \quad 1 \le j \le i-1 , \tag{4-c}$$

$$E_i = (1-k_i^2) \, E_{i-1} . \tag{4-d}$$

Equations (4-b,c,d) are solved recursively for $i=1,2,\ldots,p$. The final solution is given by

$$a_j = a_j^{(p)} , \quad 1 \le j \le p . \tag{4-e}$$

The filter H(z) with the predictor coefficients obtained from (4) is always stable, i.e. the poles of H(z) lie inside the unit circle in the z-plane. Since H(z) is an all-pole filter, stability also implies that H(z) is minimum phase.

The intermediate quantities k_i, $1 \le i \le p$, in (4) are called the reflection coefficients (or partial correlation coefficients [3,10]). Reflection coefficients occur naturally in the treatment of the vocal tract as an acoustic tube with p sections, each with a different cross-sectional area [2,9]. An important result that will be used in the sequel is that the conditions

$$-1 < k_i < 1 \quad , \quad 1 \le i \le p \quad , \tag{5}$$

are both necessary and sufficient for the stability of H(z).

The quantity E_p obtained from (4) is the minimum value of the prediction error given in (1). By expanding the squared terms in (1) and using (2), it can be shown that the minimum error is given by

$$E_p = R_o + \sum_{k=1}^{p} a_k R_k \quad . \tag{6}$$

Of interest also is the normalized error V_p which is the ratio of the minimum error to the energy of the input speech signal, i.e.

$$V_p = E_p / R_o \quad . \tag{7}$$

From (4-a), (4-d) and (7) we obtain

$$V_p = \prod_{j=1}^{p} (1-k_j^2) \quad . \tag{8}$$

The gain G of the filter H(z) is obtained by conserving the total energy between the speech signal and the impulse response of H(z). The gain can be shown to satisfy [6]

$$G^2 = E_p = R_0 V_p = R_0 + \sum_{k=1}^{p} a_k R_k \quad . \tag{9}$$

Equations (2), (3) and (9) completely specify the filter parameters. It can be shown that (for a well chosen p) the resulting linear prediction all-pole spectrum is a good match to the envelope of the signal spectrum [6].

Above we assumed that the speech signal was multiplied by a finite window. The shape of window is of importance if the signal spectrum is to approximate the transfer function of the vocal tract. This issue is discussed in detail elsewhere [7]. A smooth window such as the Hamming or Hanning window is adequate.

When applying the linear prediction method to speech compression, the model parameters - predictor coefficients, gain and pitch frequency for voiced sounds - have to be extracted, quantized and transmitted to the receiver. The rate of such parameter extraction is usually on the order of 50-100 Hz to follow the time-varying overall characteristics of the input speech signal. At the receiver, speech is reconstructed (or synthesized) using the speech production model given in Fig. 1.

The optimal choice and quantization of transmission parameters is of prime importance if the resulting synthesized speech is to be of good quality. In this paper, several alternate sets of transmission parameters are considered and their quantization properties are compared.* This comparative study has indicated that the reflection coefficients possess many desirable quantization properties. An optimal method of quantizing the reflection coefficients is derived using a spectral sensitivity measure. The sensitivity measure is also used for allocating a fixed number of bits among the various parameters in an optimal manner (in a minimax sense). Finally, the use of a second spectral sensitivity measure for the optimal quantization of the reflection coefficients is investigated.

---

*As the quantization properties of pitch and gain are well understood we have not considered them in this study.

II. ALTERNATE TRANSMISSION PARAMETER SETS

The all-pole model used in a linear predictive system has a transfer function

$$H(z) = \frac{G}{A(z)} = \sum_{n=0}^{\infty} h_n z^{-n} , \qquad (10)$$

where the inverse filter A(z) is given by

$$A(z) = 1 + \sum_{n=1}^{p} a_n z^{-n} . \qquad (11)$$

Given below is a list of possible sets of parameters for characterizing uniquely the linear prediction filter H(z):

(1) Impulse response of the inverse filter A(z), i.e. predictor coefficients a_n, $1 \le n \le p$.

(2) Impulse response of the all-pole model h_n, $0 \le n \le p$, which are easily obtained by long division. Note that the first p+1 coefficients uniquely specify the filter.

(3) Autocorrelation coefficients of $\{a_n/G\}$,

$$b_i = \frac{1}{G^2} \sum_{j=0}^{p-|i|} a_j \, a_{j+|i|} , \quad a_0 = 1 , \quad 0 \le i \le p. \qquad (12)$$

(4) Autocorrelation coefficients of $\{h_n\}$

$$r_i = \sum_{j=0}^{\infty} h_j \, h_{j+|i|} \quad , \quad 0 \le i \le p \quad . \tag{13}$$

It can be shown that r_i is equal to R_i in (3) for $0 \le i \le p$ [6,7].

(5) Spectral coefficients of $A(z)/G$, P_i, $0 \le i \le p$, (or equivalently spectral coefficients of $H(z)$, $1/P_i$)

$$P_i = b_o + 2 \sum_{j=1}^{p} b_j \cos \frac{2\pi ij}{2p+1} \quad , \quad 0 \le i \le p \quad , \tag{14}$$

where b_j are as defined in (12). In words, $\{P_i\}$ is obtained from $\{b_i\}$ through a discrete Fourier transform (DFT). Traditionally, vocoders that transmit the spectrum at selected frequencies have been known as channel vocoders. Thus, use of the spectral coefficients as transmission parameters leads to a <u>linear prediction channel vocoder</u>. While in the classical channel vocoder different channel signals are derived from contiguous band-pass filters, in the linear prediction channel vocoder a selected set of p+1 points from the all-pole spectrum constitute the "channel outputs." The main advantage of the linear prediction channel vocoder, however, is that we are able to regenerate exactly the all-pole spectrum from a knowledge of the p+1 spectral coefficients, unlike in the classical channel vocoder.

(6) Cepstral coefficients of $A(z)$, c_n, $1 \le n \le p$, (or equivalently cepstral coefficients of $H(z)/G$, $-c_n$)

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log A(e^{j\omega}) \quad e^{jn\omega} \, d\omega \quad .$$

Since $A(z)$ is minimum phase, we obtain using the results given in [8, p. 246]

$$c_n = a_n - \sum_{m=1}^{n-1} \frac{m}{n} c_m \, a_{n-m} \quad , \; 1 \le n \le p \quad . \tag{15}$$

(7) Poles of $H(z)$ (or equivalently zeros of $A(z)$).

(8) Reflection coefficients k_i, $1 \le i \le p$, or simple transformations thereof, e.g. area coefficients [2,9]. The area coefficients are given by

$$A_i = A_{i+1} \frac{1+k_i}{1-k_i} \quad , \; A_{p+1} = 1 \; , \; 1 \le i \le p \quad . \tag{16}$$

Although the reflection coefficients are obtained as a byproduct of the solution in (4), they can also be computed directly from the predictor coefficients using the following recursive relations:

$$k_i = a_i^{(i)},$$

$$a_j^{(i-1)} = \frac{a_j^{(i)} - a_j^{(i)} \; a_{i-j}^{(i)}}{1 - k_i^2}, \quad 1 \le j \le i-1, \qquad (17)$$

where the index i takes values $p, p-1, \ldots, 1$ in that order. Initially, $a_j^{(p)} = a_j$, $1 \le j \le p$.

Some of the above sets of parameters have p+1 coefficients while others have only p coefficients. However, for the latter sets the signal energy (or gain G) needs to be transmitted, thus keeping the total number of parameters as p+1 for all the cases. Although the above sets provide equivalent information about the linear predictor, their properties under quantization are different. Certain aspects of the sets (1), (4), (7) and (8) have been studied in the past [2,10]. Our purpose in this paper is to investigate the relative quantization properties of all these parameters with a particular emphasis on the reflection coefficients.

It should be emphasized that the predictor coefficients can be recovered from any of the various sets of parameters listed above. The required transformations for such a recovery are given below only for the sets (3), (5), (6) and (8) since they are well-known for the others.

The sequence $\{b_i\}$ is transformed through an FFT after appending it with an appropriate number of zeros to achieve sufficient resolution in the resulting spectrum of the filter $A(z)/G$. The spectrum of the all-pole filter $H(z)$ is then obtained by simply inverting the amplitudes of the computed spectrum. Inverse Fourier transformation of the spectrum of $H(z)$ yields autocorrelation coefficients $\{r_i\}$ defined in (13). The first p+1 autocorrelation coefficients r_i, $0 \le i \le p$, are then used to compute the predictor coefficients via the normal equations (2) with $R_i = r_i$, $0 \le i \le p$.

The predictor coefficients are recovered from the spectral coefficients $\{P_i\}$ by first taking the inverse DFT of the sequence $\{P_i\}$ to get the autocorrelation sequence $\{b_i\}$. The process of getting the predictor coefficients from $\{b_i\}$ has been discussed above.

Rearranging (15) provides the necessary transformation from cepstral coefficients to predictor coefficients:

$$a_n = c_n + \sum_{m=1}^{n-1} \frac{m}{n} c_m a_{n-m} , \quad 1 \le n \le p \quad . \tag{18}$$

Equations (15) and (18) also suggest the use of the modified cepstral coefficients $\hat{c}_n = nc_r$. as possible transmission parameters.

The predictor coefficients can be recovered from the reflection coefficients using the relations (4-c) with i=1,2,...p, then (4-e).

III. PREPROCESSING METHODS

Before we discuss the quantization properties of the different parameters we should mention that such properties can be improved by proper preprocessing, which is later undone at the synthesizer. For each set of parameters (1-8 above) we have observed that the short-time spectral dynamic range of the speech signal is the single most important factor that affects the quantization properties. We use two methods of preprocessing to reduce the spectral dynamic range and thereby to improve the quantization properties [11]. In the first method, optimal (linear predictive) preemphasis is applied to the speech signal which reduces the spectral dynamic range by reducing the general spectral slope. The second method, called the SIGMA method, involves multiplying the impulse response of the inverse filter A(z) by a decaying exponential, which increases the pole bandwidths, resulting in a reduction of the spectral dynamic range*. Preprocessing by either of these methods can be done after the linear prediction analysis, so that it can be viewed as part of the encoding process.

---

*If, however, a growing exponential is used, the pole bandwidths would be decreased thus enhancing the formant peaks in the spectrum and facilitating better formant tracking [6,7].

IV. QUANTIZATION PROPERTIES

For the purpose of quantization, two desirable properties for a parameter set to have are: (a) filter stability upon quantization and (b) a natural ordering of the parameters. Property (a) means that the poles of H(z) continue to be inside the unit circle even after parameter quantization. By (b) we mean that the parameters exhibit an inherent ordering, e.g. the predictor coefficients are ordered as $a_1, a_2, \ldots, a_p$. If a_1 and a_2 are interchanged then H(z) is no longer the same in general, thus illustrating the existence of an ordering. When such an ordering is present, a statistical study on the distribution of individual parameters can be used to develop better quantization schemes. It is clear that property (a) is more important than (b). Only the poles and the reflection coefficients ensure stability upon quantization, while all the sets of parameters except the poles possess a natural ordering. Thus, only the reflection coefficients possess both of these properties.

We have investigated experimentally the quantization properties of the sets of parameters discussed in Section II, with and without preprocessing of the speech signal. The absolute error between the log power spectra of the unquantized and the quantized linear predictors was used as a criterion in this study, since we believe that a good spectral match is necessary for synthesizing speech with good quality. A summary of the results is provided in the following.

The impulse responses $\{a_n\}$ and $\{h_n\}$ are highly susceptible to causing instability of the filter upon quantization. This is well known from discrete filter analysis. Positive definiteness of autocorrelation coefficients $\{b_i\}$ and $\{r_i\}$ is not ensured under quantization, which also leads to instabilities in the linear prediction filter. An attempt to synthesize speech with quantized autocorrelation coefficients $\{r_i\}$ resulted in distinctly perceivable "clicks" in the synthesized speech. Our conclusion is that the impulse responses and autocorrelation coefficients can be used only under minimal quantization, in which case the transmission rate would be excessive.

In the experimental investigation of the spectral and cepstral parameters, we found that the quantization properties of these parameters are generally superior to those of the impulse responses and autocorrelation coefficients. The spectral parameters often yield results comparable to those obtained by quantizing the reflection coefficients. However, for the cases when the spectrum consists of one or more very sharp peaks (narrow bandwidths), the effects of quantizing the spectral coefficients often cause certain regions in the reconstructed spectrum to become negative, which leads to instability of the filter. Preprocessing the speech signal by the SIGMA method remedies this situation, but the spectral deviation in these regions can be relatively large. Quantization of cepstral parameters can also lead to instabilities. As before, with proper preprocessing stability is restored, but at the expense of increased spectral deviation.

As mentioned earlier, the stability of the filter $H(z)$ is guaranteed under quantization of the poles. This makes the poles potentially a good set of parameters for transmission. Unfortunately, the poles do not possess a natural ordering: a property that is necessary if a low transmission rate is desired. Traditionally, poles have been ordered in terms of vocal tract resonances (formants). Since the ranges of frequencies for the various formants have been well established, their quantization can be done with improved accuracy. In addition, the formant bandwidths may be quantized less accurately than formant frequencies, which leads to further savings in transmission rate. However, experience has shown that the problem of identifying the poles as ordered formants is computationally complex and involves a fair amount of decision making which is not completely reliable. In addition, computing the poles requires finding the roots of a pth order polynomial (p~12): not a straightforward task.

Based on the results of our experimental study of the spectral deviation due to quantization, on computational considerations, and on stability and natural ordering properties, we conclude that the reflection coefficients are the best set for use as transmission parameters. The question now is, what is an optimal quantization scheme for the reflection coefficients which gives the best results in terms of the quality of the synthesized speech? To this end, we perform in the next section a spectral sensitivity analysis of the reflection coefficients, since we have assumed that good quality speech depends on an accurate representation of the power spectrum. Based on the results of this study we present in Section VI an optimal scheme for the quantization of the reflection coefficients.

V. SENSITIVITY ANALYSIS OF REFLECTION COEFFICIENTS

In order to understand the effects of parameter quantization on the all-pole model spectrum, we study in this section the sensitivity of the spectrum to small changes in the reflection coefficients. If ΔS is the spectral deviation due to a change Δk_i in the reflection coefficient k_i, then we define the spectral sensitivity for the coefficient k_i as

$$\frac{\partial S}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{\Delta S}{\Delta k_i} \quad . \tag{19}$$

The definition of spectral deviation ΔS can be arbitrary, but for it to be useful it must somehow relate in a proportional manner to the corresponding effect on perception of the synthesized speech. Here we employ a measure of spectral deviation that has been found to be useful in speech research, namely, the average of the absolute value of the difference between the two log spectra under consideration. Thus the spectral sensitivity is defined by

$$\frac{\partial S}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{1}{\Delta k_i} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \log P(k_i,\omega) - \log P(k_i+\Delta k_i,\omega) \right| d\omega \right],$$

or

$$\frac{\partial S}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{1}{\Delta k_i} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \log \frac{P(k_i,\omega)}{P(k_i+\Delta k_i,\omega)} \right| d\omega \right] , \quad (20)$$

where

$$P(\cdot,\omega) = \left| H(e^{j\omega}) \right|^2$$

is the spectrum of the all-pole model $H(z)$. The quantity between brackets in (20) is the spectral deviation ΔS due to a perturbation in the ith reflection coefficient. Experimentally, $\frac{\partial S}{\partial k_i}$ is computed by replacing the integral by a summation, and by using a sufficiently small value for Δk_i.

Typical sensitivity curves are shown in Fig. 3. (For display purposes we have plotted $10 \log_{10} \frac{\partial S}{\partial k_i}$ in decibels.) These curves were obtained from a 12-pole linear predictive analysis of a 20 msec frame from a 10 kHz sampled speech signal. Each curve in Fig. 3 is a plot of the spectral sensitivity for one of the 12 reflection coefficients as its value is varied over the range (-1,1) while the other 11 reflection coefficients are kept constant. We have performed this type of sensitivity analysis for a large number of different sounds recorded from different speakers. The resulting sensitivity curves were similar to those shown in Fig. 3. The sensitivity curves have the following

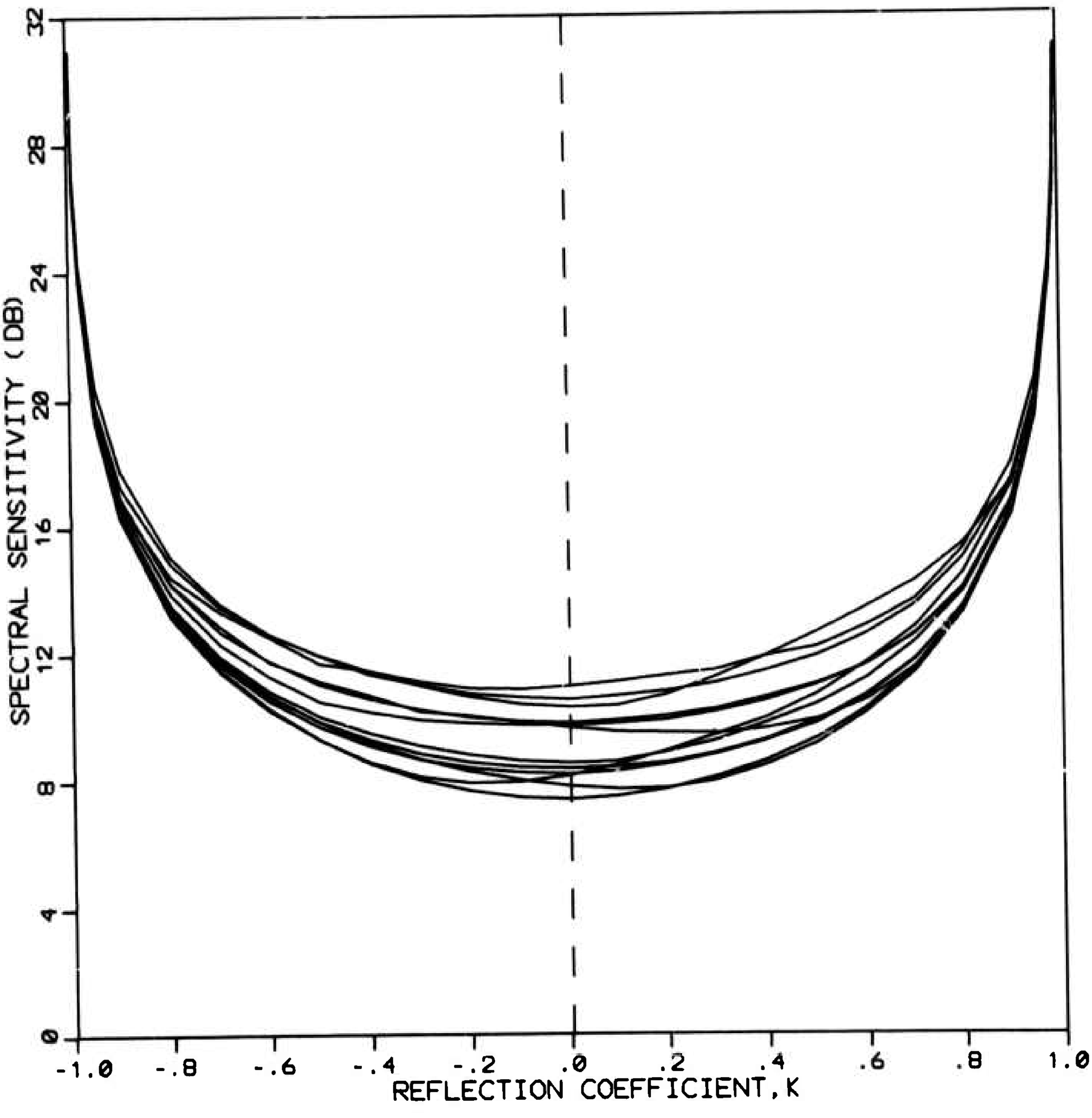


Fig. 3. Typical spectral sensitivity curves for the reflection coefficients of a 12-pole analysis of a 20 msec speech frame.

properties in common:

(i) Each sensitivity curve $\frac{\partial S}{\partial k_i}$ versus k_i has the same general shape irrespective of the index i and irrespective of the values of the other coefficients k_n, $n \neq i$, at which the sensitivity is computed.

(ii) Each sensitivity curve is U-shaped. It is even-symmetric about $k_i=0$, and has large values when the magnitude of k_i is close to 1 and small values when the magnitude of k_i is close to zero.

It has been observed by some researchers that the first few reflection coefficients are the most sensitive to the effects of quantization. While this is true, it is clear from the results of our sensitivity analysis that the high sensitivity is not due to the fact that these reflection coefficients are the leading ones but because on the average they assume magnitudes closer to 1 than the others.

The sensitivity properties given above strongly suggest the existence of a prototype sensitivity function which would apply approximately to every reflection coefficient and for different speech sounds. Such a prototype function could then be used in developing an optimal quantization scheme that would apply to all reflection coefficients all the time. Due to the above sensitivity properties, it is meaningful to obtain this prototype sensitivity function as the simple average of the sensitivity curves over different reflection coefficients and for a large number of different speech sounds. Such an averaged sensitivity function is defined below:

$$\overline{\frac{\partial S}{\partial k}} = \frac{1}{pN} \sum_{t=1}^{N} \sum_{i=1}^{p} \left. \frac{\partial S}{\partial k_i} \right|_{k_i=k} , \qquad (21)$$

where t refers to the number of the analysis frame (time averaging). The averaged sensitivity function for a representative speech sample is shown plotted as the solid curve in Fig. 4. In this plot the sensitivity values are given in decibels relative to the sensitivity at k=0. In the next section, we develop an optimal quantization scheme for the reflection coefficients using the averaged sensitivity function in Fig. 4.

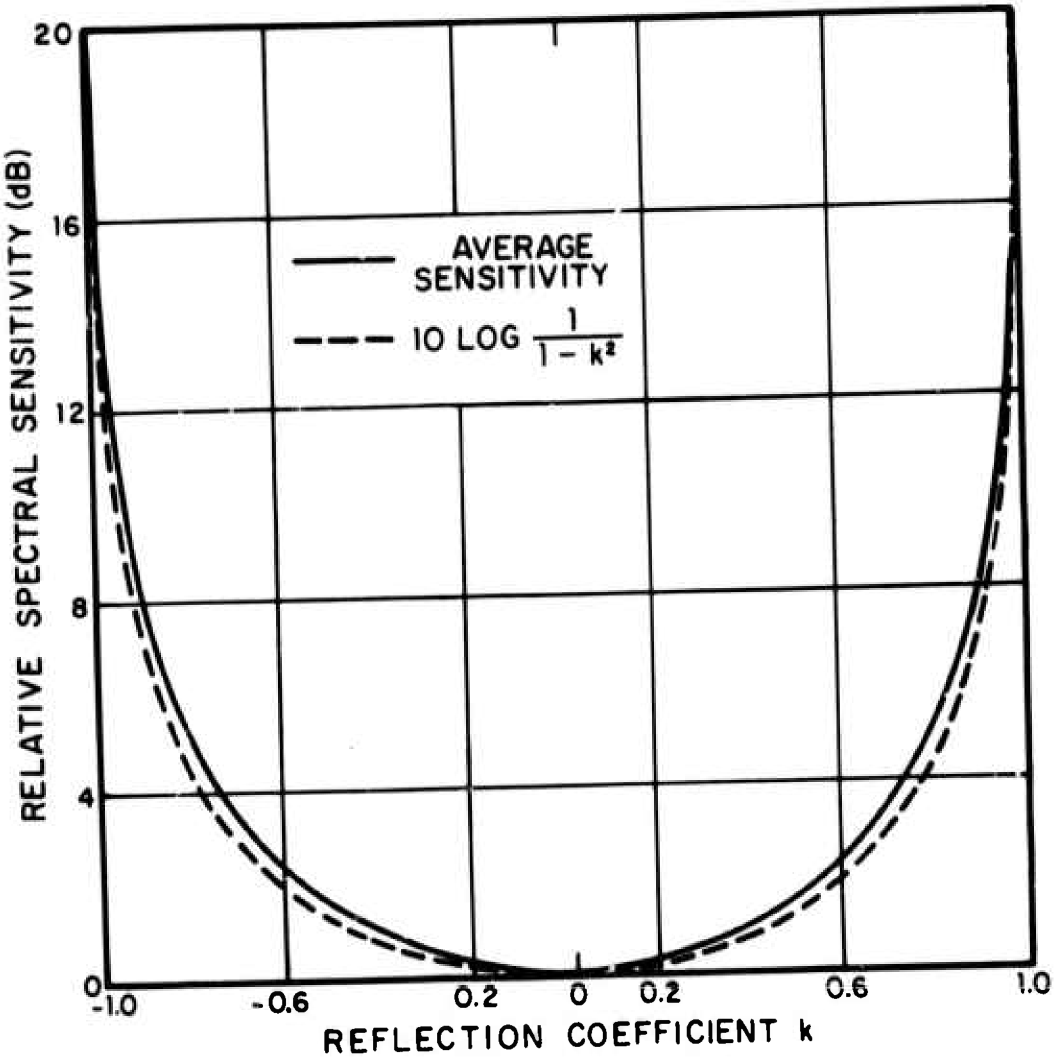


Fig. 4. Averaged spectral sensitivity curve for the reflection coefficients (solid line) and an analytical function that approximates it (dashed line).

VI. OPTIMAL QUANTIZATION OF REFLECTION COEFFICIENTS

In view of the sensitivity properties of the reflection coefficients discussed in the previous section and depicted in Figs. 3 and 4, it is clear that linear quantization of the reflection coefficients is not satisfactory, especially when some of them take values close to 1 in magnitude. What is needed is a nonlinear quantization scheme that is much more sensitive (has more steps) near ±1 than near 0. A nonlinear quantization of a reflection coefficient is equivalent to a linear quantization of a different parameter that is related to the reflection coefficient by a nonlinear transformation. We define an optimal transformation as one which results in a transformed parameter that has a flat or constant spectral sensitivity behavior. We shall now use the results of the previous section to determine this optimal transformation.

Denoting the transformed parameter as g, we have

$$g = f(k) \quad , \tag{22}$$

where $f(\cdot)$ is the underlying nonlinear mapping. The optimal mapping is one where the transformed parameter g has constant spectral sensitivity, i.e.

$$\frac{\partial S}{\partial g} = L = \text{a constant} \quad , \tag{23}$$

where the sensitivity is defined in an analogous manner to (20). Writing formally,

$$\frac{\partial S}{\partial g} = \frac{\partial S}{\partial k}\frac{dk}{dg} = \frac{\partial S}{\partial k} \Big/ \frac{df(k)}{dk} \quad . \tag{24}$$

Thus, from (23) and (24) we have

$$\frac{df(k)}{dk} = \frac{1}{L}\frac{\partial S}{\partial k} \quad . \tag{25}$$

Equation (25) provides the condition for a mapping to be optimal. The optimal mapping f(k) is obtained by simply integrating (25). It is clear that (25) may be applied to each reflection coefficient separately. However, for the reasons mentioned in the last section we shall consider the averaged sensitivity function in Fig. 4 and derive the mapping that is optimal on the average for all the reflection coefficients.

Although it is possible to obtain the optimal transformation by integrating the solid curve in Fig. 4 directly, we have found it simpler and ultimately more useful to approximate the averaged sensitivity curve by a well specified mathematical function which could then be integrated to obtain an approximately optimal f(k). An experimental fitting of the averaged sensitivity curve in

Fig. 4 has revealed that the function $1/(1-k^2)$ approximates the sensitivity function reasonably well (to within a multiplicative constant), as shown by the dashed curve in Fig. 4 (Note that the plot is given in decibels). Thus, from (25), the approximately optimal transformation is given by

$$\frac{df(k)}{dk} = \frac{1}{L(1-k^2)} \quad . \tag{26}$$

Integrating (26) we obtain

$$f(k) = \frac{1}{2L} \log \frac{1+k}{1-k} \quad . \tag{27}$$

As L is arbitrary, an interesting transformation is obtained by substituting L=1/2:

$$f(k) = \log \frac{1+k}{1-k} \quad . \tag{28}$$

From (16), the ratio of consecutive area coefficients is given by

$$\frac{A_i}{A_{i+1}} = \frac{1+k_i}{1-k_i} \ , \ A_{p+1}=1 \ , \ 1 \le i \le p \quad . \tag{29}$$

Therefore, the transformation in (28) is simply the logarithm of the area ratios. Thus, we have shown that the logarithms of the area ratios (henceforth called log area ratios) provide an approximately optimal set of coefficients for quantization.

Fig. 5 shows sensitivity curves for the log area ratios using the same example as in Fig. 3. A comparison of Figs. 3 and 5 shows that the sensitivity curves are relatively flat for the log area ratios. Our experimental investigations into the quality of the synthesized speech also indicate that the log area ratios possess good quantization properties.

Fig. 6 shows a plot of the log area ratio as a function of the reflection coefficient. We have also plotted in Fig. 6 a linear characteristic that passes through the intersection of a vertical line at k=0.7 and the log area ratio curve. For values of k less than 0.7 in magnitude, the log area ratio curve is almost linear. Thus, if a certain reflection coefficient takes values always less than 0.7 in magnitude, one could quantize it linearly to obtain approximately flat sensitivity characteristics. In practice it is found that the reflection coefficients k_i, $i>3$, have in general magnitudes less than 0.7. However, use of the log area ratios automatically leads to the desired quantization irrespective of the reflection coefficient and the range of values it spans.

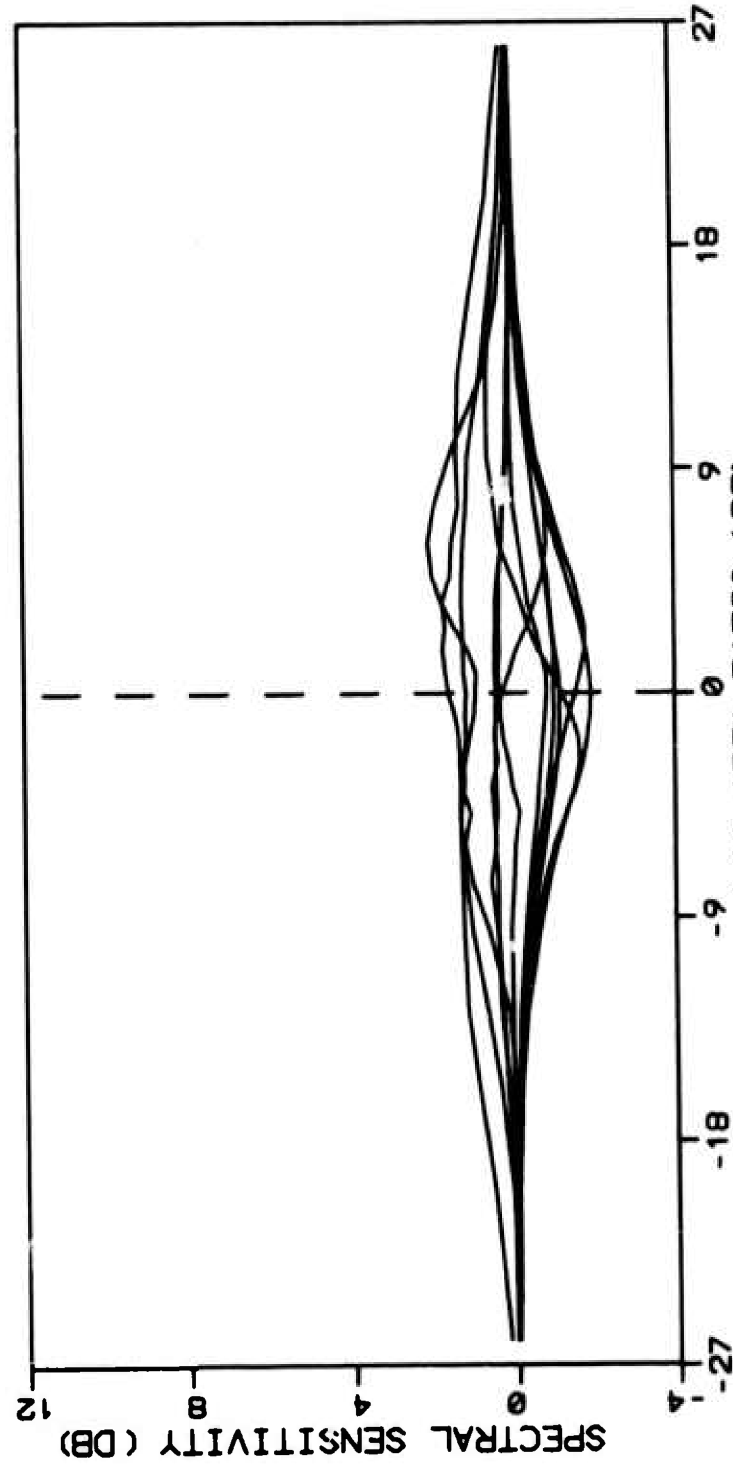


Fig. 5. Spectral sensitivity curves using the log area ratios for the same case as in Fig. 3.

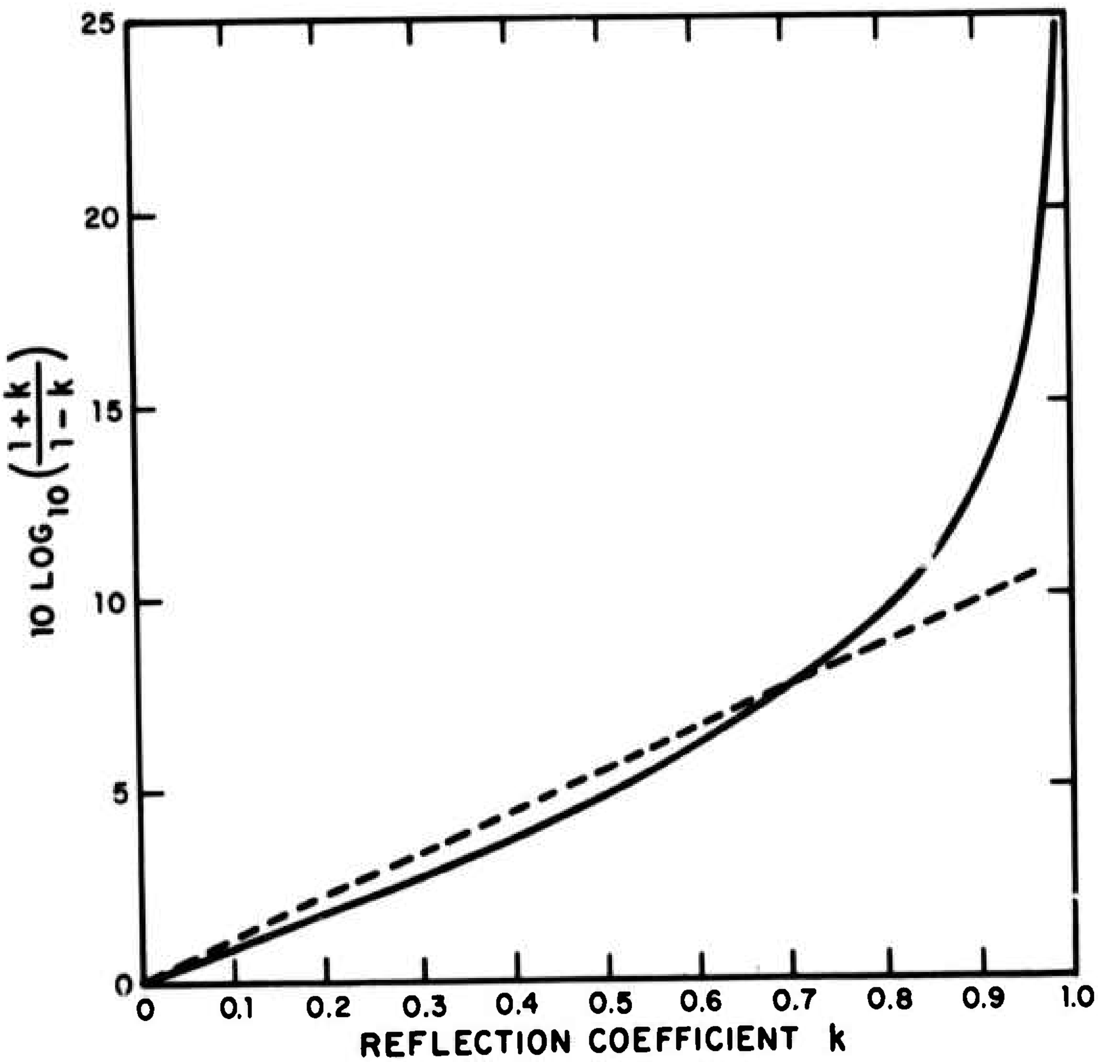


Fig. 6. Log area ratio plotted as a function of the reflection coefficient (solid line) and a linear characteristic that intersects it at k=0.7 (dashed line).

Interpretation in terms of Pole Locations

While the spectral sensitivity measure given by (20) is useful in quantifying the overall deviation in the spectrum due to perturbations in the reflection coefficients or the log area ratios, it does not, however, explain corresponding deviations in the pole locations of the linear prediction filter. Much is known about the relations between the accuracy of pole (or formant) locations and the corresponding effects on speech quality. Therefore, it would be useful to examine the pole deviations due to quantization of the transmission parameters. Unfortunately, the problem is quite untractable in general. However, some insight can still be gained by examining a 2-pole model. Although it is possible to examine this model in mathematical terms, here we shall take a graphical approach due to Kitawaki and Itakura [12].

For the second order linear predictor, the inverse filter is given by

$$A(z) = 1 + k_1(1+k_2)\ z^{-1} + k_2\ z^{-2}\ . \tag{30}$$

The zeros of $A(z)$ are the poles of our model filter $H(z)$. We shall restrict our discussion to the cases where the zeros form complex conjugate pairs. From (30) we see that

A(z) has a complex zero when

$$k_1 \in (-1,1) \quad ,$$
$$k_2 \in \left(\frac{2-k_1^2-2\sqrt{1-k_1^2}}{k_1^2} \; , \; 1 \right) . \tag{31}$$

Fig. 7 shows a plot of only the complex zeros as k_1 is varied uniformly in the interval [-.99, .99] in equal steps of .01 while k_2 is varied uniformly in the interval [0, .99] also in equal steps of .01. Let

$$g_i = \log \frac{1+k_i}{1-k_i} \; , \; i=1,2, \tag{32}$$

be the log area ratios corresponding to k_1 and k_2. Fig. 8 depicts the complex zeros of A(z) when g_1 is varied over [-log 199, log 199] and g_2 over [0, log 199] uniformly and in equal steps. The total number of steps is kept the same as in the previous case. Relative to Fig. 7, Fig. 8 shows that there is denser clustering of the zeros near the unit circle and for angles close to 0 and π. This means that in these regions, quantization errors in the log area ratios lead to a smaller deviation in the position of zeros of A(z) than that obtained by the quantization of the reflection coefficients, assuming the same number of quantization levels in both cases. Fig. 9 shows the complex roots

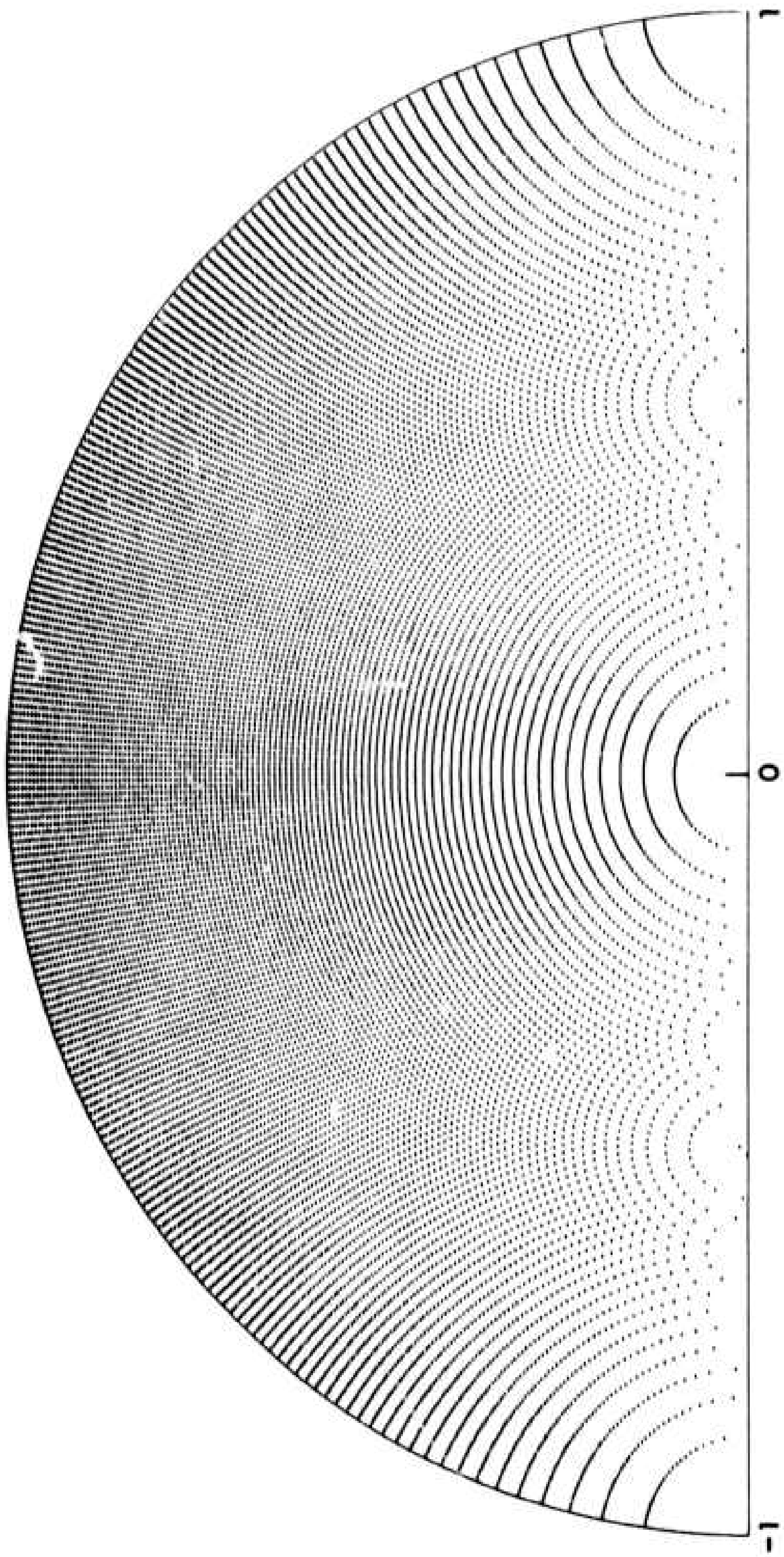


Fig. 7. Root loci for a second order all-pole system obtained by varying the two reflection coefficients in equal steps. (After Kitawaki and Itakura [12].)

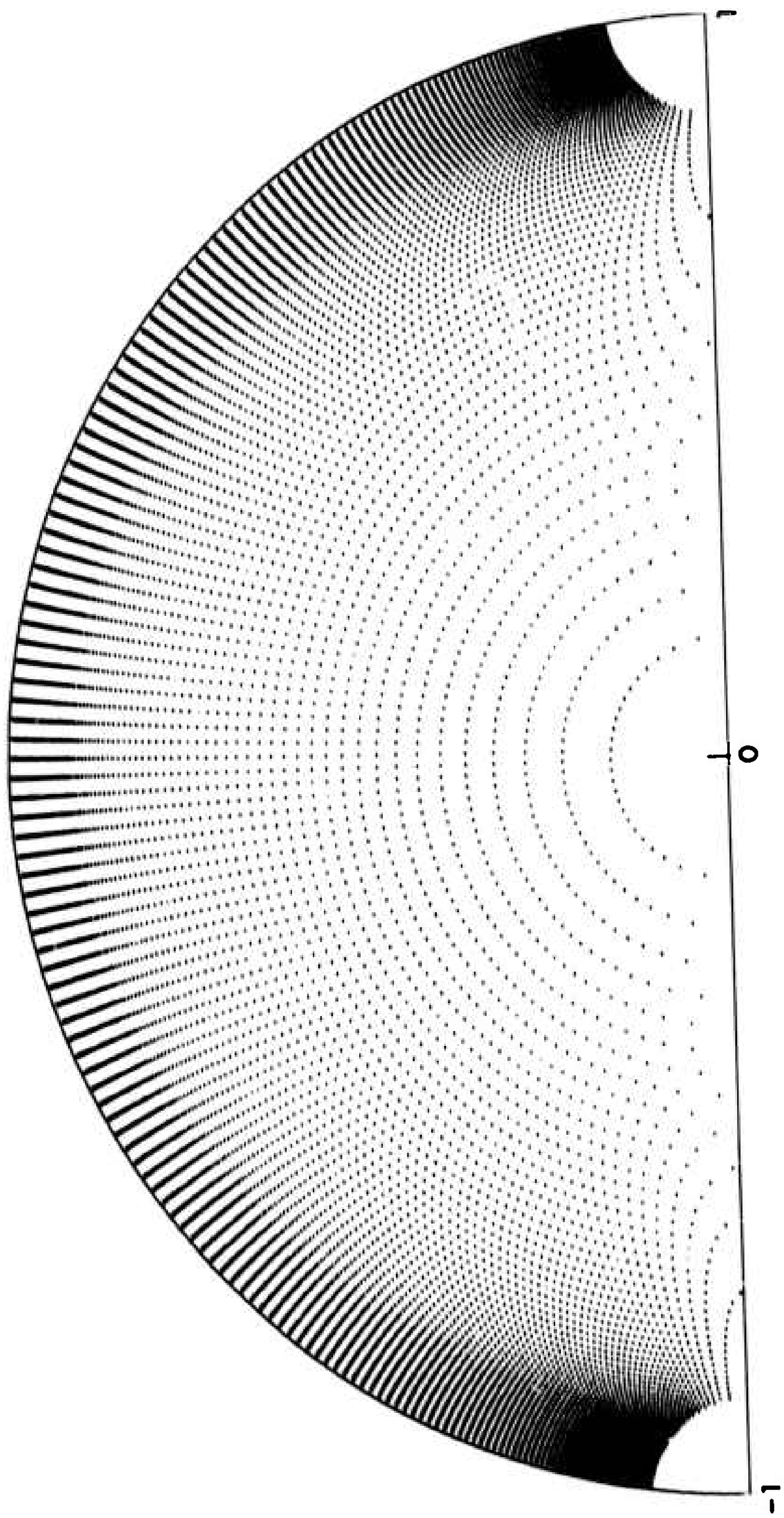


Fig. 8. Root loci for a second order all-pole system obtained by varying the two log area ratios in equal steps (the total number of steps being the same as in Fig. 7). (After Kitawaki and Itakura [12].)

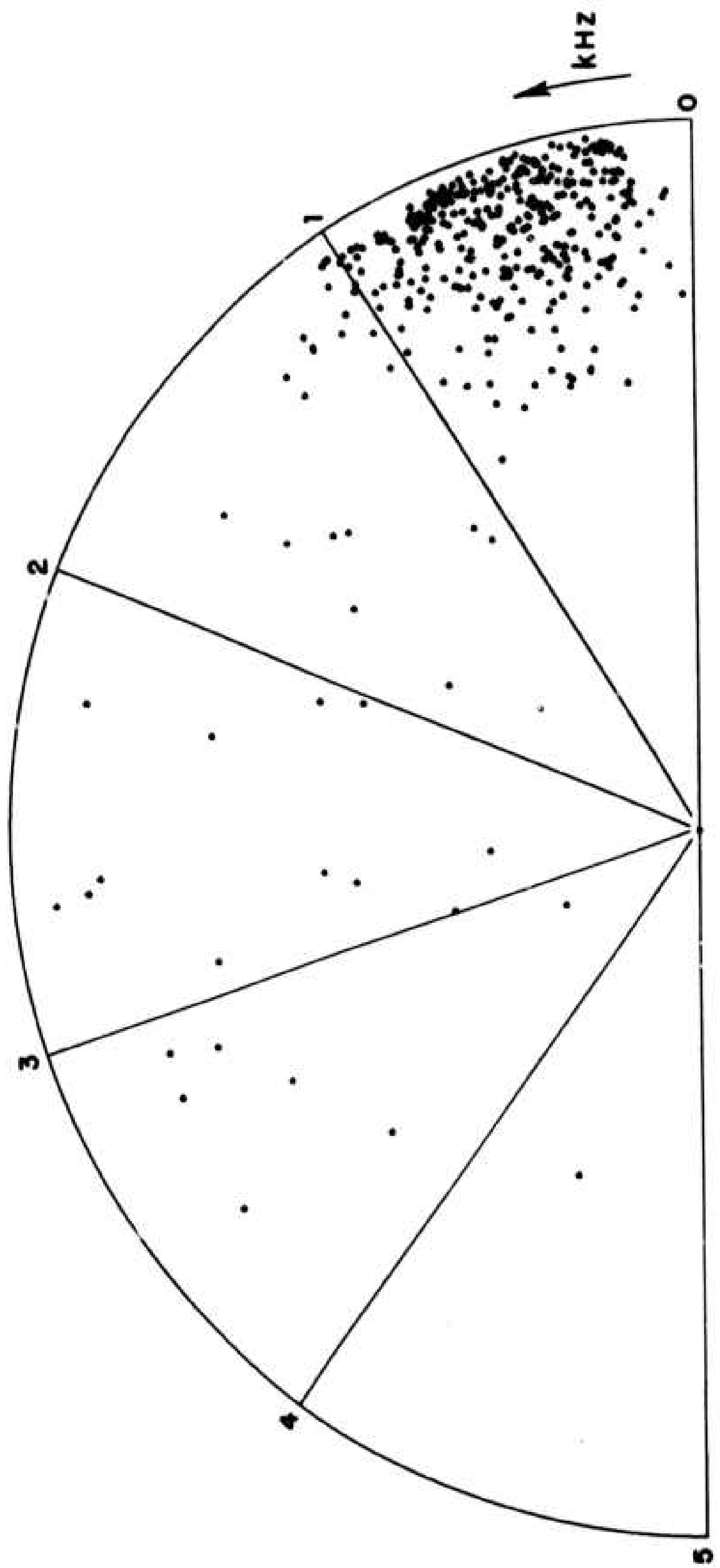


Fig. 9. Complex roots obtained by a second order linear predictive analysis of about 30 secs of continuous speech sampled at 10 kHz.

obtained by a second order linear predictive analysis of several sentences of speech material sampled at 10 kHz. An inspection of Figs. 8 and 9 reveals that the roots of the second order linear predictor for the continuous speech occur mainly in the areas where there is a dense clustering of zeros in Fig. 8. We view this as further independent evidence supporting our earlier findings of the desirable quantization properties of the log area ratios for the purpose of speech compression.

Kitawaki and Itakura considered still other nonlinear mappings of the reflection coefficients but concluded that the log area ratios lead to the best overall quantization accuracy [12]. Our results make the stronger statement that the log area ratios are actually optimal in the sense discussed earlier.

Optimum Bit Allocation

In the following we investigate the use of the spectral sensitivity measure for allocating a fixed number of bits among the various parameters. Let $q_1, q_2, \ldots, q_p$ be the parameters chosen for quantization. These may be the reflection coefficients or the log area ratios or any other set of parameters. Given the total number of bits for quantization as M, the problem is to distribute this among the p parameters as M_i, $1 \le i \le p$, in some optimal manner. In terms of quantization levels, the above problem may be restated as the allocation of $N = 2^M$ levels among the p parameters as N_i, $1 \le i \le p$, in some optimal manner. Therefore, we have

$$\sum_{i=1}^{p} M_i = \sum_{i=1}^{p} \log_2 N_i = M \quad ,$$
$$\prod_{i=1}^{p} N_i = N \; , \; N_i = 2^{M_i} \; , \; 1 \le i \le p \quad . \tag{33}$$

We propose to derive the optimal bit allocation by minimizing the maximum spectral deviation due to quantization. The total spectral deviation ΔS due to changes Δq_i in the parameters q_i, $1 \le i \le p$, is given approximately by

$$\Delta S = \sum_{i=1}^{p} \left| \frac{\partial S}{\partial q_i} \Delta q_i \right| \quad . \tag{34}$$

Define the quantization step size for q_i as

$$\delta_i = \frac{\overline{q}_i - \underline{q}_i}{N_i} \quad , \tag{35}$$

where $\overline{q}_i$ and $\underline{q}_i$ are the upper and lower bounds on q_i, respectively. Then, for a linear quantization of q_i using round-off arithmetic, the maximum quantization error is equal to half the quantization step size:

$$|\Delta q_i|_{max} = \frac{1}{2}\,\delta_i \quad .$$

Thus

$$(\Delta S)_{max} = \sum_{i=1}^{p} \left|\frac{\partial S}{\partial q_i}\right| \frac{\bar{q}_i - \underline{q}_i}{2N_i} \quad . \qquad (36)$$

Let

$$K_i = \frac{\bar{q}_i - \underline{q}_i}{2} \left|\frac{\partial S}{\partial q_i}\right| , \quad 1 \le i \le p \quad . \qquad (37)$$

Then

$$(\Delta S)_{max} = \sum_{i=1}^{p} \frac{K_i}{N_i} \quad . \qquad (38)$$

We wish to minimize $(\Delta S)_{max}$ with respect to $\{N_i\}$ subject to the constraint

$$\sum_{i=1}^{p} \log_2 N_i = M \quad . \qquad (39)$$

This is a simple problem in constrained minimization [13] and its solution is given by

$$N_1 = K_1 \left[\frac{2^M}{\prod_{i=1}^{p} K_i} \right]^{1/p} ,$$

$$N_i = \frac{K_i}{K_1} N_1 , \quad 2 \le i \le p . \tag{40}$$

The bit allocation strategy given in (40) is thus optimal in a minimax sense since it minimizes the maximum spectral deviation due to quantization. Note that if truncation arithmetic is used, the constants K_i in (37) will be doubled, but that will not affect the bit allocation results from (40).

The optimal bit allocation in (40) effectively says that the contributions of the different parameters to the maximum spectral deviation in (38) must be equal. We know that for the log area ratios the spectral sensitivity $\frac{\partial S}{\partial q_i}$ is approximately a constant and is the same for all the coefficients. From (35), (37) and (40), this implies that the quantization step size δ_i should be the same for all the log area ratios, which is intuitively clear. For this case, the bit allocation can be done as follows. Compute the constant step size δ from

$$\delta = \left[\frac{\prod_{i=1}^{p} (\bar{q}_i - \underline{q}_i)}{2^M} \right]^{1/p} . \tag{41}$$

Then the number of levels N_i for each coefficient is computed from (35). We have found it convenient and useful to begin with a particular step size. That automatically determines the total number of bits needed, as well as the maximum spectral deviation which, in turn, determines the resulting speech quality. One can then study the change in speech quality as a function of only one variable, namely the step size.

VII. COMMENTS ON ANOTHER SPECTRAL SENSITIVITY MEASURE

In Section V we introduced a spectral sensitivity measure to study the quantization properties of the reflection coefficients. Other types of sensitivity measures may also be used. In particular we have considered a measure which is similar to the total-squared error used for minimization in linear predictive analysis. By using Parseval's theorem in (1), the total-squared error is given by

$$E = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{P_o(\omega)}{P(\omega)} d\omega , \tag{42}$$

where $P_o(\omega)$ is the power spectrum of the input speech signal and $P(\omega)$ is the power spectrum of the all-pole filter:

$$P(\omega) = \left|H(e^{j\omega})\right|^2 = \frac{G^2}{\left|A(e^{j\omega})\right|^2} . \tag{43}$$

The gain G is given by (9).

Properties of the error measure E have been studied in detail elsewhere [6,7,14]. In particular, the minimization of E results in an all-pole model spectrum $P(\omega)$ that is a good approximation to the envelope of the signal spectrum $P_o(\omega)$. Because of this property, it seemed reasonable to study the use of this error E as a measure of the deviation between the two spectra. For the sake of normalization we have chosen to work with an error measure E' obtained from (42) by eliminating the factor G^2:

$$E' = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_1(\omega)}{P_2(\omega)} d\omega \quad , \tag{44}$$

where $P_1(\omega)$ and $P_2(\omega)$ are now any two spectra. Also, the two spectra are normalized such that they have equal total energy.

For our study of spectral sensitivity we let $P_1(\omega)=P(k_i,\omega)$ and $P_2(\omega)=P(k_i+\Delta k_i,\omega)$, where $P(\cdot,\omega)$ is given by (43). The error between the two spectra is then given by

$$E'(\Delta k_i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P(k_i,\omega)}{P(k_i+\Delta k_i,\omega)} d\omega \quad . \tag{45}$$

We define the spectral deviation, then, as

$$\Delta S' = \log E'(\Delta k_i) \quad . \tag{46}$$

The definition of the new measure of spectral sensitivity follows from (46) and (45) as

$$\frac{\partial S'}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{1}{\Delta k_i} \log\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P(k_i,\omega)}{P(k_i+\Delta k_i,\omega)}\right] d\omega \quad . \tag{47}$$

The spectral sensitivity in (47) can be derived analytically, without the need to resort to experimental data as was the case for the study of $\frac{\partial S}{\partial k_i}$ in (20). This is done below.

$E'(\Delta k_i)$ in (45) can be interpreted as the arithmetic mean of the ratio of the two spectra. For small Δk_i, the arithmetic mean is approximately equal to the geometric mean, which is given by

$$E''(\Delta k_i) = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{P(k_i,\omega)}{P(k_i+\Delta k_i,\omega)}\, d\omega\right]. \qquad (48)$$

As $\Delta k_i \to 0$, the arithmetic mean becomes equal to the geometric mean. Using this result, we have from (45), (47) and (48),

$$\frac{\partial S'}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{1}{\Delta k_i}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{P(k_i,\omega)}{P(k_i+\Delta k_i,\omega)}\, d\omega\right]. \qquad (49)$$

Substituting (9) and (43) in (49), there results

$$\frac{\partial S'}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{1}{\Delta k_i}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{V_p(k_i)}{V_p(k_i+\Delta k_i)} + \log\left|\frac{A(k_i+\Delta k_i, e^{j\omega})}{A(k_i, e^{j\omega})}\right|^2 d\omega\right]. \qquad (50)$$

It can be shown [7] that if the zeros of A(z) lie inside the unit circle, then

$$\int_{-\pi}^{\pi} \log|A(e^{j\omega})|^2 \; d\omega = 0 \quad . \qquad (51)$$

Substituting (51) in (50), and noting that V_p is independent of ω, we obtain

$$\frac{\partial S'}{\partial k_i} = \lim_{\Delta k_i \to 0} \frac{\log V_p(k_i) - \log V_p(k_i+\Delta k_i)}{\Delta k_i} \quad , \qquad (52)$$

or

$$\frac{\partial S'}{\partial k_i} = - \frac{\partial[\log V_p(k_i)]}{\partial k_i} \quad .$$

Using (8) in (52) we obtain the desired result

$$\frac{\partial S'}{\partial k_i} = \frac{2k_i}{1-k_i^2} \quad . \qquad (53)$$

It is important to note that this is an exact result and it is true for each reflection coefficient, independent of the values of the other coefficients. Also, a plot of $\left|\frac{\partial S'}{\partial k}\right|$ versus k gives a U-shaped curve. Therefore, the spectral sensitivity in (53) has the same general properties as the spectral sensitivity $\frac{\partial S}{\partial k}$ obtained experimentally in Section V. The only difference between the two is the actual shape of the sensitivity curve.

Substituting (53) in the optimality condition (25) and integrating it with L=1, we obtain the following optimal mapping for the sensitivity measure (47):

$$f'(k) = \log \frac{1}{1-k^2} \quad . \tag{54}$$

From (8) and (54), it is interesting to observe that $f'(k_i)$ is equal to the logarithm of the ratio of the normalized errors (or log error ratio) associated with the linear predictors of orders i-1 and i,

$$f'(k_i) = \log \frac{V_{i-1}}{V_i} \quad . \tag{55}$$

We have experimentally investigated the quantization properties resulting from the mappings given by (28) and (55). Through informal listening tests we have found that the use of the log area ratios for quantization leads to uniformly better speech quality than that obtained using the log error ratios.

It is interesting to note that the only difference between the two sensitivity measures given by (20) and (49) is the lack of an absolute value sign inside the integral in

(49). This makes the sensitivity measure in (49) less powerful, because spectral deviations when $P(k_i,\omega)>P(k_i+\Delta k_i,\omega)$ can cancel deviations when $P(k_i+\Delta k_i,\omega)>P(k_i,\omega)$. Both of these cases contribute to the total spectral deviation in (20). This is another reason why (20) is to be preferred over (49) as a definition of spectral sensitivity, and therefore why the log area ratios are to be preferred over the log error ratios as transmission parameters. (See [14] for further comparison of the spectral deviations in (20) and (47).)

VIII. CONCLUSIONS

We have dealt with the problem of quantization of transmission parameters in linear predictive speech compression systems. Several alternate sets of transmission parameters were considered and their relative quantization properties were presented. The results of this study have shown that the reflection coefficients are the best set for use as transmission parameters. Specifically, the reflection coefficients preserve the stability of the linear predictor under quantization, and possess a natural ordering which property can be used in the design of better quantization schemes. The quantization of the reflection coefficients was then examined in more detail using a spectral sensitivity measure.

The spectral sensitivity of a given reflection coefficient was defined in terms of the absolute spectral deviation due to a small perturbation in the reflection coefficient. Experimental study of this spectral sensitivity measure has shown that a reflection coefficient has a high sensitivity for magnitudes close to 1 and a low sensitivity near 0. Further, all the reflection coefficients have approximately the same sensitivity behavior, irrespective of the particular speech sound to which they correspond. A prototype sensitivity function was obtained experimentally by averaging the sensitivity values over the various reflection coefficients and over a large number of speech sounds. We have then developed an optimal quantization procedure for the reflection coefficients. This consisted of finding a suitable mapping that transforms the reflection coefficients to a set of parameters having a flat or constant sensitivity behavior. Using an analytical function that well approximates the averaged sensitivity of the reflection coefficients, we demonstrated that the logarithms of the ratios of area coefficients (or log area ratios) possess approximately optimal quantization properties.

An optimal solution was then derived for the problem of bit allocation among the different parameters. This was done by minimizing the maximum spectral deviation due to quantization. For the log area ratios, this optimal solution reduces to using equal quantization steps for all the parameters.

Finally, motivated to use an error measure similar to the one used in linear predictive analysis, we have provided an alternate definition of spectral sensitivity. An analytical evaluation of this spectral sensitivity for the reflection coefficients has shown that the logarithms of the ratios of normalized errors of linear predictors of successive orders (on log error ratios) exhibit optimal quantization properties. However, informal listening tests have indicated that the use of log area ratios for quantization leads to better synthesis than the use of log error ratios. This further implies that the definition of spectral sensitivity that resulted in the log area ratios gives a superior measure of spectral sensitivity for the purpose of quantization studies.

ACKNOWLEDGMENT

The authors would like to thank F. Itakura for bringing his work [12] to their attention.

REFERENCES

[1] Itakura, F. and S. Saito, "Analysis Synthesis Telephony based upon the Maximum Likelihood Method," Reports of the 6th International Congress on Acoustics, edited by Y. Kohasi, Tokyo, C-5-5, Aug. 21-28, 1968.

[2] Atal, B.S. and S.L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am., vol. 50, pp.637-655, 1971.

[3] Itakura, F., et al., "An Audio Response Unit Based on Partial Autocorrelation," IEEE Trans. Comm., vol. COM-20, no. 4, pp. 792-797, August 1972.

[4] Haskew, J.R., J.M. Kelly, and T.H. McKinney, "Results of a Study of the Linear Prediction Vocoder," IEEE Trans. Comm., vol. COM-21, no. 9, pp. 1008-1014, Sept. 1973.

[5] Markel, J.D. and A.H. Gray, Jr., "A Linear Prediction Vocoder Simulation Based Upon the Autocorrelation Method," Trans. Acoustics, Speech and Signal Processing, vol. ASSP-22, pp. 124-134, April 1974.

[6] Makhoul, J., "Spectral Analysis of Speech by Linear Prediction," IEEE Trans. Audio and Electroacoustics, vol. AU-21, pp. 140-148, June 1973.

[7] Makhoul, J.I. and J.J. Wolf, "Linear Prediction and the Spectral Analysis of Speech," BBN Report No. 2304. Bolt Beranek and Newman Inc., Cambridge, Mass., August 1972.

[8] Gold, B. and C.M. Rader, Digital Processing of Signals, McGraw-Hill, New York, 1969.

[9] Wakita, H., "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms," IEEE Trans. Audio and Electroacoustics, vol. AU-21, pp. 417-427, Oct. 1973.

[10] Itakura, F. and S. Saito, "On the Optimum Quantization of Feature Parameters in the PARCOR Speech Synthesizer," Conference Record, 1972 Conf. on Speech Comm. and Processing, Newton, Mass., pp. 434-437, April 1972.

[11] Makhoul, J. and R. Viswanathan, "Adaptive Preprocessing for Linear Predictive Speech Compression Systems," Presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Oct. 30-Nov. 2, 1973.

[12] Kitawaki, N. and F. Itakura, "Nonlinear Coding of PARCOR Coefficients," Meeting of the Acoustic Society of Japan, Oct. 1973, pp. 449-450, (in Japanese).

[13] Bryson, A.E., Jr. and Y.C. Ho, Applied Optimal Control, Blaisdell Publishing Co., Waltham, Mass., 1969.

[14] Makhoul, J., "Selective Linear Prediction and Analysis-by-Synthesis in Speech Analysis," BBN Report No. 2578, Bolt Beranek and Newman Inc., Cambridge, Mass., April 1974.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Bolt Beranek and Newman Inc. 50 Moulton Street Cambridge, Mass. 02138 AD-A108 350	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE
"Quantization Properties of Transmission Parameters in Linear Predictive Systems"

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (First name, middle initial, last name)
John Makhoul
R. Viswanathan

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1974	51	14

8a. CONTRACT OR GRANT NO. DAHC-71-C-0088 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) BBN Report No. 2800 A.I. Report No. 14 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

Several alternate sets of parameters that represent the linear predictor are investigated as transmission parameters for linear predictive speech compression systems. Although each of these sets provides equivalent information about the linear predictor, their properties under quantization are different. The results of a comparative study of the various parameter sets are reported. Specifically it is concluded that the reflection coefficients are the best set for use as transmission parameters. A more detailed investigation of the quantization properties of the reflection coefficients is then carried out using a spectral sensitivity measure. A method of optimally quantizing the reflection coefficients is also derived. Using this method it is demonstrated that logarithms of the ratios of the familiar area functions possess approximately optimal quantization properties. Also, a solution to the problem of bit allocation among the various parameters is presented, based on the sensitivity measure.

The use of another spectral sensitivity measure renders logarithms of the ratios of normalized errors associated with linear predictors of successive orders as the optimal quantization parameters. Informal listening tests indicate that the use of log area ratios for quantization leads to better synthesis than the use of log error ratios.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Unclassified
Security Classification

A-31408

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Linear Prediction						
Speech Compression						
Vocoders						
Speech Analysis						
Speech Synthesis						
Quantization						
Filter Stability						
Spectral Sensitivity						
Bit Allocation						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-607-6821

Unclassified
Security Classification

A-31409