

- [2] W. A. Ainsworth, "A system for converting English text into speech," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 288-290, June 1974.
- [3] M. D. McLroy, "Synthetic English speech by rule," Bell Tele. Lab., Inc., Murray Hill, NJ, Mar. 1974.
- [4] F. F. Lee, "Machine to man communication by speech Part I: Generation of segmental phonemes from text," in *1968 Spring Joint Computer Conf.*, pp. 333-338.
- [5] J. Allen, "Machine to man communication by speech, Part II: Synthesis of prosodic features of speech by rule," in *1968 Spring Joint Computer Conf.*, pp. 339-344.
- [6] F. F. Lee, "Reading machine: From text to speech," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 275-282, Dec. 1969.
- [7] J. Allen, "Speech synthesis from unrestricted text," *IEEE Convention Digest*, 1971, pp. 108-109.
- [8] —, "Reading machines for the blind: The technical problems and the methods adopted for their solution," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 259-264, June 1973.
- [9] H. Kucera and W. N. Francis, *Computational Analysis of Present-Day American English*. Providence, RI: Brown Univ. Press, 1967.
- [10] H. Elovitz, R. Johnson, A. McHugh, and J. Shore, "Automatic translation of English text to phonetics by means of letter-to-sound rules," Naval Res. Lab., Washington, DC, NRL Rep. 7948, Jan. 1976.
- [11] W. A. Ainsworth, private communication, Apr. 1974.
- [12] J. S. Kenyon and T. A. Knott, *A Pronouncing Dictionary of American English*. Springfield, MA: Merriam, 1951.
- [13] P. J. Santos, Jr., "FASBOL II, a SNOBOL compiler for the PDP-10," Digital Equipment Computer Users' Society, DECUS 10-179, Dec. 1972.
- [14] R. Venezky, "A study of English spelling-to-sound correspondence on historical principles," Stanford Univ., Stanford, CA, 1965.
- [15] —, *The Structure of English Orthography*. The Hague, The Netherlands: Mouton, 1970.
- [16] N. Chomsky and M. Halle, *The Sound Patterns of English*. New York: Harper & Row, 1968.
- [17] A. McHugh, "Listener preference and comprehension tests of stress algorithms for a text-to-phonetic speech synthesis program," Naval Res. Lab., Washington, DC, NRL Rep. 8015, Sept. 1976.
- [18] —, "Frequency ranked and alphabetical lists of words judged by navy officers to be frequent in navy usage," to be published.

Quantization and Bit Allocation in Speech Processing

AUGUSTINE H. GRAY, JR., MEMBER, IEEE, AND JOHN D. MARKEL, MEMBER, IEEE

Abstract—The topic of quantization and bit allocation in speech processing is studied using an L_2 norm. Closed-form expressions are derived for the root mean square (rms) spectral deviation due to variations in one, two, or multiple parameters. For one-parameter variation, the reflection coefficients, log area ratios, and inverse sine coefficients are studied. It is shown that, depending upon the criterion chosen, either log area ratios or inverse sine quantization can be viewed as optimal. From a practical point of view, it is shown experimentally that very little difference exists among the various quantization methods beyond the second coefficient.

Two-parameter variations are studied in terms of formant frequency and bandwidth movement and in terms of a two-pair quantization scheme. A lower bound on the number of quantization levels required to satisfy a given maximum spectral deviation is derived along with the two-pair quantization scheme which approximately satisfies the bound. It is shown theoretically that the two-pair quantization scheme has a 10-bit superiority over other above-mentioned quantization schemes in the sense of theoretically assuring that a maximum overall log spectral deviation will not be exceeded.

Manuscript received January 12, 1976; revised June 7, 1976. This work was supported by the Department of Defense under Contract MDA904-76-C-0259.

A. H. Gray, Jr. is with the Department of Electrical Engineering, University of California, Santa Barbara, CA 93106 and the Speech Communications Research Laboratory, Inc., Santa Barbara, CA 93109.

J. D. Markel is with the Speech Communications Research Laboratory, Inc., Santa Barbara, CA 93109.

I. INTRODUCTION

THE QUANTIZATION properties of transmission parameters in linear prediction speech compression systems have been discussed in a recent paper by Viswanathan and Makhoul [1] on the basis of a spectral sensitivity analysis using an L_1 norm on log spectral differences. The emphasis was on sensitivity to the reflection coefficient parameter set, $\{k_1, k_2, \dots, k_M\}$, which can define an all-pole filter model. The reflection coefficients have the desirable property of retaining stability under quantization. Spectral sensitivity curves were numerically obtained and used to show that log area quantization of the reflection coefficients was optimal in the sense of minimizing a maximum spectral error over the entire range of possible values for the reflection coefficients, -1 to $+1$. In addition, based upon the assumption that total spectral deviation due to changes in all the parameters can be approximated by the sum of individual deviations, a bit-allocation scheme was proposed for minimizing the total spectral deviation.

The topic of quantization and bit allocation is also studied here, but using an L_2 as opposed to the L_1 norm. This seemingly minor change has a significant effect in that tractable

analytical results can be used without approximations or frequency-domain calculations. Within this framework, the results of Viswanathan and Makhoul [1] are reinterpreted and a number of new or previously unpublished results are shown. In terms of single-parameter variations which cover all possible values of each reflection coefficient from -1 to $+1$, it is shown that the log area quantization of [1] theoretically results from the choice of a gain term which normalizes the all-pole model energies. If the log spectral means are normalized, then inverse sine quantization becomes optimal in the above-defined sense. In a practical sense it is shown that with the exception of the first two-reflection coefficients, simple linear quantization produces essentially equivalent results and, in fact, slightly superior results over the range of reflection coefficients measured from actual speech.

II. THEORY

Assume a linear prediction all-pole model $H(z, \lambda)$ of the form

$$H(z, \lambda) = \sigma(\lambda)/A(z, \lambda), \quad (1)$$

where λ is a parameter row vector

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_L). \quad (2)$$

The inverse filter polynomial $A(z, \lambda)$ has the form

$$A(z, \lambda) = \sum_{i=0}^M a_i z^{-i} \text{ with } a_0 = 1, \quad (3)$$

where $\sigma(\lambda)$ is a gain term. Both the gain terms $\sigma(\lambda)$ and the filter coefficients $a_1, a_2, \dots, a_M$ may be functions of the parameter vector λ . It is assumed that $A(z, \lambda)$ has all of its roots within the unit circle so that $H(z, \lambda)$ is minimum phase. Analogously, a model $H(z, \lambda + \Delta\lambda)$ is defined by replacing λ with $\lambda + \Delta\lambda$ where

$$\Delta\lambda = (\Delta\lambda_1, \Delta\lambda_2, \dots, \Delta\lambda_L). \quad (4)$$

It is assumed that the elements of $\Delta\lambda$ are sufficiently small to insure that $H(z, \lambda + \Delta\lambda)$ remains a minimum-phase function.

The log spectral difference between $H(z, \lambda)$ and $H(z, \lambda + \Delta\lambda)$ is defined as

$$\begin{aligned} \Delta V(\theta) &= \ln |H(e^{j\theta}, \lambda + \Delta\lambda)|^2 - \ln |H(e^{j\theta}, \lambda)|^2 \\ &= \ln |H(e^{j\theta}, \lambda + \Delta\lambda)/H(e^{j\theta}, \lambda)|^2, \end{aligned} \quad (5)$$

where θ is the normalized frequency (the angle on the unit circle in the z plane). A distance between the log spectra, Δd , or equivalently, the length of $\Delta V(\theta)$, can be defined in terms of an L_p norm as

$$\Delta d = \left(\int_{-\pi}^{\pi} |\Delta V(\theta)|^p \frac{d\theta}{2\pi} \right)^{1/p}, \quad (6)$$

where p is a positive integer.

Viswanathan and Makhoul [1] used the L_1 norm, $p = 1$, so that Δd represented the mean absolute value of $\Delta V(\theta)$. It has been noted [2] that when analyzing real speech, the choice of p is not a major consideration in terms of the end results. Although the L_p norms increase monotonically with p , they are

all highly correlated for differences in all-pole model log spectra over reasonable ranges expected in speech processing. As a result, we shall restrict ourselves to the value of $p = 2$, so that the norm is a root mean square (rms) value. This leads to mathematically tractable results which can be numerically evaluated without frequency-domain calculations.

For the remainder of this study, unless otherwise stated, $p = 2$ and L_2 norms are implied. The rms log spectral deviation Δd will be called the spectral deviation. The norm square $(\Delta d)^2$ will be called the mean-square deviation. When only one parameter is being considered, independent of the others ($L = 1$), the parameter row vector λ will be replaced by the scalar λ . For more general values of p other than $p = 2$, Δd of (6) can only be estimated by using discrete Fourier transforms (DFT's) and numerical approximations of the integral. For $p = 2$ and incremental changes in the parameters, known properties of analytic functions can be used to evaluate Δd in terms of finite summations without approximations.

A. Spectral Deviation Evaluation

Equation (5) can be rewritten using (1) as

$$\Delta V(\theta) = 2 \ln [\sigma(\lambda + \Delta\lambda)/\sigma(\lambda)] - \Delta G(e^{j\theta}) - \Delta G^*(e^{j\theta}), \quad (7)$$

where

$$\Delta G(z) = \ln [A(z, \lambda + \Delta\lambda)/A(z, \lambda)], \quad (8)$$

and the asterisk denotes a complex conjugate. As $A(z, \lambda)$ and $A(z, \lambda + \Delta\lambda)$ are minimum phase, $\Delta G(1/z)$ is analytic on and within the unit circle. This fact can be used with (6) and (7) to yield

$$(\Delta d)^2 = 4 \{ \ln [\sigma(\lambda + \Delta\lambda)/\sigma(\lambda)] \}^2 + 2 \int_{-\pi}^{\pi} |\Delta G(e^{j\theta})|^2 \frac{d\theta}{2\pi}. \quad (9)$$

The mean-square deviation is thus expressed as the sum of the mean-square deviation arising from the gain term $\sigma(\lambda)$, and the integral of $|\Delta G(e^{j\theta})|^2$. The effects of the gain term are seen to be additive in the mean-square deviation, and will be considered separately later. For purposes of the present discussion, the gain term is defined by

$$\sigma(\lambda) = 1,$$

which minimizes the mean-square deviation and normalizes the log spectra [2] so that

$$\int_{-\pi}^{\pi} \ln |H(e^{j\theta}, \lambda)|^2 \frac{d\theta}{2\pi} = 0. \quad (10)$$

Discussion of other gain normalizations is found elsewhere [2].

By defining an incremental polynomial

$$\Delta A(z) = A(z, \lambda + \Delta\lambda) - A(z, \lambda)$$

and substituting for $A(z, \lambda + \Delta\lambda)$ in (8), the approximation $\ln(1 + \delta) \cong \delta$ for small δ gives

$$\Delta G(z) = \Delta A(z)/A(z, \lambda) \quad (11)$$

for sufficiently small incremental parameter changes. The incremental polynomial $\Delta A(z)$ can be evaluated by expanding $A(z, \lambda + \Delta \lambda)$ in a Taylor series and retaining only first-order terms as

$$\Delta A(z) = \sum_{i=1}^L \Delta \lambda_i \frac{\partial A(z, \lambda)}{\partial \lambda_i} \quad (12)$$

with unity gain. The mean-square deviation can be expressed from (9) and (11) as twice a total-squared integral in the form

$$(\Delta d)^2 = 2 \int_{-\pi}^{\pi} \left| \frac{\Delta A(e^{j\theta})}{A(e^{j\theta}, \lambda)} \right|^2 \frac{d\theta}{2\pi} \quad (13)$$

The polynomial $\Delta A(z)$ contains at most the powers z^{-1} through z^{-M} based upon the form of (3). The integral of (13) can be evaluated as a summation involving no more than M terms by two different methods. First, one can use a step-down approach [4] utilizing a set of orthogonal polynomials. The polynomials $\{B_0(z), B_1(z), \dots, B_M(z)\}$ used for the backwards prediction error which are defined by

$$B_0(z) = z^{-1} \quad (14a)$$

and

$$B_m(z) = z^{-1} B_{m-1}(z) + k_m z^{-(m+1)} B_{m-1}(1/z) \quad (14b)$$

for $m = 1, 2, \dots, M$, form an orthogonal set. As shown in Appendix A, these can be used to expand $\Delta A(z)$ in the form

$$\Delta A(z) = \sum_{m=0}^{M-1} \nu_m B_m(z), \quad (15)$$

leading to the mean-square deviation

$$(\Delta d)^2 = 2 \sum_{m=0}^{M-1} \nu_m^2 \alpha_m / \alpha_M, \quad (16)$$

where the coefficients ν_m are defined in Appendix A, and

$$\alpha_m = (1 - k_m^2) \alpha_{m-1} \quad \text{for } m = 1, 2, \dots, M. \quad (17)$$

When using the autocorrelation method of linear prediction [which is assumed here based upon the minimum-phase requirement for $H(z, \lambda)$], α_m is the minimum prediction error energy obtainable with an m th-order predictor, so that α_0 is the input signal energy.

An alternate approach to the evaluation of (13) is similar to that used in the evaluation of likelihood ratios [2]. Let the autocorrelation sequence $\{r_x(n)\}$ represent the autocorrelation sequence for the unit sample response of a filter $\sqrt{\alpha_M}/A(z, \lambda)$. In the autocorrelation method, this sequence is also the autocorrelation sequence for the input data used in calculating $A(z, \lambda)$ [3]. Let the autocorrelation sequence $\{r_{\Delta}(n)\}$ represent the autocorrelation sequence for the polynomial coefficients of $\Delta A(z)$. Applying the discrete form of Parseval's Theorem to the integral of (13) then gives

$$(\Delta d)^2 = \frac{2}{\alpha_M} \sum_{n=-\infty}^{\infty} r_x(n) r_{\Delta}(n).$$

As $\Delta A(z)$ contains at most the powers z^{-1} through z^{-M} , $r_{\Delta}(n)$ must be zero for $|n| > M - 1$. In addition, the autocorrelation sequences are even, so that a second closed-form expression for the mean-square deviation is

$$(\Delta d)^2 = \frac{2}{\alpha_M} \left[r_x(0) r_{\Delta}(0) + 2 \sum_{n=1}^{M-1} r_x(n) r_{\Delta}(n) \right]. \quad (18)$$

Either (16) or (18) can be used to evaluate the mean-square deviation, once the incremental polynomial $\Delta A(z)$ is found. The former expression appears to be most useful for theoretical discussions while the latter expression appears most useful for numerical evaluation using the autocorrelation method of linear prediction since the autocorrelation sequence $\{r_x(n)\}$ is part of the analysis.

B. Single-Parameter Variations ($L = 1$)

Let λ represent either a linear or nonlinear transformation of the single-filter coefficient a_l in (3). From (12),

$$\Delta A(z) = (\Delta \lambda) \frac{\partial A(z, \lambda)}{\partial \lambda} = (\Delta \lambda) \frac{\partial a_l}{\partial \lambda} z^{-l}, \quad (19)$$

so that the autocorrelation coefficients corresponding to

$$R_{\Delta}(z) = [\Delta A(z)] [\Delta A(1/z)]$$

are then

$$r_{\Delta}(n) = (\Delta \lambda)^2 (\partial a_l / \partial \lambda)^2 \delta_{n,0},$$

where $\delta_{n,k}$ is the Kronecker delta. Therefore, from (18), the mean-square deviation is

$$(\Delta d)^2 = 2 [r_x(0)/\alpha_M] (\Delta \lambda)^2 (\partial a_l / \partial \lambda)^2. \quad (20)$$

The ratio $r_x(0)/\alpha_M$ is the ratio of the input signal energy to the prediction error energy and can be expressed in several forms [3] as

$$\begin{aligned} r_x(0)/\alpha_M &= \alpha_0/\alpha_M = \int_{-\pi}^{\pi} \frac{1}{|A(e^{j\theta}, \lambda)|^2} \frac{d\theta}{2\pi} \\ &= \prod_{i=1}^M (1 - k_i^2)^{-1}. \end{aligned} \quad (21)$$

For the case where λ equals the filter coefficient a_l , (20) and (21) show that the spectral deviation due to a change in the coefficient a_l is proportional to the magnitude of the change as

$$\Delta d = \sqrt{2\alpha_0/\alpha_M} |\Delta a_l|.$$

The proportionality constant, $\sqrt{2\alpha_0/\alpha_M}$, is the square root of twice the energy of the unit sample response for the all-pole model, $1/A(z, \lambda)$. As any reflection coefficient approaches unity magnitude (21) becomes unbounded. This result illustrates what has been experimentally known for some time—quantizing the filter coefficients for parameter transmission is unadvisable [4], independent of stability problems.

When λ is a linear or nonlinear function of a single reflection coefficient, k_l , then from (19),

$$\Delta A(z) = (\Delta \lambda) \frac{\partial A(z, \lambda)}{\partial \lambda} = (\Delta \lambda) (\partial k_l / \partial \lambda) [\partial A(z, \lambda) / \partial k_l]. \quad (22)$$

The polynomial $A(z, \lambda)$ can be recursively generated from the reflection coefficients using a step-up procedure [3] which effectively corresponds to the solution for the recursive filter coefficients with unnecessary steps removed. Each individual reflection coefficient appears only once in the recursion, in a linear manner. Starting with

$$A_0(z) = 1, B_0(z) = z^{-1}, \quad (23a)$$

the recursion is specified by

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \quad (23b)$$

and

$$B_m(z) = z^{-1} [k_m A_{m-1}(z) + B_{m-1}(z)] \quad (23c)$$

for $m = 1, 2, \dots, M$, terminating with

$$A(z, \lambda) = A_M(z). \quad (23d)$$

Thus $A(z, \lambda)$ is a linear function of each individual reflection coefficient.

Using the recursion relationships of (23), closed-form results can be obtained for $\partial A(z, \lambda) / \partial k_l$ which increase in complexity as l decreases. For example,

$$\frac{\partial A(z, \lambda)}{\partial k_M} = B_{M-1}(z) \quad (24a)$$

$$\frac{\partial A(z, \lambda)}{\partial k_{M-1}} = B_{M-2}(z) + k_M z^{-1} A_{M-2}(z) \quad (24b)$$

$$\begin{aligned} \frac{\partial A(z, \lambda)}{\partial k_{M-2}} &= (k_{M-1} z^{-1} + k_M z^{-2}) A_{M-3}(z) \\ &+ (1 + k_M k_{M-1} z^{-1}) B_{M-3}(z). \end{aligned} \quad (24c)$$

Since $A(z, \lambda)$ is a linear function of each individual reflection coefficient, the partial derivative of $A(z)$ can be numerically obtained directly from the difference between two polynomials. Let $A(z, \lambda')$ be another polynomial generated from the reflection coefficient set, but with k_l replaced by $k_l + 1$,

$$k_l(\lambda') = k_l(\lambda) + 1,$$

with all other reflection coefficients unchanged. Therefore,

$$\frac{\partial A(z, \lambda)}{\partial k_l} = A(z, \lambda') - A(z, \lambda). \quad (25)$$

For $l = M$ and $l = M - 1$, convenient closed-form results can be obtained for the mean-square deviation $(\Delta d)^2$ by using (24), as shown in Appendix B. These results are presented in Table I. For $l = M - 2$, the closed form results is quite complex in general, and is thus shown in the table only for the case $M = 3$. A common characteristic of the mean-square deviation is that for any l , the mean-square deviation has the form

TABLE I
MEAN-SQUARE DEVIATIONS DUE TO k_l QUANTIZATION

Restrictions	$(\Delta d)^2$
$l = M, \quad M > 0$	$2(\Delta \lambda)^2 (\partial k_M / \partial \lambda)^2 \frac{1}{1 - k_M^2}$
$l = M - 1, \quad M > 2$	$2(\Delta \lambda)^2 (\partial k_{M-1} / \partial \lambda)^2 \left[\frac{1 + 2k_M k_{M-2} + k_M^2}{(1 - k_{M-1}^2)(1 - k_M^2)} \right]$
$l = M - 1 = 1, \quad M = 2$	$2(\Delta \lambda)^2 (\partial k_1 / \partial \lambda)^2 \left[\frac{1 + k_2}{(1 - k_1^2)(1 - k_2)} \right]$
$l = M - 2 = 1, \quad M = 3$	$2(\Delta \lambda)^2 (\partial k_1 / \partial \lambda)^2 \left[\frac{(1 - 2k_1 k_3 + k_3^2)(1 + k_2)}{(1 - k_1^2)(1 - k_2)(1 - k_3^2)} \right]$

$$(\Delta d)^2 = (\Delta \lambda)^2 (\partial k_l / \partial \lambda)^2 f_l(k_1, k_2, \dots, k_M) / (1 - k_l^2), \quad (26)$$

where $f_l(\cdot)$ is a well-behaved function of k_l (it does not go to infinity). It is shown in Appendix C that (26) is a general result. Although $f_l(\cdot)$ does not depend upon k_l for the first three entries in Table I, it is a function of k_l for the last entry. The importance of this relation is that it shows how to define a transformation which removes the singular characteristic at $k_l = \pm 1$.

In order to remove the singularity at $k_l = \pm 1$, it is necessary that $(\partial k_l / \partial \lambda)^2$ be proportional to $(1 - k_l^2)$ near $k_l = \pm 1$. This is accomplished if λ is proportional to the inverse sine of k_l , i.e.,

$$\lambda = c_l \sin^{-1} k_l, \quad (27a)$$

for then

$$k_l = \sin(\lambda / c_l), \quad (27b)$$

and

$$(\Delta d)^2 = (\Delta \lambda / c_l)^2 f_l(k_1, k_2, \dots, k_M).$$

The inverse sine quantization suggested by (27) differs from the log area ratio quantization of [1] because of the gain normalization. In [1] the gain was normalized to yield unity energy

$$\int_{-\pi}^{\pi} |H(e^{j\theta}, \lambda)|^2 \frac{d\theta}{2\pi} = 1. \quad (28)$$

Therefore, from (1) and (21),

$$\sigma(\lambda) = \sqrt{\alpha_M / \alpha_0} = \prod_{i=1}^M (1 - k_i^2)^{1/2}. \quad (29)$$

From (9) we see that there is an additive effect from the gain term, which for incremental variations is given by

$$4[\ln [\sigma(\lambda + \Delta \lambda) / \sigma(\lambda)]]^2 = (\Delta \lambda)^2 \left\{ \frac{\partial}{\partial \lambda} [\ln \sigma^2(\lambda)] \right\}^2.$$

Using this result with (29) gives the additional term in the mean-square deviation as

$$4[\ln \{\sigma(\lambda + \Delta\lambda)/\sigma(\lambda)\}]^2 = (\Delta\lambda)^2 (\partial k_l / \partial \lambda)^2 \frac{4k_l^2}{(1 - k_l^2)^2}.$$

When this result is added to (26), the mean-square deviation takes on the form

$$(\Delta d)^2 = (\Delta\lambda)^2 (\partial k_l / \partial \lambda)^2 g_l(k_1, k_2, \dots, k_M) / (1 - k_l^2)^2, \quad (30)$$

where $g_l(\cdot)$ is a well-behaved function of k_l .

Following the same logic leading to the inverse sine quantization of (27), we find that to eliminate the singularities at $k_l = \pm 1$ in (30), λ must be proportional to the log area ratio

$$\lambda = c_l \ln [(1 + k_l)/(1 - k_l)] \quad (31)$$

near the extremes $k_l = \pm 1$. If k_l is allowed to cover the full range from -1 through $+1$, then the singularities from the denominators of (26) and (30) predominate in the behavior of Δd , leading to either the inverse sine or log area quantization schemes as being optimal. The difference in the two results lies in the choice of the gain terms, not in the choice of a norm (L_1 or L_2).

It should be noted that the "optimality" of either the log area of inverse sine quantization is specifically addressed to finding elementary nonlinear transformations of variables which depend only upon individual reflection coefficients. These transformations remove any singular behavior at the end points, where the magnitudes of the reflection coefficients approach unity.

We prefer to use the gain normalization $\sigma(\lambda) = 1$ which normalizes the log spectra (10), so that the gain term is treated as an independent parameter, as is customary in most speech compression systems. Using this choice, inverse sine quantization is superior to log area quantization when the full range of values for k_l from -1 to $+1$ is covered. In Section III we present experimental results which show that when the reflection coefficients are restricted to more practical ranges, particularly for $l > 2$, the choice of quantization makes little difference, for the singularities at $k_l = \pm 1$ are never approached. We shall see that with the exception of k_1 and k_2 , uniform quantization is quite adequate (and, in fact, is slightly superior!). For k_1 and k_2 , the inverse sine quantization or the log area quantization results are very similar and slightly superior to uniform quantization.

C. Two Parameter Variations ($L = 2$)

Assume that $A(z, \lambda)$ is factored into a product of first-order terms representing real roots and second-order terms representing complex root pairs or pairs of real roots. If parameter variations occur within only a single product term, then all other factors of $A(z, \lambda)$ and $A(z, \lambda + \Delta\lambda)$ will be equal, and will thus cancel in (8). To study such parameter changes, it is, therefore, sufficient to look at first- and second-order polynomials only. First-order polynomials can be treated by applying the first entry in Table I, with $M = 1$. Following the results of the preceding section, it can be seen that the maximum spectral deviation for a first-order polynomial is minimized with inverse sine quantization of $k_1 = a_1$.

The second-order polynomials which can represent either complex or real root pairs are given by

$$\begin{aligned} A(z, \lambda) &= 1 + a_1 z^{-1} + a_2 z^{-2} \\ &= 1 + k_1(1 + k_2)z^{-1} + k_2 z^{-2}. \end{aligned} \quad (32)$$

When the roots are complex conjugates occurring at $re^{\pm j\phi}$ then

$$a_1 = -2r \cos(\phi), \quad a_2 = r^2, \quad (33a)$$

and

$$k_1 = -2r \cos(\phi)/(1 + r^2), \quad k_2 = r^2. \quad (33b)$$

Assuming that the row vector λ has elements λ_1 and λ_2 , all spectral deviations can be expressed in terms of Δk_1 and Δk_2 as

$$\Delta k_1 = (\partial k_1 / \partial \lambda_1)(\Delta \lambda_1) + (\partial k_1 / \partial \lambda_2)(\Delta \lambda_2) \quad (34a)$$

and

$$\Delta k_2 = (\partial k_2 / \partial \lambda_1)(\Delta \lambda_1) + (\partial k_2 / \partial \lambda_2)(\Delta \lambda_2). \quad (34b)$$

From (12) and (32), with the reflection coefficients as parameters, the incremental polynomial is found to be

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

In terms of the orthogonal polynomials defined by (14), the incremental polynomial has the form

$$\Delta A(z) = (1 + k_2) B_0(z)(\Delta k_1) + B_1(z)(\Delta k_2).$$

As in the derivation of (16), the orthogonality of the polynomials $B_m(z)$ leads to a mean-square deviation given by

$$(\Delta d)^2 = 2(1 + k_2)^2 (\Delta k_1)^2 (\alpha_0/\alpha_2) + 2(\Delta k_2)^2 (\alpha_1/\alpha_2).$$

By applying (17), this expression takes on the form

$$(\Delta d)^2 = 2 \frac{1 + k_2}{1 - k_2} \frac{(\Delta k_1)^2}{1 - k_1^2} + 2 \frac{(\Delta k_2)^2}{1 - k_2^2}. \quad (35)$$

This result can be used to study the effect of pole location quantization in terms of resultant spectral deviation, and to obtain an incremental lower bound on the minimum number of bits needed to quantize any pole pair for a maximum allowable spectral deviation.

D. Formant Movement and Spectral Deviation

A problem of basic research interest is to know how the spectral deviation is related to formant movement. Equation (35) can be directly applied to solve this problem for single-formant movement. The frequency F and bandwidth B of a formant is defined in terms of a complex pole pair $r \exp(\pm j\phi)$ with sampling frequency f_s as

$$r = e^{-\pi B/f_s}, \quad \phi = 2\pi F/f_s. \quad (36)$$

Equations (33b) and (34) can be used with (35) to express the mean-square deviation in terms of the pole location incremental changes. After some tedious trigonometric manipulations, the mean-square deviation takes on the form

$$(\Delta d)^2 = [u_{rr}(\Delta r)^2 - 2u_{r\phi}(\Delta r)(\Delta \phi) + [u_{\phi\phi}(\Delta \phi)^2] / [(1 - k_1^2)(1 - k_2^2)]$$

where

$$u_{rr} = [(r^2 - 1)^2 + (3r^2 - 1) \sin^2 \phi] / (1 + r^2),$$

$$u_{r\phi} = r \sin(\phi) \cos(\phi) (1 - r^2) / (1 + r^2),$$

$$u_{\phi\phi} = r^2 \sin^2 \phi,$$

and

$$(1 - k_1^2)(1 - k_2^2) = [(1 - r^2)^2 + 4r^2 \sin^2 \phi] (1 - r^2) / (1 + r^2)^2.$$

As the formant behavior is highly resonant [5], $B \ll F$ and the above expressions can easily be approximated by using the assumption

$$(1 - r^2) = 2\pi B / f_s \ll \sin(\phi) = \sin(2\pi F / f_s), \quad (37)$$

giving

$$(\Delta d)^2 \cong \frac{2f_s}{\pi B} [(r\Delta\phi)^2 + (\Delta r)^2], \quad (38a)$$

or

$$\Delta d \cong |\Delta z_p| \sqrt{2f_s / \pi B}, \quad (38b)$$

where $|\Delta z_p|$ is the actual incremental distance in the z plane for pole motion resulting from incremental parameter changes.

In terms of single-formant movement, the spectral deviation is found to be related to changes in formant movement ΔF and formant bandwidth ΔB by

$$(\Delta d)^2 = \frac{8\pi}{Bf_s} [(\Delta F)^2 + (\Delta B/2)^2]. \quad (39)$$

From (38) it is seen that the spectral deviation resulting from an incremental change in a narrow bandwidth pole location is directly proportional to the distance moved and to the square root of the ratio of sampling frequency to bandwidth. Increasing the sampling frequency while holding the bandwidth fixed increases Δd , a result which can intuitively be seen by noting that the poles are closer to the unit circle. From (39) it can be noted that for a fixed bandwidth $\Delta B = 0$, Δd is proportional to the absolute change in formant frequency $|\Delta F|$. This result can be compared to the perceptual experiments of Stevens [5] where distance limens (values of ΔF which resulted in barely perceptible changes in sounds) were essentially constant for one formant synthesis models. Using Stevens' bandwidths of 130 Hz and a difference limen of 20 Hz, a typical spectral deviation for a 6.5-kHz sampling frequency corresponding to one distance limen is $4.34 (\Delta d) = 0.47$ dB.

While (38) and (39) lend insight into how formant movement relates to spectral deviation, it must be kept in mind that the results are based upon the assumption of (37). In considering general pole-pair quantization where wide bandwidth shaping terms are used, for example, it is theoretically preferable to

return to (35) and the reflection coefficients, and to use the two-parameter coding scheme which is now described.

E. Minimum Bits to Implement a Pole Pair

Let $\Delta \bar{d}$ represent an upper limit to be allowed on the spectral deviation arising from quantization errors in a single pole pair. We wish to determine how to quantize the pole pair of (32) so that the range of reflection coefficients

$$-1 \leq k_1 \leq 1, \quad -1 \leq k_2 \leq \bar{k}_2$$

is covered by the smallest number of quantization points, without exceeding a maximum spectral deviation $\Delta \bar{d}$ due to quantization error, where $\bar{k}_2$ is the maximum allowable value of k_2 based upon values observed from speech. For complex pole pairs, $\bar{k}_2$ is the square of the maximum pole magnitude, and is related to a minimum allowable bandwidth B by

$$\bar{k}_2 = e^{-2\pi B / f_s}.$$

If β is the number of bits used for quantization, we shall show that an incremental lower bound, valid for very small $\Delta \bar{d}$, is given by

$$2^\beta \geq \pi \frac{\ln [2/(1 - \bar{k}_2)]}{(\Delta \bar{d})^2}. \quad (40)$$

In addition, a quantizing scheme which approximately satisfies the lower bound is derived.

The first step is to change variables using inverse sine transformations

$$\psi_i = \sin^{-1} k_i \quad \text{for } i = 1, 2, \quad (41)$$

so that (35) can be rewritten as

$$(\Delta d)^2 = 2 \frac{1 + \sin(\psi_2)}{1 - \sin(\psi_2)} (\Delta \psi_1)^2 + 2(\Delta \psi_2)^2. \quad (42)$$

Using this transformation, we see that (42) describes incremental ellipses. The locus of points which satisfies the inequality $\Delta d \leq \Delta \bar{d}$ is then given by the interiors of the ellipses, centered at ψ_1, ψ_2 . Two such ellipses are shown in Fig. 1. Their principal axes are parallel to the coordinate axes. The heights of the ellipses are equal to $\sqrt{2}(\Delta \bar{d})$. The ellipse widths depend upon the vertical coordinate of the ellipse centers, and are given by

$$w = \sqrt{2}(\Delta d) \{ [1 - \sin(\psi_2)] / [1 + \sin(\psi_2)] \}^{1/2}.$$

These ellipses are assumed to be incrementally small, so that the value of ψ_2 used in determining the width can be taken as the center value, since it differs only incrementally within the ellipse. More precise curves would taper towards the top, giving an egg-shaped appearance. Using more precise curves only slightly modifies the end results while substantially increasing the complexity of the derivation.

In order that the rms deviation due to quantization error be less than $\Delta \bar{d}$, it is necessary that every point in the range

$$-\pi/2 \leq \psi_1 \leq \pi/2, \quad -\pi/2 \leq \psi_2 \leq \sin^{-1}(\bar{k}_2) \quad (43)$$

lie within at least one such ellipse, centered at a quantization

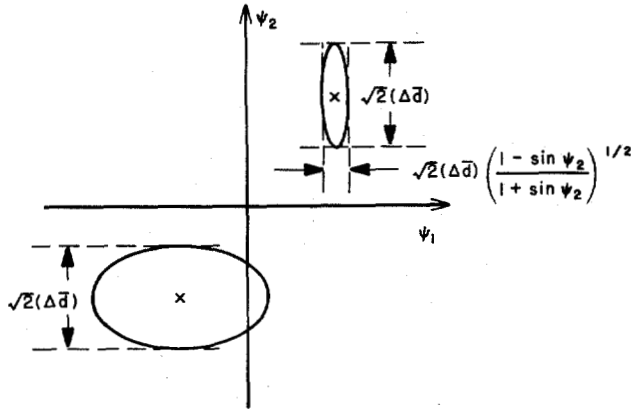


Fig. 1. Illustration showing the locus of points satisfying $\Delta d \leq \Delta \bar{d}$ in the ψ_1, ψ_2 plane.

point. As the ellipses do not have rectangular boundaries, there will be some overlap, effectively wasting area. To minimize the waste, we obtain the largest possible rectangle (in terms of area) within an ellipse. This maximization problem leads to rectangles with dimensions as shown in Fig. 2. All numerical values that fall within the rectangle are represented by the single numerical value corresponding to its center. The shaded region between the rectangle and the ellipse indicates the wasted area that could be assigned to more than one quantization point.

To cover the full range of ψ_1 and ψ_2 , values by quantization points without exceeding a spectral deviation of $\Delta \bar{d}$, we can divide the plane into horizontal strips with heights not less than the rectangle height $\Delta \bar{d}$ of Fig. 2. From the range of ψ_2 value of (43), it is seen that at least N_s strips are required where

$$N_s = [\sin^{-1}(\bar{k}_2) + \pi/2] / (\Delta \bar{d}). \quad (44)$$

Within each strip, numbered $n = 1, 2, \dots, N_s$, the vertical ordinate for the quantization points (or rectangle centers) is given by

$$\psi_2(n) = -\frac{\pi}{2} + \left(n - \frac{1}{2}\right) (\Delta \bar{d}), \quad \text{for } n = 1, 2, \dots, N_s. \quad (45)$$

From (43), the width of each strip is π . With reference to Fig. 2, the number of quantization points along a strip, $N(n)$, must at least equal the strip width divided by the rectangle width, so that $N(n)$ must at least equal

$$N(n) = \frac{\pi}{\Delta \bar{d}} \left[\frac{1 + \sin [\psi_2(n)]}{1 - \sin [\psi_2(n)]} \right]^{1/2}.$$

The horizontal ordinate for the quantization points is then given by

$$\psi_1(n, l) = -\frac{\pi}{2} + \left(l - \frac{1}{2}\right) \frac{\pi}{N(n)} \quad \text{for } l = 1, 2, \dots, N(n). \quad (46)$$

The total number of quantization points must at least equal the sum

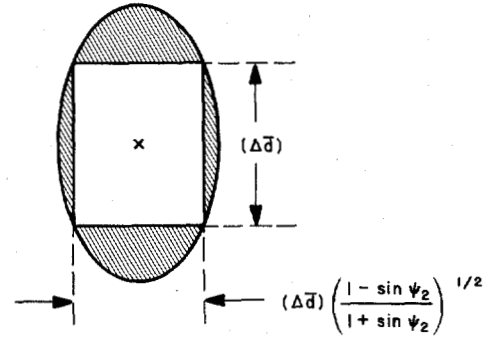


Fig. 2. Illustration showing the largest rectangle fitting within an ellipse.

$$N = \sum_{n=1}^{N_s} N(n) = \frac{\pi}{\Delta \bar{d}} \sum_{n=1}^{N_s} \left[\frac{1 + \sin [\psi_2(n)]}{1 - \sin [\psi_2(n)]} \right]^{1/2}. \quad (47)$$

From the incremental nature of $\Delta \bar{d}$, this summation can be approximated by an integral, or by slightly changing the limits, (47) can be bounded by an integral. Using (45) to define

$$\Delta \psi_2 = \psi_2(n+1) - \psi_2(n) = \Delta \bar{d},$$

(47) can be rewritten as

$$N = [\pi / (\Delta \bar{d})^2] \sum_{n=1}^{N_s} g[\psi_2(n)] \Delta \psi_2,$$

where

$$g(\psi) = \left[\frac{1 + \sin(\psi)}{1 - \sin(\psi)} \right]^{1/2}.$$

The integral approximation to the summation is then

$$\begin{aligned} \sum_{n=1}^{N_s} g[\psi_2(n)] \Delta \psi_2 &\cong \int_{-\pi/2}^{\sin^{-1}(\bar{k}_2)} g(\psi) d\psi \\ &= \ln [2 / (1 - \bar{k}_2)]. \end{aligned}$$

This result can be combined with (48) to give the bound of (40) where $N = 2^B$. It is a bound, because in each step of the derivation the minimum number of quantization levels was used. If, for example, (44) specifies 23.7 strips, then at least 24 strips must be used.

The transformation that approximately attains equality in (40) is given by uniform quantization along the ψ_2 axis and nonuniform quantization along the ψ_1 axis, where the nonuniformity is specified by Fig. 2. Equivalently, the transformation is given by uniform quantization of

$$\lambda_2 = \sin^{-1}(\bar{k}_2) \quad (48a)$$

$$\lambda_1 = \left[\frac{1 + \bar{k}_2}{1 - \bar{k}_2} \right]^{1/2} \sin^{-1}(\bar{k}_1). \quad (48b)$$

This end result has apparently also been developed by Itakura from the likelihood function of an all-pole stochastic signal [6].

III. EXPERIMENTAL RESULTS

The sentences "Thieves who rob friends deserve jail," "Open the crate but don't break the glass," "Oak is strong and also gives shade," and "It's time we rounded up that herd of Asian cattle," each spoken by a different person, were analyzed by the autocorrelation method of linear prediction. The fundamental frequency ranged from 50 Hz to 220 Hz. The analysis windows of 100 samples (15.38 ms) moved every 10 ms for a total of 912 analysis frames (9.12 s). A constant preemphasis coefficient of 0.9 was applied to all frames.

A. Single-Parameter Variations

For determining bit allocation based upon single-parameter variations, the following approach was used. For each frame a set of 10 reflection coefficients was obtained. Histograms were made of the frequency of occurrence of the reflection coefficient over all reflection coefficients for determining reasonable ranges for each coefficient value. The histograms incorporate all of the speech sounds but silence. The bit-allocation procedure to be described can be directly applied to finding efficient quantization for the separate categories of voiced and unvoiced sounds at the cost of additional storage for quantization tables. The results are shown in Fig. 3 for the 10 reflection coefficients. The first two coefficients show a very strong skewing effect towards -1 and $+1$, respectively. This characteristic has been theoretically shown for voiced sounds having high correlation [7]. The bimodal distribution of k_2 is due to the inclusion of unvoiced sounds which results in a high frequency of occurrence near $k_2 = 0$. The remaining histograms are reasonably well modeled by Gaussian probability density curves as shown. The maxima and minima of these parameter distributions are shown in Table II, along with means and standard deviations. By combining (12), (13), and (25), the expression for $(\Delta d)^2$ and $L = 1$ can be written in a normalized form:

$$I_1 = 2 \int_{-\pi}^{\pi} \left| \frac{A(e^{j\theta}, \lambda') - A(e^{j\theta}, \lambda)}{A(e^{j\theta}, \lambda)} \right|^2 \frac{d\theta}{2\pi}$$

$$= \frac{[\Delta d(l)]^2 (\partial \lambda / \partial k_l)^2}{(\Delta \lambda)^2} \quad (49)$$

where $\Delta d(l)$ defines the spectral deviation due to a single-parameter variation. The normalized spectral deviation is computed from (49) using the computer program in Appendix D. The transformations and partial derivatives for the three cases of interest here are as follows:

reflection coefficient quantization:

$$\lambda = k_l, \quad \partial k_l / \partial \lambda = 1;$$

log area quantization:

$$\lambda = \ln \left(\frac{1 + k_l}{1 - k_l} \right), \quad \partial k_l / \partial \lambda = (1 - k_l^2)/2; \text{ and}$$

inverse sine quantization:

$$\lambda = \sin^{-1}(k_l), \quad \partial k_l / \partial \lambda = \sqrt{1 - k_l^2}.$$

Given the min-max values $\underline{k}_l, \bar{k}_l$ from Table II, $\underline{\lambda}_l$ and $\bar{\lambda}_l$ are

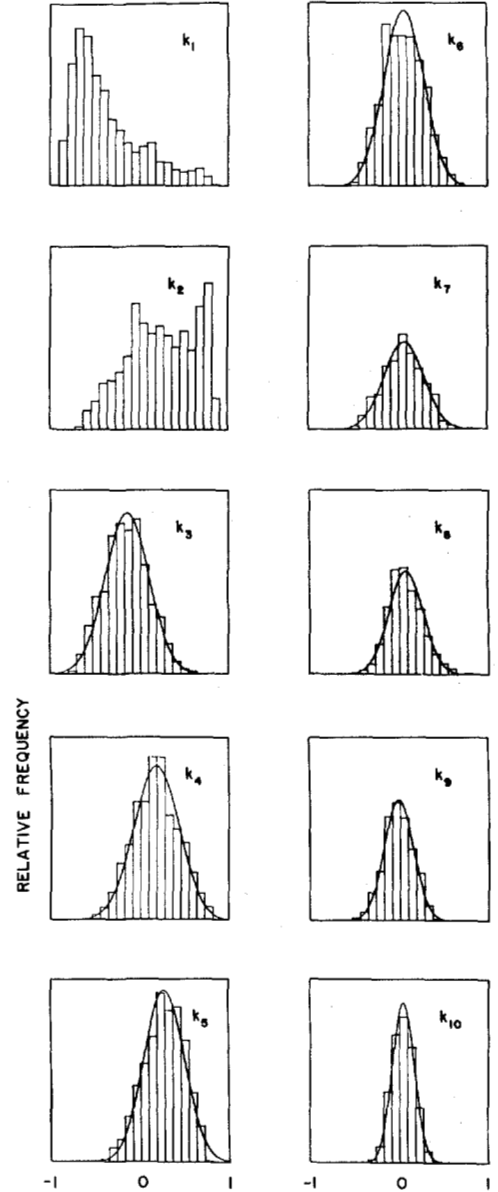


Fig. 3. Histograms of reflection coefficients.

computed from the above expressions. Linear quantization of the parameter λ into N_l levels (or $\log_2(N_l)$ bits) produces a maximum quantization error of $\Delta \bar{\lambda} = (\bar{\lambda}_l - \underline{\lambda}_l)/2N_l$ assuming that rounding is used. Therefore, the maximum value of the mean-square spectral deviation is given by

$$[\Delta d(l)]^2 = \left(\frac{\bar{\lambda}_l - \underline{\lambda}_l}{2N_l} \right)^2 \left(\frac{\partial k_l}{\partial \lambda} \right)^2 I_l.$$

The resultant scatter plots of normalized spectral deviation versus k_l for the above quantization methods are presented in Fig. 4. Each graph shows the quantity $2(4.34 \Delta d(l))N_l$ (decibels) versus k_l over the range -1 to 1 . For each index the corresponding plots are scaled identically. For discussion purposes, the graphs will be referenced in the form of matrix elements as G_{ij} where $i = 1, 2, \dots, 10$ defines the coefficient index and $j = 1, 2, 3$ corresponds to reflection coefficient, log

TABLE II
REFLECTION COEFFICIENT DISTRIBUTION

i	Minimum Reflection Coefficient k_i	Maximum Reflection Coefficient k_i	Mean	Standard Deviation
1	-0.9729	0.9243	-0.40630	0.4334
2	-0.7288	0.9503	0.26920	0.4219
3	-0.8774	0.6411	-0.17430	0.2721
4	-0.5789	0.8622	0.18550	0.2819
5	-0.5237	0.7200	0.05539	0.2470
6	-0.4762	0.8539	0.28130	0.2517
7	-0.6444	0.8049	0.05886	0.2487
8	-0.5146	0.8562	0.07080	0.2121
9	-0.5876	0.4532	-0.02784	0.1801
10	-0.3972	0.5099	0.03665	0.1366

area, and inverse sine quantization, respectively. On the right side of each graph two numbers are shown. The top number is the largest value computed for G_{ij} over all 912 frames (rounded to an integer value). Only one percent of the frames produce values which exceed the lower number.

In G_{11} a u-shaped characteristic is easily seen for k_1 near -1. The peak deviation is $4.34\Delta d = 277/(2N)$ dB. To insure that the spectral deviation over all frames in the data base does not exceed 0.3 dB due to quantization of k_1 requires $N_1 = 462$ or 9 bits. If it is required that the spectral deviation not exceed 0.3 dB more than one percent of the time, then $N_1 \geq 248$ or 8 bits. In G_{12} and G_{13} , the peak spectral deviation is substantially reduced. Although G_{12} shows a greater reduction of spectral deviation near $|k_1| = 1$ than G_{13} , it does so at the expense of slightly increased errors away from these regions. Both spectral deviation values are slightly less for G_{13} . By quantizing with $j = 2$ or $j = 3$ for either a peak or one percent exceeded criterion, it is seen that the quantization requirements are reduced by approximately one bit for the first coefficient, with respect to reflection coefficient coding.

For $i = 2$, G_{21} shows a moderate amount of u-shaped behavior near $k_2 = 1$. Again G_{22} shows a substantial reduction in spectral deviation near the boundaries $|k_2| = 1$, but slightly higher peak and one percent values than for G_{23} . For $i > 2$ the possibly surprising result is seen that log area and inverse sine quantization are theoretically inferior to direct reflection coefficient quantization. For example, to insure that the spectral deviation due to quantization of k_3 does not exceed 0.3 dB, log area quantization requires $N_3 \geq 136$ or 7.08 bits whereas reflection coefficient quantization requires $N_3 \geq 102$ or 6.67 bits. For inverse sine quantization, $N_3 \geq 115$ or 6.84 bits. The reason for this behavior is that for G_{12} or G_{13} , higher resolution is guaranteed as $|k_i|$ increases towards unity at the expense of less resolution for $|k_i|$ near zero. A graph of $(\lambda/\bar{\lambda})$ versus k for $0 \leq k \leq \bar{k}$ will show that log area quantization has the most extreme curvature from the $\lambda = k$ line and, therefore, the highest resolution for $|k|$ near one and lowest resolution for $|k|$ near zero. From a practical viewpoint, it can be seen that for $i > 2$, very little difference exists among the quantization schemes.

It is interesting to note that beyond $i = 6$, a deterministic trend begins to appear in each respective graph G_{ij} . At the

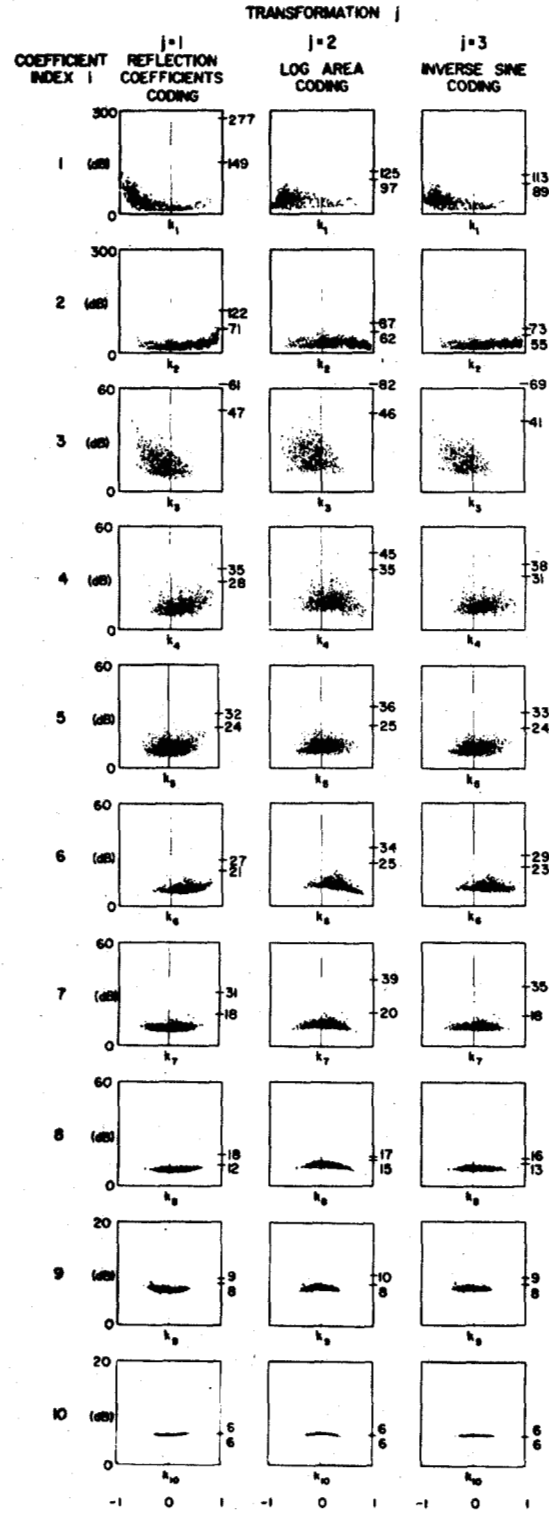


Fig. 4. Scatter plots for normalized spectral deviation.

final coefficient $i = M = 10$, the variations of the data become deterministic. This behavior was predicted by the first entry of Table I as

$$[\Delta d(M)]^2 = 2(\Delta\lambda)^2 (\partial k_M / \partial \lambda)^2 (1 - k_M^2)^{-1}.$$

Therefore,

$$[\Delta d(M)]^2 = \begin{cases} \sqrt{2} \left(\frac{\bar{\lambda}_M - \lambda_M}{2N_M} \right) (1 - k_M^2)^{-1/2} & \text{reflection coefficient quantization} \\ \sqrt{2} \left(\frac{\bar{\lambda}_M - \lambda_M}{4N_M} \right) (1 - k_M^2)^{1/2} & \text{log area quantization} \\ \sqrt{2} \left(\frac{\bar{\lambda}_M - \lambda_M}{2N_M} \right) & \text{inverse sine quantization.} \end{cases}$$

Whereas $\Delta d(M)$ becomes unbounded as k_M approaches unity for reflection coefficient quantization, it approaches zero for log area quantization and is constant for inverse sine quantization. The constant line for inverse sine quantization intersects the convex and concave functions from the reflection coefficient and log area quantization.

Specific bit allocations are obtained from Fig. 4 by the following procedure. First, a maximum decibel rms spectral deviation $\max(\Delta d)$ is chosen along with a criterion of 1) never being exceeded (or more precisely for the number of frames of data, approximately 0.1 percent of the time) or 2) being exceeded 1.0 percent of the time. Applying the triangle inequality, we know that the overall maximum spectral deviation satisfies

$$\max(\Delta d) \leq \sum_{l=1}^M \Delta \bar{d}(l).$$

The minimum spectral deviation for a given number of levels or bits is obtained by choosing each of the $\Delta \bar{d}(l)$ terms to be equal. Therefore, choose

$$\Delta \bar{d}(l) = \max(\Delta d)/M.$$

If an overall maximum spectral deviation of 3 dB is desired, with 10 coefficients, then $\Delta \bar{d}(l) = 0.3$ dB. The number of levels N_l for each coefficient is defined by the peak or one percent number for the specified G_{ij} divided by $2\Delta d(l)$. The required bit allocations for each quantization method based upon peak values and $\max(\Delta d) = 3$ dB and Fig. 4 are given in Table III. The number of levels, fractional bits and integer number of bits obtained by arbitrarily rounding upward if the fraction exceeds 0.25, are presented. Beyond $l > 2$ all three quantization methods are essentially equivalent, although for a detailed comparison the direct reflection coefficient quantization is superior for $l > 2$. For $l = 1$ and 2, one additional bit is saved by using either log area or inverse sine quantization. The inverse sine quantization for $l \leq 2$ is seen to be slightly superior to log area quantization and significantly superior to the reflection coefficient quantization.

It is important to note that a halving of the spectral deviation requires one additional bit per parameter or M total bits. Conversely, a reduction of one bit per parameter (for example, a reduction to 49 bits) will double the decibel spectral deviation.

It should be emphasized that the results presented are for

sampling frequencies in the vicinity of 6.5 kHz and analysis using the autocorrelation method with preemphasis near 0.9. The normalized bit allocations described can be used directly to find the bit allocation and necessary number of bits to guarantee that no frames or one percent of the frames analyzed (statistically speaking) will exceed a specified maximum spectral deviation. If no preemphasis is used, somewhat different results are obtained due to the fact that a significant percentage of the first few reflection coefficients will lie closer to unity magnitude. An important practical implication is that for the same maximum rms spectral deviation, a larger number of bits will generally be necessary and, in addition, a larger percentage of the available bits will have to be allocated to the first few coefficients.

Similar statements hold for higher sampling frequencies. For a maximum desired spectral deviation, a larger (although possibly not significant) number of bits will have to be allocated. The bit allocation will also change in the direction of requiring a larger percentage of the total bits to be used for the first few coefficients. For any desired analysis condition it is only necessary to analyze a reasonably long segment of data such as 5–10 s with representative speech behavior, and then perform the same straightforward computation for N_l based upon the maximum overall allowable rms spectral deviation.

B. Two-Parameter Variations

In this section, the two-parameter quantization scheme is compared with quantizing methods of the previous section. For each of the 912 frames of data, the polynomial roots z_i , $i = 1, 2, \dots, M$, were obtained from $A(z)$. Since M is even ($M = 10$), second-order polynomial pairs can be constructed from the real and imaginary parts of the roots in terms of reflection coefficients. Assuming $z_i = x_i + jy_i$,

$$\left. \begin{aligned} k_2 &= (x_i)^2 + (y_i)^2 \\ k_1 &= -2x_i/(1 + k_2) \end{aligned} \right\} z_i \text{ complex}$$

$$\left. \begin{aligned} k_2 &= x_i x_j \\ k_1 &= -(x_i + x_j)/(1 + k_2) \end{aligned} \right\} z_i, z_j \text{ real,}$$

or every frame of analysis data, each pair is ordered from highest to lowest k_2 . Since $k_2 = \exp(-2\pi B/f_s)$ for complex conjugate root pairs, the ordering corresponds to ranking pairs in order of increasing bandwidth. The parameter transformation used is

$$\lambda_2 = \sin^{-1}(k_2)$$

$$\lambda_1 = \left[\frac{1 + k_2}{1 - k_2} \right]^{1/2} \sin^{-1}(k_1).$$

For a given maximum $\Delta \bar{d}$, this transformation will result in the minimum number of quantization levels over all possible two-parameter transformations. Scatter plots of the root pairs, reflection coefficient pairs (k_1, k_2) , and two-parameter quantization pairs (λ_1, λ_2) for each second-order polynomial are shown in Fig. 5(a)–(c), respectively. For $l = 1, 2$ in Fig. 5(a), the roots are seen to cluster near the unit circle. For $l = 4, 5$, the samples appear to be somewhat randomly scattered around

TABLE III
BIT ALLOCATIONS BASED UPON SPECTRAL DEVIATION GRAPHS AND A MAXIMUM DEVIATION NOT TO EXCEED 3 dB

Index l	Reflection Coefficient Quantization			Log Area Quantization			Inverse Sine Quantization		
	N_l	β (exact)	β (rounded)	N_l	β (exact)	β (rounded)	N_l	β (exact)	β (rounded)
1	462	8.85	9	208	7.70	8	188	7.55	8
2	203	7.66	8	145	7.18	7	121	6.92	7
3	101	6.66	7	136	7.09	7	115	6.84	7
4	58	5.86	6	75	6.23	6	63	5.98	6
5	53	5.73	6	60	5.91	6	55	5.78	6
6	45	5.49	6	56	5.81	6	48	5.58	6
7	51	5.67	6	65	6.02	6	58	5.86	6
8	30	4.91	5	28	4.81	5	26	4.70	5
9	15	3.91	4	16	4.00	4	15	3.91	4
10	10	3.32	4	10	3.32	4	10	3.32	4
Total		58.0	61		58.1	59		56.4	59

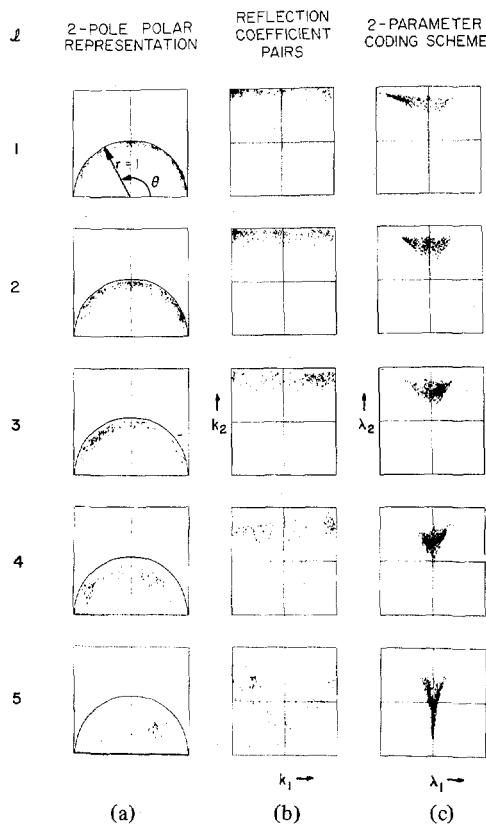


Fig. 5. Scatter plots of parameters for two-pair quantization. (a) Pole pairs, (b) reflection coefficients, and (c) two-parameter quantization pairs λ_1, λ_2 .

the interior of the unit circle. In Fig. 5(b), the values along the k_1 axis appear somewhat randomly distributed over the range $-1 < k_1 < 1$. The k_2 axis values, however, tend to lie within restricted ranges of k_2 with largest values of k_2 occurring for the first few ordered pairs. For the first pair ($l=1$), there is a concentration of samples near $k_2 = +1$ and $k_1 = -1$. In this region, $k_1 \approx -\cos(\phi)$ for complex pole pairs and $k_1 = -(z_i + z_j)$ for real pole pairs. This concentration is thus primarily due to the large number of narrow bandwidth terms at, or near, zero frequency. For $-1 < k_1 < 1$, the points indicated correspond to the narrowest bandwidth formant along the warped fre-

quency scale $\cos(\phi)$ for $0 < \phi < f_s/2$. Without preemphasis, very few samples would exist beyond $k_1 = 0$ (corresponding to $f_s/4$) since the narrowest bandwidths will usually be attributed to the highest energy terms. For the remaining pairs, there do not appear to be any particularly clustered regions along the warped frequency axis.

In Fig. 5(c) the data seem to lie approximately within horizontal strips bounded by a wedge. Since $|k_1| < 1$, the wedge is defined by

$$\lambda_1 = \pm \frac{\pi}{2} \left[\frac{1 + \sin \lambda_2}{1 - \sin \lambda_2} \right]^{1/2}$$

If the λ_1, λ_2 plane is uniformly quantized in both dimensions with quantization intervals of length Δd , we are assured that the spectral deviation due to quantization of any second-order polynomial (using this quantization scheme) will not exceed $4.34 \Delta d$ (decibels). These figures illustrate the important fact that as k_2 decreases (the bandwidths increase) fewer quantization levels are necessary to represent the log spectrum with the same degree of accuracy. As λ_2 approaches unity (zero-bandwidth), the necessary number of levels becomes unbounded. It is interesting to note that the tightly clustered samples about $k_1 = -1$ and $k_1 = 1$ in Fig. 5(b) (for $l=1$) are spread out over a fairly wide range in Fig. 5(c). From Section II, the necessary number of levels to insure that $\Delta \bar{d}$ is not exceeded is easily generalized to allow for a maximum range of $\underline{k} \leq k_2 \leq \bar{k}$ with the result

$$N \geq \frac{\pi}{(\Delta \bar{d})^2} \ln \left[\frac{1 - \underline{k}}{1 - \bar{k}} \right] \quad \text{for } \underline{k} \leq k_2 \leq \bar{k}.$$

Using the same criterion of choosing extrema values for $\underline{k}$ and $\bar{k}$ as in the previous section, and a specified maximum overall spectral deviation, we can compare the different quantization methods theoretically. For a maximum total spectral deviation of 3 dB, $\Delta \bar{d}$ for each pair must not exceed 0.6 dB or $\Delta \bar{d} = 0.1382$. Using the above expression and the k_2 extrema measured from the data, results are summarized in Table IV. If 45 bits are uniformly allocated over the five ordered pairs with their corresponding extrema ranges, by making use of the triangle inequality, we are assured that the total rms log spectral deviation between the original spectrum and the spectrum

TABLE IV
BIT ALLOCATION FOR ORDERED PAIRS USING TWO-PARAMETER
QUANTIZATION METHOD

l	$\bar{k}_2$	k_2	N	Bits	[Bits]
1	0.976	0.7	417	8.70	9
2	0.960	0.6	379	8.57	9
3	0.930	0.3	379	8.57	9
4	0.920	-0.3	462	8.85	9
5	0.710	-0.8	300	8.23	9
Totals				42.92	45

reconstructed from quantized parameters will not exceed 3 dB. If the total range $\bar{k} \leq k_2 \leq \bar{k}$ were quantized for each pair without ordering terms, the required number of bits would increase to 50. From (40) it can be seen that if the maximum spectral deviation Δd is doubled, the required number of bits per pole pair is reduced by two. In the one-parameter quantization methods, the required number of bits was only reduced by one per parameter. However, since only $M/2$ pairs are quantized, all methods described have the same overall decibel versus $\max(\Delta d)$ behavior.

The improvement of this quantization method over reflection, inverse sine or log area coefficient quantization is significant. For the same maximum overall rms log spectral deviation of 3 dB, the other methods all require approximately 59 bits. Assume representative values for a linear prediction vocoder as 5 bits for pitch, 5 bits for gain, and a fixed transmission frame rate of 50 Hz, with 10 coefficients. Using a criterion of quality as not exceeding an rms log spectral deviation of 3 dB, with all else equal, reflection coefficient, log area, or inverse sine quantization require 3500 bits/s whereas the two-parameter quantization scheme required only 2800 bits/s. When the quantization efficiency of only the quantized 10 coefficients is considered, a savings of 14 bits per frame can be obtained without any sacrifice in quality. The price paid for this increased efficiency is the requirement of polynomial root solving. Although encoding/decoding details have not been presented here, they are relatively straightforward and do not require complex numerical operations or excessive computer storage.

C. Bit Allocation for Fixed Bandwidth Formant Vocoder

Based upon the theoretical developments of Section II, we can also determine the necessary number of levels or bits for satisfying a maximum spectral deviation in formant vocoders. The procedure is illustrated here for the fixed bandwidth case.

For fixed bandwidths, $\Delta B_n = 0$ so that (39) reduces to

$$\Delta d = \sqrt{\frac{8\pi}{B_n f_s}} |\Delta F_n|.$$

Now if F_n is restricted to the range $(F_n, \bar{F}_n)$ and N_n levels are used to quantize F_n , the maximum frequency error ΔF_n will be

$$\Delta F_n = \frac{\bar{F}_n - F_n}{2N_n}.$$

The maximum spectral deviation is then

$$\Delta \bar{d} = \sqrt{\frac{8\pi}{B_n f_s}} \left[\frac{\bar{F}_n - F_n}{2N_n} \right].$$

Typical ranges for F_n $n = 1, 2, 3$ are given in Table IV. To guarantee less than 3 dB total maximum spectral deviation,

$$\max(\Delta d) < \Delta d(1) + \Delta d(2) + \Delta d(3) = 0.690775$$

or $\Delta \bar{d} = 0.23026$. Therefore, the necessary number of levels is

$$N_n \geq \frac{1}{\Delta \bar{d}} \sqrt{\frac{8\pi}{B_n f_s}} \left[\frac{\bar{F}_n - F_n}{2} \right].$$

The assumed min-max range of formant frequencies and fixed bandwidth values, along with the minimum number of bits $\log_2(N_1 N_2 N_3)$ and integer number of bits for each parameter to insure that $4.34 \max(\Delta d) < 3$ dB, is shown in Table V. When rounded, the allocation corresponds to five bits per parameter. Assuming a 50-Hz frame rate and pitch and gain quantized to 6 bits and 5 bits, respectively, the formant vocoder bit rate would be 1300 bits/s for the same quality, as the other systems. The subjective quality, however, will be considerably less than that of the other systems simply because of the fixed bandwidth definition and implied fixed spectral slope for glottal driving function and lip radiation characteristics.

IV. CONCLUSIONS

A theoretical basis for studying the effects of quantization and bit-allocation methods in speech processing has been developed and applied to speech data. The three most widely used parameters in linear prediction vocoder systems are the reflection coefficients, log area coefficients, and inverse sine coefficients. It was shown that in the sense of minimizing a maximum spectral deviation, either the log area ratios or inverse sine coefficients could be considered optimal, depending on the definition of a particular gain term. Although either of these quantization methods is superior to direct reflection coefficient quantization for the first two coefficients, the surprising result is shown that reflection coefficients are either nearly equal or superior to log-area coefficients or inverse sine coefficients beyond the first two terms. Using the number of levels as specified in the section on experimental results for insuring against a specified overall maximum spectral deviation, it was informally verified that the synthesized speech using each quantization method was of approximately the same perceived quality.

A two-pair quantization method was derived which approximately satisfies a lower bound on the necessary number of levels to insure that a specified maximum spectral deviation is not exceeded. Based upon use of the triangle inequality, the two-pair quantization scheme was shown to be significantly better than the reflection coefficient, log area, or inverse sine quantization at the expense of root solving.

In all of the quantization methods discussed, it is necessary to apply the triangle inequality for insuring against an overall maximum spectral deviation since only one-parameter or two-parameter variations can be theoretically studied in a useful manner. Presently, we are performing analysis/synthesis ex-

TABLE V
BIT ALLOCATION FOR FIXED BANDWIDTH FORMANT VOCODING

n	F_n	$\bar{F}_n$	N_n (min)	Bits
1	2.000	1.200	19.11	5
2	5.000	2.400	30.23	5
3	1.500	3.000	22.67	5
			13.67 bits	15 bits

periments to test how conservative the results are for the one-parameter and two-parameter variations based upon quantization of all parameters. Preliminary studies indicate that it is quite conservative, with the two-parameter results being much more conservative than the one-parameter results.

APPENDIX A

Evaluation of Mean-Square Spectral Deviation

This Appendix summarizes the relations needed to evaluate the mean-square deviation (16) from $\Delta A(z)$ and $A(z, \lambda)$. Details may be found elsewhere [3]. For notational convenience we define the inner product of two functions of z by the integral

$$\langle F(z), G(z) \rangle = \alpha_M \int_{-\pi}^{\pi} \frac{F(e^{j\theta}) G^*(e^{j\theta})}{|A(e^{j\theta}, \lambda)|^2} \frac{d\theta}{2\pi}, \quad (A1)$$

so that (13) can be written in the form of an inner product

$$(\Delta d)^2 = \frac{2}{\alpha_M} \langle \Delta A(z), \Delta A(z) \rangle. \quad (A2)$$

The polynomials $B_m(z)$ defined by (14) are each of the form

$$B_m(z) = \sum_{i=1}^{m+1} b_{mi} z^{-i} \quad \text{with} \quad b_{m,m+1} = 1, \quad (A3)$$

and are orthogonal to each other, so that

$$\langle B_m(z), B_n(z) \rangle = \delta_{mn} \alpha_m. \quad (A4)$$

As $\Delta A(z)$ contains at most the powers z^{-1} through z^{-M} , it can be expressed as a linear combination of the $B_m(z)$ for $m = 0, 1, \dots, M-1$. This expansion is readily carried out by using the fact that the coefficient of $z^{-(m+1)}$ in $B_m(z)$ is unity. Thus if we define

$$\Delta A(z) = P_{M-1}(z), \quad (A5a)$$

$$P_m(z) = \sum_{i=1}^{m+1} p_{mi} z^{-i}, \quad (A5b)$$

and

$$P_{m-1}(z) = P_m(z) - p_{m,m+1} B_m(z), \quad (A5c)$$

for $m = M-1, M-2, \dots, 0$, we obtain

$$A(z) = P_{M-1}(z) = \sum_{m=0}^{M-1} p_{m,m+1} B_m(z), \quad (A6)$$

which is identical to (18) with $v_m = p_{m,m+1}$. Equation (16) follows from (15), (A2), and the orthogonality relation of (A3).

APPENDIX B

Evaluation of Mean-Square Deviations in Table I

The derivation of the results in Table I is most efficiently carried out by using (24) and the inner product notation of Appendix A. In addition to the use of (17), one must use other properties derived elsewhere [3], but summarized as follows:

$$A_m(z) = z^{-(m+1)} B_m(1/z) = \sum_{i=0}^m a_{mi} z^{-i} \quad \text{with} \quad a_{m0} = 1 \quad (B1)$$

$$\langle A_m(z), z^{-l} \rangle = \langle B_m(z), z^{-l} \rangle = 0 \quad \text{for } l = 1, \dots, m, \quad (B2)$$

$$\langle A_m(z), A_m(z) \rangle = \langle B_m(z), B_m(z) \rangle = \alpha_m, \quad (B3)$$

$$\langle A_m(z), B_m(z) \rangle = -k_{m+1} \alpha_m, \quad (B4)$$

$$\langle z^{-l} F(z), z^{-l} G(z) \rangle = \langle F(z), G(z) \rangle. \quad (B5)$$

Furthermore, if $F(z)$ and $G(z)$ are polynomials with only real coefficients, then the order is immaterial, i.e.,

$$\langle F(z), G(z) \rangle = \langle G(z), F(z) \rangle. \quad (B6)$$

Using (A1) and (24) along with the properties given, the results of Table I follow directly.

APPENDIX C

The General Form of the Mean-Square Deviation

Using the reflection coefficients, the polynomials $A_m(z)$ can be recursively generated through (23). The coefficients of these polynomials a_{mi} of (B1), can be used to recursively generate the autocorrelation sequence $\{r_x(n)\}$ used in the summation of (18) for the evaluation of mean-square deviations. From [3], this recursion has the form

$$r_x(0) = \alpha_0, \quad (C1a)$$

$$r_x(n) = - \sum_{k=1}^n a_{nk} r_x(n-k) \quad \text{for } n = 1, 2, \dots, M. \quad (C1b)$$

If a reflection coefficient is allowed to approach plus or minus unity, it can be seen from the relations generating both the autocorrelation sequences, $\{r_x(n)\}$ and $\{r_\Delta(n)\}$, that these sequences remain well behaved and bounded, so that the bracketed summation of (18) will approach some finite limit. It is only the effect of the α_M in the denominator of (18) that lends a singular behavior to the mean-square deviation as a reflection coefficient approaches unity. From (21) it is seen that the singularity is of the nature of a simple pole as a reflection coefficient approaches either plus or minus one, hence (26) follows directly.

APPENDIX D

Computer Program for Evaluating Spectral Deviation

The evaluation of the spectral deviation due to changes in filter coefficients and reflection coefficients is carried out by the Fortran subroutine SENS of Fig. 6. The approach is quite

```

01300      SUBROUTINE SENS(M,R,SM,RC)
01600      DIMENSION R(1),SM(1),RC(1)
01700      DIMENSION A(21),AQ(21),RAQ(21),RQ(21)
01900      DATA DBFAC/4.342944819/
02100      MP=M+1
02300      A(1)=1.
02600      AQ(1)=1.
02700      A(2)=-R(2)/R(1)
02800      RC(1)=A(2)
02900      ALPHA=R(1)*(1.-A(2)*A(2))
03000      DO 450 J=2,M
03100      MH=J/2
03200      JM=J-1
03300      Q=R(J+1)
03400      DO 420 L=1,JM
03500      LB=J+1-L
03600      420 Q=Q+A(L+1)*R(LB)
03700      Q=-Q/ALPHA
03800      RC(J)=Q
03900      DO 430 K=1,MH
04000      KB=J-K+1
04100      AT=A(K+1)+Q*A(KB)
04200      A(KB)=A(KB)+Q*A(K+1)
04300      430 A(K+1)=AT
04400      A(J+1)=Q
04500      ALPHA=ALPHA*(1.-Q*Q)
04600      C      KILL JOB IF UNSTABLE FILTER
04700      450 IF(ALPHA.LE.0.) CALL EXIT
04800      C      .....
04900      DO 600 I=1,M
05000      DO 460 K=1,M
05100      RQ(K)=RC(K)
05200      460 RQ(I)=RC(I)+1.
05300      AQ(2)=RQ(1)
05400      DO 480 J=2,M
05500      MH=J/2
05600      JM=J-1
05700      DO 470 K=1,MH
05800      KB=J-K+1
05900      AT=AQ(K+1)+RQ(J)*AQ(KB)
06000      AQ(KB)=AQ(KB)+RQ(J)*AQ(K+1)
06100      470 AQ(K+1)=AT
06200      480 AQ(J+1)=RQ(J)
06300      C      EVALUATION OF POLYNOMIAL AUTOCORRELATION
06400      DO 500 L=1,MP
06500      K=MP+L-1
06600      RAQ(L)=0.
06700      DO 500 J=1,K
06800      JL=L+J-1
06900      500 RAQ(L)=RAQ(L)+AQ(J)*AQ(JL)
07000      DEL=R(1)*RAQ(1)
07100      DO 590 J=2,MP
07200      DEL=DEL+2.*R(J)*RAQ(J)
07300      590 SM(1)=DBFAC*SQRT(2.*DEL/ALPHA-2.)
07400      SM(MP)=DBFAC*SQRT(2.*R(1)/ALPHA)
07500      RETURN
07600      END

```

Fig. 6. Fortran subroutine SENS for computing spectral deviations.

similar to that used for the Fortran evaluation of likelihood measures [2]. Many of the lines of the programs are identical and have the same statement numbers. Inputs to the subroutine SENS are the filter order, M , and the dimensioned array $R(\cdot)$, which contains the autocorrelation sequence for the data being analyzed, where

$$R(n) = r_x(n-1) \quad \text{for } n = 1, 2, \dots, M+1.$$

The outputs are the dimensioned array $RC(\cdot)$ containing the reflection coefficients

$$RC(n) = k_n \quad \text{for } n = 1, 2, \dots, M,$$

and the array $SM(\cdot)$ containing the sensitivities or normalized spectral deviations given by

$$SM(n) = 4.34294418 \cdot \Delta d,$$

where Δd is based upon the change in a single-reflection coefficient

$$\Delta k_n = 1 \quad \text{for } n = 1, 2, \dots, M.$$

The final array element based upon a unit change in a single-filter coefficient is computed as

$$SM(M+1) = 4.34294418 \sqrt{2 \cdot [r_x(0)/\alpha_M]},$$

as indicated by (20). All outputs are in units of decibels.

As in the subroutine FIND of [2], the subroutine SENS begins by solving the autocorrelation equations within DO LOOP 450, and terminates for the case of unstable model filters. A step-up procedure [3] is then followed to generate the polynomial $A(z, \lambda')$ as required for the evaluation of $\partial A(z, \lambda)/\partial k_l$ in (25). Since Δk_l is chosen as unity, the incremental polynomial $A(z)$ needed in (13) can be expressed as

$$\lambda A(z) = \frac{\partial A(z, \lambda)}{\partial k_l} A(z, \lambda') - A(z, \lambda). \quad (D1)$$

The orthogonality of $A(z, \lambda)$ and the powers $z^{-1}, z^{-2}, \dots, z^{-M}$ are used to note that

$$\int_{-\pi}^{\pi} \left| \frac{\Delta A(e^{j\theta})}{A(e^{j\theta}, \lambda)} \right|^2 \frac{d\theta}{2\pi} = \int_{-\pi}^{\pi} \left| \frac{A(e^{j\theta}, \lambda')}{A(e^{j\theta}, \lambda)} \right|^2 \frac{d\theta}{2\pi} - 1. \quad (D2)$$

```

C      TEST FOR PROGRAM
      DIMENSION R(5), SM(9), RC(4)
      DATA R/2., 1.6., 776., 15568006., -.0098079/
      M=4
      MP=M+1
      CALL SENS(M, R, SM, RC)
      WRITE(5, 50) (SM(J), J=1, MP)
      WRITE(5, 50) (RC(J), J=1, M)
50     FORMAT(5F10.5)
      CALL EXIT
      END

17.74352  15.62195  9.88957  6.43840  18.78245
-0.80000  0.70000  -0.60000  0.30000

```

Fig. 7. Sample test program for subroutine SENS with output results.

Therefore, the integral of (13) can be evaluated by taking a likelihood ratio minus one. The portions of FIND in [2] which evaluate the likelihood ratio are then used to obtain (D2).

A sample test program is shown in Fig. 7. The input test data are identical to an autocorrelation sequence used in [2]. The output data give the normalized spectral deviations in decibels that would arise from reflection coefficient variations

$k_n = 1$ for $n = 1, 2, 3, 4$, and that which would arise for a unity filter coefficient variation.

REFERENCES

- [1] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing (Special Issue on 1974 Arden House Workshop on Digital Signal Processing)*, vol. ASSP-23, pp. 309-321, June 1975.
- [2] A. H. Gray, Jr. and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 380-391, Oct. 1976.
- [3] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. Berlin, Germany: Springer-Verlag, 1976.
- [4] 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. 637-644, 1971.
- [5] K. N. Stevens, "The perception of sounds shaped by resonant circuits," Sc.D. dissertation, Massachusetts Inst. Technol., Cambridge, 1952.
- [6] F. Itakura, "Optimal nonlinear transformation of LPC's to improve quantization properties," *J. Acoust. Soc. Amer.*, vol. 56 (supplement), Paper H14, p. S16.
- [7] J. D. Markel and A. H. Gray, Jr., "A linear prediction vocoder simulation based upon the autocorrelation method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 124-134, Apr. 1974.

A Hearing Aid for Subjects with Extreme High-Frequency Losses

SIEGFRIED G. KNORR

Abstract—A speech processing system has been developed in which the unvoiced portion of speech is bandwidth compressed from an original bandwidth of 4000 Hz into a low-frequency band not exceeding 1000 Hz, in which hearing impaired subjects with severe high-frequency hearing losses still possess some residual speech perception. The basic compression operation is based upon a time-domain time expansion technique, and the resulting reduction in bandwidth is accomplished without relinquishing the essential information contained in unvoiced speech. Thus, subjects are able again to perceive unvoiced speech of fair intelligibility where conventional hearing aids normally fail to be of any assistance. The imposition of stringent operating requirements such as portability, real-time operation, and functionality in a real-listening

environment composed of many competing speech and noise sources, eliminated numerous elegant speech processing approaches.

I. INTRODUCTION

A FAIR number of subjects with severe hearing losses can be helped only to a limited degree with an ordinary hearing aid if they exhibit very extreme, if not complete, loss of hearing at higher frequencies, and no amount of amplification can significantly improve their speech perception. The typical losses as depicted from the audiogram may be -50 dB at $f = 500$ Hz, -80 dB at $f = 1000$ Hz, and -120 dB at $f = 1500$ Hz. In many instances, sensori-neural losses at the cochlea are the cause for this kind of hearing impairment. If these losses are very severe, subjects may gain little by wearing conventional hearing aids.

Extensive acoustical experimentation lead to the design of a

Manuscript received February 17, 1976; revised August 3, 1976. This work was supported in part by the Joint Services Electronics Program under Contract F44620-71-C-0087 and in part by the Computer Center of the University of California, Berkeley, CA.

The author is with the Department of Electrical Sciences and Engineering, University of California, Los Angeles, CA 90024.