Optimal Quantizers in Linear Predictive Coding of Speech

Marc Belleau and Peter Kabal

80-23

INRS-Telecommunications
c/o Bell-Northern Research Ltd
3, Place du Commerce
Nuns' Island, Que.
H3E 1H6

INRS-Telecommunications Technical Report No. 80-23

September 1980

OPTIMAL QUANTIZERS IN LINEAR PREDICTIVE CODING OF SPEECH

M. Belleau
Electrical Engineering
McGill University
Montreal, Quebec H3A 2A7

P. Kabal
Electrical Engineering
McGill University
Montreal, Quebec H3A 2A7
and
INRS-Telecommunications
University of Quebec
Verdun, Quebec H3E 1H6

ABSTRACT

The linear predictive coding of speech produces a set of parameters (the reflection coefficients) for each analysis frame. Quantization of these parameters is often performed using an inverse-sine quantizer which minimizes the maximum spectral deviation bound. Quantizers based on the expected spectral deviation bound have been proposed in the literature but not evaluated experimentally. This paper evaluates this quantization scheme and in addition examines the effect of decorrelating the coefficients within each frame. It was found that; i) the cross-correlation of the reflection coefficients is small; ii) no bit rate reduction is achieved by decorrelation; iii) using the expected spectral deviation bound criterion, the quality of the reconstructed speech is higher when "inverse-sine" quantizing of the reflection coefficients is performed than when the quantizer minimizing the expected spectral deviation bound for either the reflection coefficients or the decorrelated coefficients is chosen.

I Introduction

This paper examines the reflection coefficients for linear predictive coding of speech. It is desired to find a quantization scheme which reduces the bit rate for the transmission of a speech signal while retaining acceptable quality in the output waveform. Since the perception mechanism of the ear is not well understood, modelling speech quality in mathematical terms is difficult. One empirical measure of speech quality that has been proposed is the L_p norm of the difference between the log spectra of the original and coded speech. The spectrum referred to is the frequency response of the linear predictor [1]. Throughout this paper, we refer to L_2 norms by the term spectral deviation. An expected spectral deviation which depends on many parameters can be upper bounded by the sum of the single parameter expected spectral deviation [3]. The present paper deals with the application of these fidelity criteria to several reflection coefficient quantization schemes. It will be shown that minimum expected spectral deviation bound quantization (MEDQ) of the decorrelated reflection coefficients does not result in a lower bit rate than MEDQ of the original reflection coefficients. Also, in either of the above two schemes, there is a noticeable degradation in the quality of the output speech when compared to speech obtained using inverse sine quantization of the reflection coefficients under the expected spectral deviation bound criterion. In the following section, these ideas will be made more precise.

II The Expected Spectral Deviation Fidelity Criterion

Let the M parameters of the system, x_m , m=1,...,M be denoted by the vector $\underline{x}$. Parameter x_m will be quantized with N_m levels to form a new parameter vector $\underline{x}'$. The spectral deviation due to quantization is $d(\underline{x},\underline{x}')$. By applying the triangle inequality, the following bound on the expected spectral deviation is derived in [3] in the asymptotic limit of a large number of quantization levels N_m,

$$ED(\underline{x}, \underline{x}') \leq E(\bar{D}_{tot})$$

$$\triangleq \sum_{m=1}^{M} \frac{1}{4N_m} \int_{\underline{x}_m}^{\bar{x}_m} \frac{Es_{X_m}(\nu)\; p_{X_m}(\nu)d\nu}{dU/d\nu} \quad . \qquad (1)$$

$U(\nu)$ is the companding law. There is one such function for each parameter. $Es_{x_m}(\nu)$ is a conditional expectation of the sensitivity over all parameters given the sample ν for X_m. $p_{x_m}(\nu)$ is the probability density function (p.d.f.) of X_m and $(\underline{x}_m,\bar{x}_m)$ is the range of quantization of the m^{th} parameter. The derivation of (1) requires that $U(\underline{x}_m) = 0$ and $U(\bar{x}_m) = 1$.

Minimizing the total bit rate $\sum_{i=1}^{M} \log N_i$ while keeping $E(\bar{D}_{tot})$ fixed, yields

$$N_m = \int_{\underline{x}_m}^{\bar{x}_m} \frac{Es_{X_m}(\nu)\; p_{X_m}(\nu)d\nu}{dU/d\nu} \Bigg/ \frac{4E(\bar{D}_{tot})}{M} \quad . \qquad (2)$$

The integral in (2) is minimized by a compander U(.) whose derivative is proportional to $\sqrt{Es_{x_m}(\nu)p_{x_m}(\nu)}$. Using this, in conjunction with the constraints on U(.) at the limits of integration,

$$N_m = \left[\int_{\underline{x}_m}^{\bar{x}_m} \sqrt{Es_{X_m}(\nu)\ p_{X_m}(\nu)}\,d\nu\right]^2 \Bigg/ \frac{4E(\bar{D}_{tot})}{M} \quad . \qquad (3)$$

Denote the input speech autocorrelation coefficients by r(n) and the resulting mean square prediction error by α . Consider the sequence formed by the difference $a_i(x_m+\Delta x_m) - a_i(x_m)$ where the a_i´s are the linear prediction filter coefficients. Denote the autocorrelation of this sequence at lag n by $r_\Delta(n)$. Then, for the spectral deviation, the expression for the sensitivity function $s_{x_m}(\nu)$ can be shown to be [2]

$$s_{X_m}(\nu) = \left(\frac{2}{\alpha}\lim_{\Delta\nu\to 0} \frac{r(0)r_\Delta(0) + 2\sum_{n=1}^{M-1} r(n)\ r_\Delta(n)}{(\Delta\nu)^2}\right)^{1/2} . \qquad (4)$$

Equation (4) involves partial derivatives of the form $\frac{\partial a_i}{\partial x_m}$.

III A Comparison of Three Quantization Schemes

(1) The reflection coefficients (denoted by k_i).

For the k_i´s one can easily prove that (see [2])

$$\frac{\partial a_i}{\partial k_m} = a_i(k_m + 1) - a_i(k_m) \quad .$$

Using such a set of parameters greatly simplifies the computation of (4). Two quantization schemes were tested.

a) Inverse sine quantization (referred to as scheme 1a):

$$U(k_i) = c_i \mathrm{Sin}^{-1} k_i$$

Knowledge of U(x) and the number of levels allows one to compute the quantized value of a variable x. The above choice of U(x) minimizes the maximum spectral deviation bound when the gain of the linear prediction filter is independently quantized. Letting the limits of integration in (2) be the points $\underline{k}_i$ and $\bar{k}_i$, the normalization of U(.) requires that

$$1/c_i = \mathrm{Sin}^{-1}\bar{k}_i - \mathrm{Sin}^{-1}\underline{k}_i \quad .$$

Then $dU/dk_i = c_i/(1-k_i^2)$ is used in (2) to calculate the the number of quantization levels.

b) Minimum $E(\bar{D}_{tot})$ quantization (MEDQ for short) of the k_i's (referred to as scheme 1b):

In this scheme the upper bound on the spectral deviation is minimized using (3). Tests based upon this scheme will complement the theoretical development in [3].

(2) The decorrelated reflection coefficients (denoted by θ_i)

For the θ_i's, only MEDQ was performed (scheme 2). It has been observed that the reflection coefficients are not independent, with the dependency being greatest between k_1 and k_2 [3][*]. Decorrelation can be used to achieve a certain measure of independence between them [4]. The decorrelation operation can be viewed as a transformation,

$$\theta_i = g_i(k_1,\ldots,k_M) \quad .$$

Note that

$$\frac{\partial a_j}{\partial \theta_m} = \sum_{i=1}^{M} \frac{\partial a_j}{\partial k_i} \frac{\partial k_i}{\partial \theta_m} \quad .$$

This partial derivative can be computed since $g_i^{-1}(.)$ can be readily determined. The transformation which decorrelates the reflection coefficients is orthogonal. As a result, the following relation can be shown to hold

$$\sum_{i=1}^{M} \sqrt{\mathrm{Var}\theta_i} \leq \sum_{i=1}^{M} \sqrt{\mathrm{Var}k_i} \quad .$$

Since the trace of a matrix is constant under a similarity transformation (here it represents the sum of the variances) this implies that the variation becomes more concentrated in a few of the θ_i´s.

Sambur has applied decorrelation to the log area ratios (where an area ratio = $(1-k_i)/(1+k_i)$) as well as to the k_i´s [4,5]. Using 12 parameters, he observed that 90% of the total variance is contained in 5 or 6 of them. If $\log N_i$ were proportional to

[*] This study considers only intra-frame correlations.

$\sqrt{\text{Var } X_i}$, then the above inequality would imply that decorrelation reduces the bit rate. Decorrelation would then be attractive as long as the gathering of the necessary statistics is done infrequently. Fortunately, the p.d.f. of the k_i's is not very speaker and context dependent. In fact, McCandless [7] reports that the statistics are much more dependent on the amount of background noise in the input speech. Keeping the noise to a minimum, and assuming that the correlation among different k_i's is also reasonably independent of speaker and content, the required eigenvector analysis can be predetermined.

IV Experimental Results

Analysis Conditions

In all, 14 utterances of approximately 2 to 3 seconds in duration by 3 male and 2 female speakers, were used for these tests. The input speech was bandlimited to 5 kHz and sampled at 10 kHz. Adaptive pre-emphasis and windowing (using a Hamming window) were then applied. This was followed by an autocorrelation analysis (50 frames per second) with a filter order of 14 as suggested in [6].

Parameter Statistics

Statistics necessary in the evaluation of the covariance matrix were gathered about the k_i's. Then in order to study the dependence of the k_i's on the text and speakers, statistics were gathered for single utterances and also for all 14 utterances (referred to as composite statistics). The mean of all k_i's was computed using a time average over N frames, and the cross-correlation was obtained by a time average over N-1 frames. The values of Ek_i and $Vark_i$ are shown in Table I for single utterance and composite statistics. Comparing the 12 variances from Table I with those given in [5], it is found that the sum of the 12 variances is roughly the same and is also distributed in a similar way. The probability distribution of the k_i's does not depend on the filter order M for $i \leq M$, i.e. taking two arbitrary filter orders M_1 and M_2 , the distributions are the same for $1 \leq i \leq \min(M_1, M_2)$ [6]. Figures 1a and 1b give the histograms of the first and second reflection coefficient, respectively. Note that the histogram is zero over part of the allowed ranges of the parameters. The general shape of these histograms is in agreement with those of [2,3].

Assuming that $\log N_i$ is directly proportional to the standard deviation of X_i, Table II gives the potential percentage reduction in bit rate that can be achieved when the k_i's are decorrelated. For composite statistics, the potential bit reduction is less than for the single utterance statistics. From these percentages, it can be seen that even for single utterance statistics, the k_i's are not very correlated. Table III lists characteristics of the θ_i's. The θ_i's are

TABLE I

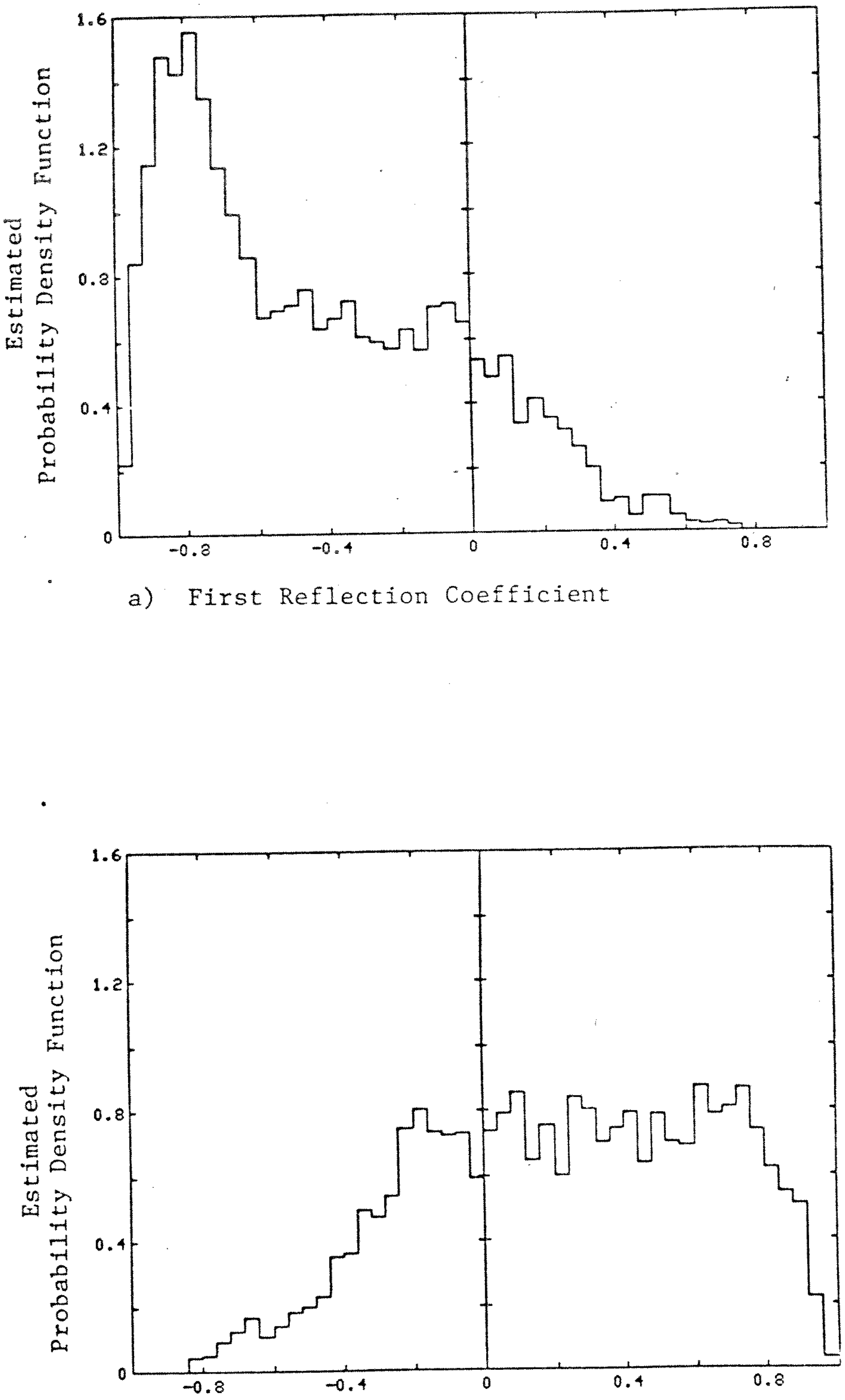
THE MEAN AND VARIANCE OF THE REFLECTION COEFFICIENTS

SINGLE UTTERANCE STATISTICS (UTTERANCE #1)

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Ek_i	-.359	.067	-.212	.157	-.169	.173	.025	.163	.135	.202	.009	-.044	-.107	-.019
$Vark_i$	.133	.138	.055	.072	.082	.039	.037	.038	.035	.056	.028	.022	.015	.012

COMPOSITE STATISTICS

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
EK_i	.-430	.228	-.198	.089	-.071	.125	.038	.157	.070	.108	.008	.014	-.004	.033
$Vark_i$	.152	.172	.091	.070	.077	.060	.037	.044	.056	.048	.028	.021	.017	.013



a) First Reflection Coefficient

b) Second Reflection Coefficient

Figure 1 Relative frequency of occurence of the reflection coefficients.

TABLE II

SUM OF STANDARD DEVIATIONS

	Single Utterance Statistics		Composite Statistics
	Utterance #1	Utterance #2	
$\Sigma \text{Var}\theta_i$	.783	.919	.888
A $\triangleq$ $\Sigma\sqrt{\text{Vark}_i}$	3.085	3.336	3.308
$\Sigma\sqrt{\text{Var}\theta_i}$	2.881	2.976	3.181
B $\triangleq$ $\Sigma(\sqrt{\text{Vark}_i}-\sqrt{\text{Var}\theta_i})$	.204	.360	.127
Potential Bit rate reduction B/A (%)	7	11	4

TABLE III

SINGLE UTTERANCE STATISTICS

(Utterance #1)

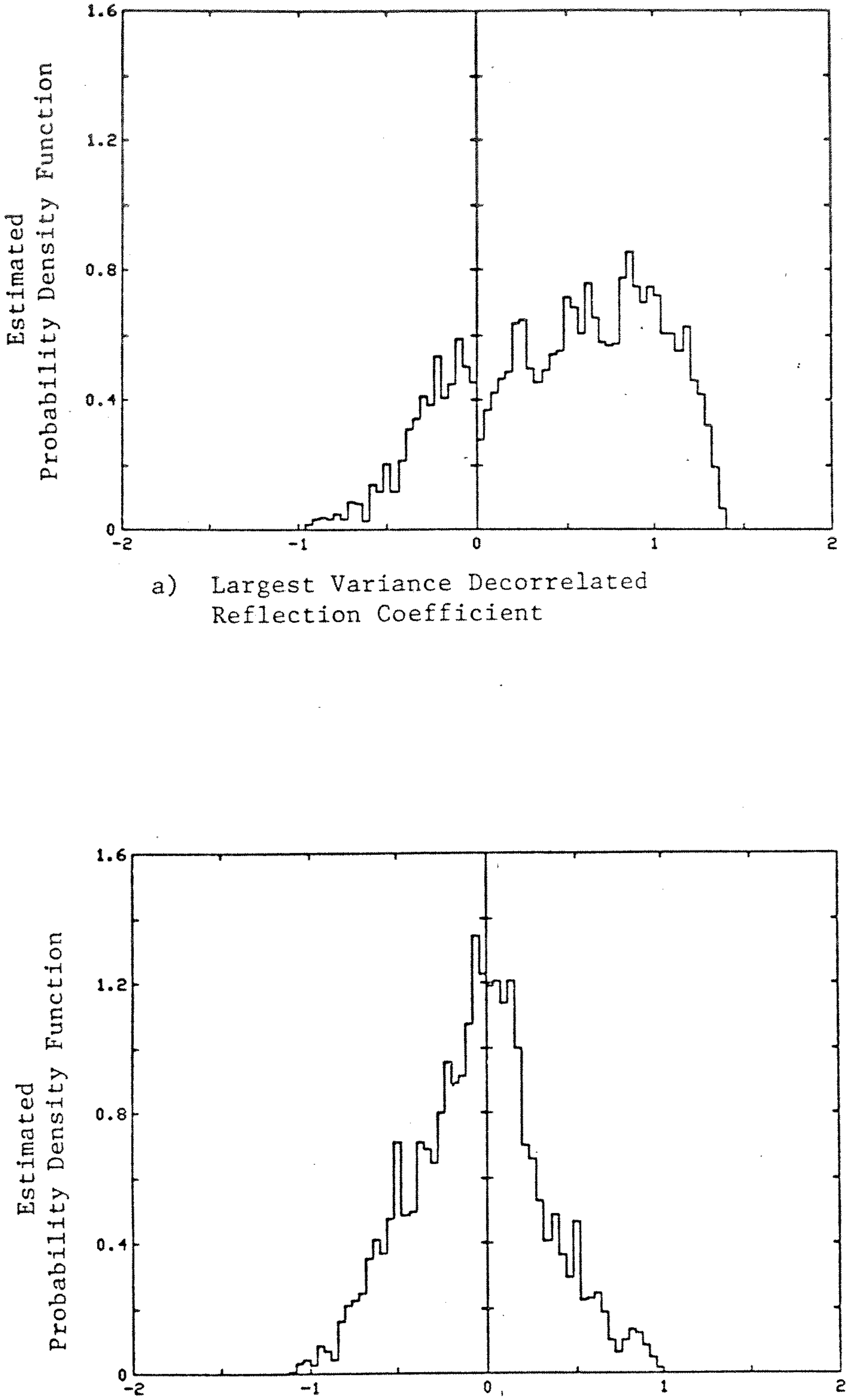
Mean $E\theta_i$	Variance λ_i	Allowable Range $(-R_i,+R_i)$
.451	.224	± 2.695
-.033	.160	± 2.619
.190	.091	± 2.954
-.124	.055	± 3.072
.243	.049	± 2.722
.008	.036	± 2.947
.070	.033	± 2.930
.103	.027	± 3.171
.055	.023	± 2.696
-.081	.019	± 2.892
-.027	.016	± 2.871
-0.30	.013	± 2.364
-.090	.010	± 2.994
.025	.007	± 2.598

listed in order of their decreasing variance λ_i. Notice that their range is always much larger than λ_i. For the smallest λ_i, it is in fact larger than λ_i by a factor of 30. Figures 2a and 2b are the relative frequency of occurrence histograms of the two largest variance decorrelated coefficients using composite statistics. It is expected that the probability distributions of the θ_i's also do not depend very much on the value of M if the latter is large because the variance and the cross-correlation of the k_i's decrease as i increases. Again, as for the k_i's, the histogram is zero over part of the allowed range of the parameters.

Parameter Quantization

Figures 3a and 3b are the min $E(\overline{D}_{tot})$ quantizer companding curves for the first and second reflection coefficients. The general shape of the plots is in agreement with those of [2,3]. As i increases, the quantizer curves of the k_i's become more symmetrical about Ek_i. In order to determine the levels and boundaries, it is only necessary to know the shape of the quantizer curve although its correct normalization is required in computing the number of levels. It can be seen from Fig. 2a that the quantizer curve of Fig. 3b is flat outside the range over which the histogram is nonzero. (Recall that $dU/d\upsilon$ is proportional to $\sqrt{Es_{X_m}(\upsilon)p_{X_m}(\upsilon)}$) .

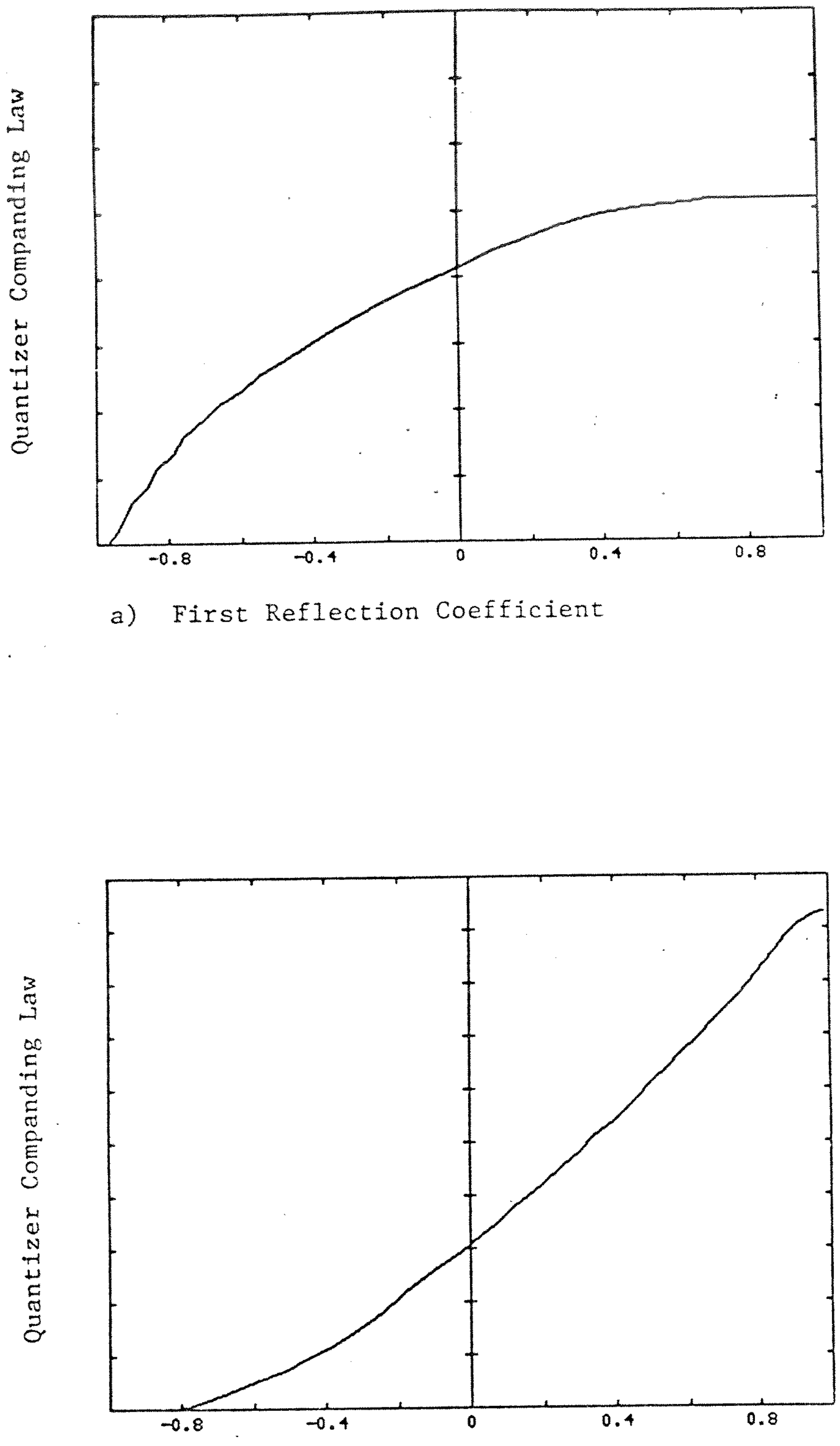
Fig. 4 is the min $E(\overline{D}_{tot})$ quantizer companding curve of the largest variance θ_i. The quantizer curves for the smaller variance



a) Largest Variance Decorrelated Reflection Coefficient

b) Second Largest Variance Decorrelated Reflection Coefficient

Figure 2 Relative frequency of occurence of the decorrelated reflection coefficients.



a) First Reflection Coefficient

b) Second Reflection Coefficient

Figure 3 Quantizer companding curves for the reflection coefficients.

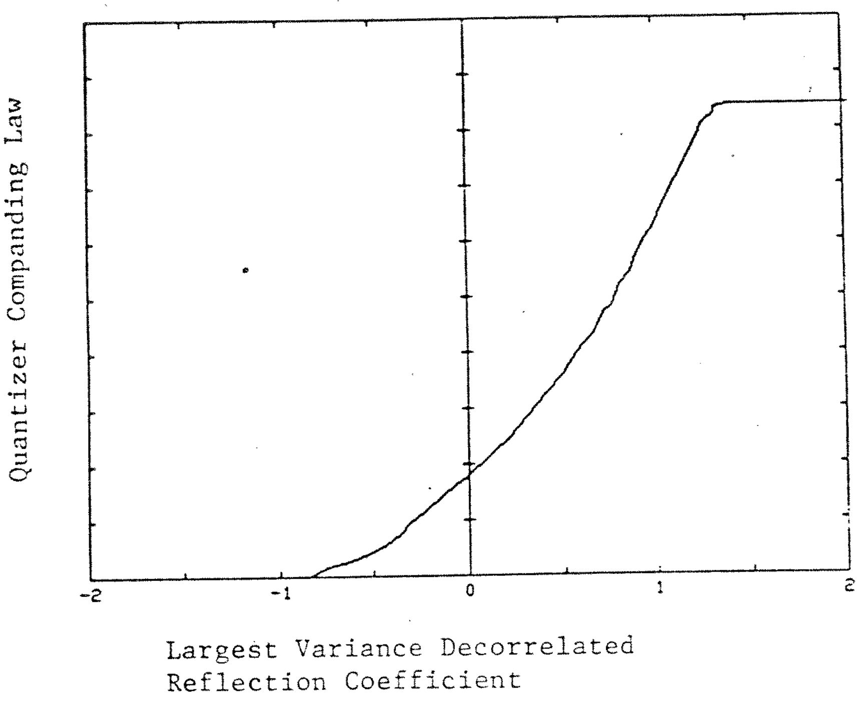


Figure 4 Quantizer companding curve for the decorrelated reflection coefficient

parameters are even more symmetrical about $E\theta_i$ and in addition their shape becomes more similar to that of the quantizer curves for the k_i´s.

The number of quantization levels necessary to achieve a 3.5 dB* upper bound, $E(\overline{D}_{tot})$, on the spectral deviation was calculated from (2) or (3) as appropriate. Table IVa is for scheme 1a (inverse sine quantization of the k_i´s) and scheme 1b (MEDQ of the k_i´s). Table IVb shows the same information for scheme 2 (MEDQ on the decorrelated coefficients) and, in addition, the mean, the variance and the allowed range of θ_i´s. These results shown are for composite statistics.

Converting levels to bits, with a frame rate f_r, the total bit rate is

$$\left(\sum_{i=1}^{M} \log_2 N_i\right) f_r \qquad .$$

When the error signal is left unquantized, the total number of bits required if $E(\overline{D}_{tot})$ is not to be exceeded, is 2570 bits/sec for scheme 1a, 2250 bits/sec for scheme 1b, and 2380 bits/sec for scheme 2. Scheme 1b is therefore slightly superior to scheme 1a as predicted in the

* Spectral deviations in the range 3 to 4 dB have been found to give acceptable results [1,2,3].

TABLE IVa

NUMBER OF LEVELS AND THE NON-ZERO INTERVAL

FOR THE REFLECTION COEFFICIENTS

	Scheme 1a	Scheme 1b	Lower Limit	Upper Limit
i	N_i	N_i	$\underline{k}_i$	$\overline{k}_i$
1	33	27	-.98	.72
2	26	22	-.85	.97
3	21	16	-.86	.76
4	18	13	-.70	.85
5	17	13	-.79	.73
6	13	10	-.53	.77
7	13	8	-.57	.73
8	11	8	-.50	.75
9	12	9	-.60	.80
10	11	8	-.52	.84
11	9	6	-.53	.65
12	7	5	-.48	.52
13	7	4	-.57	.38
14	5	4	-.36	.41

TABLE IVb

COMPOSITE STATISTICS AND NUMBER OF LEVELS FOR THE DECORRELATED REFLECTION COEFFICIENTS

Mean	Variance	Lower Limit	Upper Limit	Allowable Range	Scheme 2
$E\theta_i$	λ_i	$\underline{\theta}_i$	$\bar{\theta}_i$	$(-R_i,+R_i)$	N_i
.495	.251	-0.856	1.376	± 2.085	35
-.027	.138	-1.055	1.001	± 2.075	25
-.231	.111	-1.122	.734	± 2.158	18
-.130	.080	-1.048	.728	± 2.912	14
.104	.062	-.756	.952	± 2.800	14
-.022	.056	-.871	.813	± 2.904	11
.032	.037	-.653	.823	± 2.839	8
.069	.033	-.512	.796	± 2.844	9
.039	.031	-.580	.603	± 2.233	8
.054	.026	-.427	.655	± 2.848	8
-.078	.019	-.524	.360	± 3.276	8
-.302	.018	-.507	.531	± 2.414	6
.061	.015	-.425	.454	± 2.836	6
-.012	.011	-.376	.430	± 2.688	5

theoretical study of [3]. Unfortunately even though

$$\sum_{i=1}^{M} \sqrt{\mathrm{Var} k_i} \geq \sum_{i=1}^{M} \sqrt{\lambda_i} ,$$

the bit rate for scheme 1b is still less than that of scheme 2 given a fixed bound $E(\overline{D}_{tot})$. The final conclusion must however, be based on subjective assessment of the speech quality.

Subjective Results

This experiment allows a subjective comparison of processed speech in which only the reflection coefficients are changed due to different quantization strategies. For one particular utterance, scheme 1a, 1b and 2 all resulted in quality almost indistinguishable from that of the original utterance. However, for most of the utterances that were processed, it was found that scheme 1a produces speech of quality close to that of the original, scheme 1b resulted in the most discernable degradation while scheme 2 was only slightly superior to scheme 1b. Note that the synthesizer used the error signal as the driving function, and the quantizers were designed based on the composite statistics. Then, to check for dependence on the method of determining the statistics, the composite statistics were replaced by single utterance statistics. Among all utterances, the worst performer under composite statistics was selected for this study. The performance did not improve, indicating that the that speech quality does not depend on which statistics are used to compute the θ_i's. This should not be surprising in view of the fact that the k_i's are never very correlated.

V Conclusions

Using the $E(\overline{D}_{tot})$ fidelity criterion, it has been verified that scheme 1b results in a slightly lower bit rate than scheme 1a, as is expected from the results of [3]. Decorrelation of the k_i's results in a total bit rate which is lower than that using scheme 1a but unfortunately, is higher than that using scheme 1b. Recall that when the number of bits is proportional to the standard deviation, the percentage reduction in bit rate is not substantial for either single utterance or composite statistics since the cross-correlations in the reflection coefficients are not pronounced. It is found that scheme 2 results in speech quality slightly superior to that using scheme 1b, while scheme 1a surpasses scheme 2. In fact, the latter method results in speech quality fairly close to that of the original utterance. However, recall that scheme 1a was not designed to minimize the $E(\overline{D}_{tot})$ criterion but rather, to minimize the $\max(\overline{D}_{tot})$ criterion. The fact that, under the $E(\overline{D}_{tot})$ criterion, scheme 1a is subjectively better than scheme 1b indicates that $\max(\overline{D}_{tot})$ may be a better objective error criterion.

The frame to frame dependence of the k_i's is even more significant than the cross-correlation within a frame [3]. Variable frame rate techniques [6,7] and frame to frame DPCM coding [4] can be used to take advantage of the frame to frame correlations. Hence, if decorrelation is to be performed, it should be followed by variable rate and/or DPCM on the decorrelated coefficients [4].

This study examined only a few of the possible quantization strategies for the reflection coefficients. However, it is felt that it is unlikely that other transformations of the k_i's will result in any significant bit reduction for <u>independent</u> quantization of the resulting coefficients.

References

1. Markel, J.D., Gray, A.H., Jr., "Distance measures for speech processing", IEEE Transactions on ASSP, Vol.24, No.5, October 1976 pp. 380-390.
2. Gray, A.H., Jr., Markel, J.D., "Quantization and bit allocation in speech processing", IEEE Transactions on ASSP, Vol.24, No.6, December 1976, pp. 459-473.
3. Gray, A.H., Jr., Gray, R.M., Markel, J.D., "Comparison of optimal quantizations of speech reflection coefficients", IEEE Transactions on ASSP, Vol.25, No.1, February 1977, pp. 9-22.
4. Sambur, M.R., "An efficient linear prediction vocoder", Bell System Technical Journal, Vol.54, December 1975, pp. 1693-1723.
5. Sambur, M.R., "Speaker recognition using orthogonal linear prediction", IEEE Transactions on ASSP, Vol.24, No.4, August 1976, pp. 283-289.
6. Markel, J.D., Gray, A.H., Jr., Linear Prediction of Speech, Springer-Verlag, 1976.
7. McCandless, S.S., "A new encoding technique for the k-parameters: A statistical approach", NSC Note #53, MIT Lincoln Laboratory, December 1974.