Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ADAPTIVE, DIFFERENTIAL PULSE-CODE MODULATION FOR SPEECH PROCESSING

BY

PETER CUMMISKEY

A DISSERTATION

PRESENTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF SCIENCE IN ELECTRICAL ENGINEERING

AT

NEWARK COLLEGE OF ENGINEERING

Newark, New Jersey
1973

APPROVAL OF DISSERTATION

ADAPTIVE, DIFFERENTIAL PULSE-CODE MODULATION FOR SPEECH PROCESSING

BY

PETER CUMMISKEY

FOR

DEPARTMENT OF ELECTRICAL ENGINEERING

NEWARK COLLEGE OF ENGINEERING

BY

FACULTY COMMITTEE

APPROVED: _____ CHAIRMAN

_____

_____

_____

_____

NEWARK, NEW JERSEY

JUNE, 1973

ABSTRACT

The objective of the research reported here is the design of efficient speech coders that can easily be implemented in integrated circuit hardware. Companding techniques like those introduced by M. R. Winkler, J. A. Greefkes, F. DeJager, A. Tomozawa and H. Kaneko were explored along with a large body of theory concerning the application of linear prediction to speech coding.

The best features of the speech signal to be measured and coded are the overall amplitude, the resonant frequencies and dampings of the vocal cavity and the fundamental frequency of the vocal cord oscillations. Adaptive quantization was used to track variations in overall amplitude, and adaptive prediction was used to track the frequencies and dampings of the cavity resonances. No attempt was made to exploit redundancies related to the vocal cord oscillations, however.

An adaptive differential pulse code modulator (i.e., an ADPCM coder) with a fixed integrator was simulated first. Later a hardware model was constructed, signal to noise measurements were taken and subjective tests

Abstract (continued)

conducted. When operating at 4 bits per sample, speech of a quality nearly equal to that of 7 bit log PCM was regenerated by the ADPCM encoder. At 3 bits per sample speech quality was nearly equal to 6 bit log PCM.

Further improvements were achieved with the application of adaptive predictors in place of the integrator. The predictor coefficients form a vector which is adapted in a direction away from the gradient with respect to the error power. By applying this technique to the quantized signals occurring in the coder, the coefficients are derived from the quantized error signal; hence, there is no need to transmit them.

ACKNOWLEDGMENTS

The author wishes to acknowledge many useful discussions and suggestions offered by Dr. N. S. Jayant and Dr. J. L. Flanagan of Bell Laboratories. He also wishes to thank Dr. L. R. Rabiner and Dr. R. W. Schafer for the use of two of their programs. The advice of Miss B. McDermott was partially responsible for the success of the subjective tests, and her help in running the multi-dimensional scaling programs is also appreciated. He would also like to thank Dr. W. Ball of Newark College of Engineering for his counsel and his editorial comments. Finally, the extensive editorial comments contributed by Dr. R. W. Schafer are also acknowledged.

CONTENTS

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF TABLES

INTRODUCTION

Many long distance telephone calls are transmitted in a digital form called pulse code modulation, (PCM). This process involves log compression [22] of the speech signal, quantizing to 128 levels, conversion to a 7 bit binary code and transmission (see Fig. 1a). Compression provides a signal with a more nearly uniform amplitude probability distribution than that of natural speech. At the receiving end of the channel the binary code words are converted into signal levels and the reconstructed signal is processed by an exponential expander and a desampling filter. Hence, low amplitude regions of high probability are finely quantized and high amplitude, low probability regions are coarsely quantized. An 8 kHz rate is generally used when sampling telephone speech; therefore, the information rate required is 56 kilobits/sec.

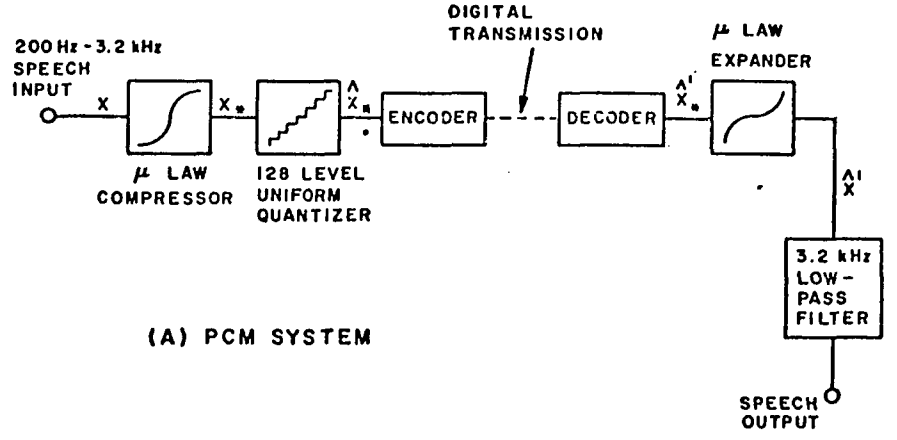
The only features of the PCM system that are tailored to the speech signal as opposed to other signals are the sampling rate and the compressor and expander characteristics. The sampling rate is dictated by the bandwidth of the speech signal and the requirements on the desampling filter. (The desampling filter must pass all signal components in the 200 Hz to 3.2 kHz telephone band and

must reject aliasing components above 4.8 kHz.) By providing the analog to digital converter in the PCM system with a nearly uniform probability distribution across the 128 quantizing levels, approximately 2 bits per word or 16 kilobits/sec. of information rate is saved.

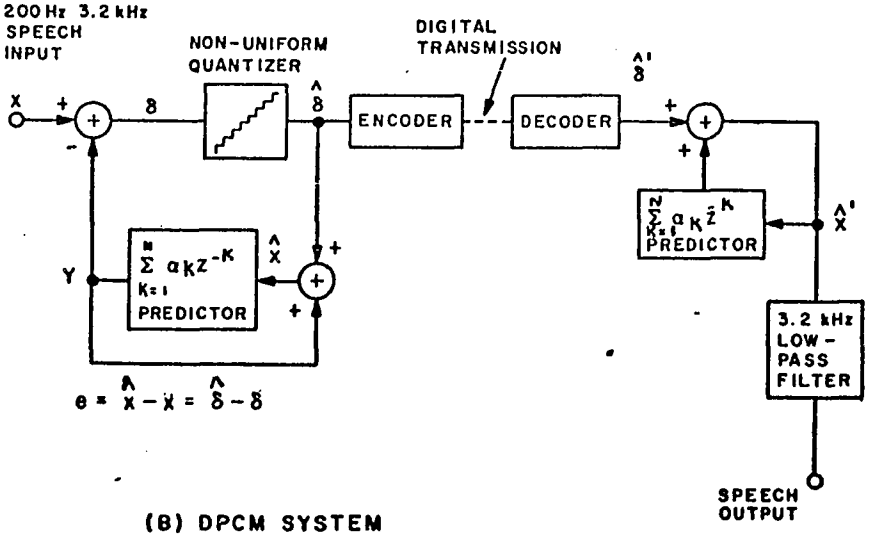
Another parameter which should be considered when coding a signal, X(t), is its autocovariance at various sample delays (T, 2T, 3T, etc.). At 8 kHz sampling the first autocovariance of telephone speech, ρ_1, is approximately equal to 0.8, where

$$\rho_1 = \lim_{j\to\infty} \frac{\sum_{i=-j}^{0} X(iT) \cdot X((i-1)T)}{\sum_{i=-j}^{0} (X(iT))^2} .$$

For this reason differential coding (DPCM) is more efficient than PCM. In the DPCM encoder shown in Fig. 1b, the difference between the input signal and the output of an integrator is coded and transmitted. Then the quantized difference is added to the integrator to obtain an approximation to the current input sample.



(A) PCM SYSTEM



(B) DPCM SYSTEM

FIG. 1

A signal to noise improvement of 3 to 7 db over PCM or a saving of one bit per word is thus achieved. Hence, with conventional DPCM the bit rate may be reduced to 48 kilobits/sec. without a sacrifice in quality. The benefits obtained from nonuniform quantizing in DPCM are the same as those obtained from compression and expansion in the PCM system, because the probability distribution of the first difference of speech is about the same as the speech amplitude distribution.

By applying the Wiener-Kolmogorov [14] method of linear prediction to a sampled signal, two or more terms in the autocovariance function may be used to design a more efficient speech coder. In this case the quantity to be quantized and coded is the predictor error. Unfortunately, a single, fixed, higher-order predictor derived from long-term, speech statistics will not work reliably for all talkers or all circuit conditions. A 4 db improvement in signal to noise ratio over conventional DPCM was reported by R. A. McDonald [16] for a fixed third order predictor operating on a particular, low-passed speech sample. These results could not be duplicated with the band-limited speech sample used in the author's simulations, where the SNR was increased by only 0.4 db.

Speech is not a long-term, stationary process, but rather a short-term quasi-stationary phenomena. Hence, higher order variable predictors like those suggested by B. S. Atal and M. R. Schroeder [2] have been used to obtain better estimates of the input signal from previous samples. In these schemes, predictor parameters are computed periodically from past samples of the input signal. In one of these coders the predictor error was coded with only one bit per sample and transmitted along with the predictor coefficients and a volume parameter. Speech with a quality better than 5 bit log PCM but inferior to 6 bit log PCM was regenerated at a data rate of only 10 kilobits/sec. (6 kilobits/sec. for error data and 4 kilobits/sec. for volume data and predictor coefficients.) It should be noted that the data rate for the predictor parameters is low because they need to be readjusted only once every 5 ms or 10 ms, the interval over which they are calculated to minimize the mean square error. Among the predictor parameters used by Schroeder and Atal there is a delay related to the pitch interval. Due to the large number of computations required to obtain this parameter the scheme is presently too complex for most communications systems applications.

Much attention has been given to delta modulation (i.e., DPCM with a two level quantizer). In these coders

the input signal is grossly over sampled, the sampling rate being equal to the bit rate. F. DeJager [5] showed that a delta modulator with fixed integration networks and a fixed step size can regenerate speech with telephone quality when operating at approximately double the bit rate of 7 bit log PCM. Following M. R. Winkler's [25] work with instantaneous adaptation, several delta modulators were developed where the quantization step is varied as a function of the history of the bit stream. Excellent quality speech is regenerated by most of these adaptive delta modulators while operating at 40 to 60 kilobits/sec. In these cases performance is improved by tracking the nonstationary amplitude characteristics of the input signal.

Of the coding techniques mentioned thus far, non-adaptive, linear delta modulation is the least efficient with respect to the transmitted bit rate. Hence it is reasonable to expect that if adaptation of the quantization step greatly improves the performance of delta modulators, adaptation of the quantizer in PCM and DPCM system will also improve their performance. Therefore, it was decided that adaptive DPCM (ADPCM) should be investigated. A DPCM coder with an adaptive quantizer was simulated first. Later a hardware model was constructed, signal to noise measurements were taken and a subjective test conducted.

When operating at 4 bits per sample, speech of a subjective quality nearly equal to that of 7 bit log PCM was regenerated by the ADPCM coder. At 3 bits per sample, speech quality was nearly equal to 6 bit log PCM.

Further improvements were achieved with the application of adaptive predictors in place of the integrator. The predictor coefficients form a vector which is adapted in a direction opposite to the gradient with respect the error power. By applying this technique to the quantized signals occurring in the coder, the predictor coefficients are derived indirectly from the quantized error signal, and the need to transmit predictor coefficients is eliminated.

Although more accurate calculation of the coefficients than that described above yields greater reduction in the prediction error, there is very little real gain because the coefficients must be transmitted. In either case steepest - descent, gradient techniques remain an excellent method for determining the predictor coefficients.

There are many digital channels where virtually error free performance exists. The T1 and T2 systems used for transmitting PCM coded messages have error rates less than one in 10^6 bits. In many other applications such as voice-answer-back systems where speech is stored on computer

memories, the channel is also error free. For these reasons a low priority was given to the study of coder performance in the presence of transmission errors.

In this paper, the parameters considered in the design of an efficient speech coder are the bit rate, the signal to quantizing noise ratio, the subjective quality of the regenerated speech and the complexity of the hardware involved.

PART I

A STATEMENT OF THE THEORY

A theoretical foundation is layed in this part of the dissertation so that the reader can better understand the author's research and the motivation for doing it.

The design of differential coders (DPCM coders and delta modulators) can be divided into two parts (prediction of the input sample based on previous samples and optimum quantization of the prediction error). In Sections 1 and 2 the theory of optimum, linear prediction is presented. A fixed first order predictor is described in Section 1 and higher order prediction is discussed in Section 2. In Section 3 the design of optimum quantizers and μ law quantizers is described.

Speech is not a long-term stationary process but rather a short-term quasi-stationary phenomena. (See Fig. 2). During voicing, when the vocal cavity is excited by pulses of air emitted at the vocal chords, the waveform is almost periodic. A sequence of damped oscillations appears in the time waveform. The frequencies and dampings of these oscillations are a function of the shape of the vocal cavity and of other slowly changing characteristics of the vocal organs. During fricative sounds like "sh"

and "ch" the vocal cavity is excited by turbulent air at a constriction. In these cases high frequency noise appears in the time waveform and the amplitude of the signal is greatly diminished.

If the pitch, the modes of oscillation and/or the overall amplitude are specified for each of the stationary intervals a more efficient coder can be realized than one which utilizes only long-term average statistics. In delta modulators the quantization step has been adapted to match the changing amplitude characteristics of the speech signal. These adaptive delta modulators are reviewed in Section 4 because similar techniques can easily be applied to PCM and DPCM coders. A higher order predictor which adapts to track the changing modes in the speech waveform can also be incorporated in a DPCM coder; however, this modification greatly increases the complexity of the coder. Finally, the pitch interval redundancies can be exploited at an even greater cost.

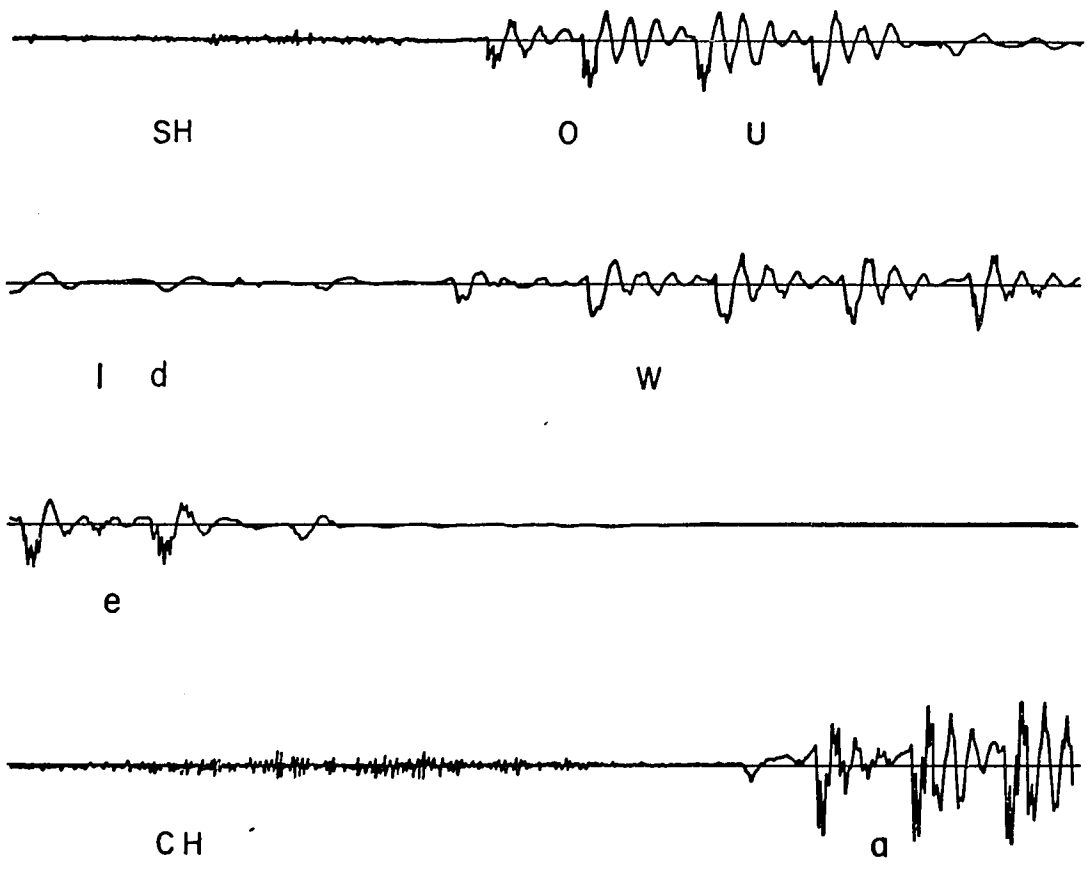


FIG. 2

Section 1 Analysis of a DPCM Coder With a Fixed First Order Predictor

In the conventional DPCM system shown in Fig. 3 the difference between the ith sample of the input signal, X_i, and a predicted value Y_i: is quantized and coded for transmission. The quantized difference signal, $\hat{\delta}_i$, is added to Y_i to obtain a corrected approximation to the input.

$$\hat{X}_i = Y_i + \hat{\delta}_i \qquad (1.1.1)$$

The quantizing error is given by

$$e_i = \hat{\delta}_i - \delta_i \qquad (1.1.2)$$

It may also be inferred from Fig. 3 that

$$\delta_i = X_i - Y_i \quad . \qquad (1.1.3)$$

For M = 1,

$$Y_i = a_1 \hat{X}_{i-1} \qquad (1.1.4)$$

Substitution of (1.1.1) and (1.1.3) into (1.1.2) shows that the system error and the quantization error are the same.

$$e_i = \hat{X}_i - X_i \qquad (1.1.5)$$

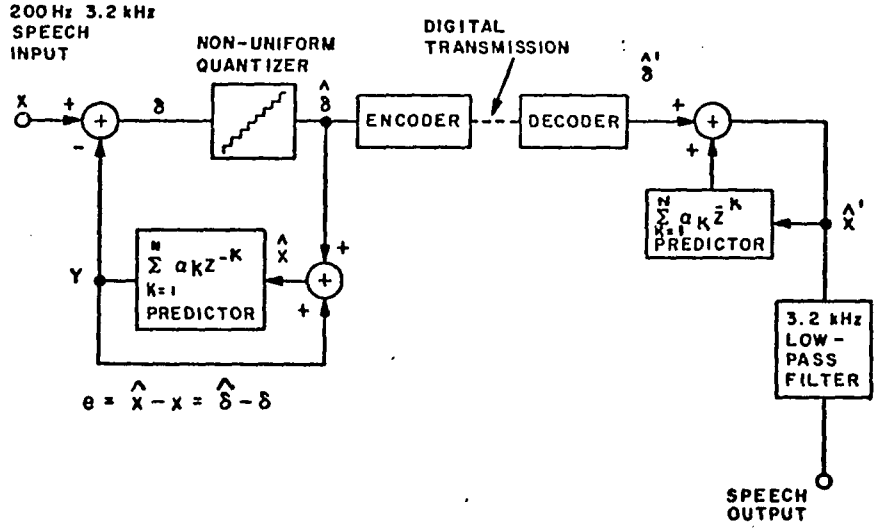

200 Hz 3.2 kHz
SPEECH
INPUT
NON-UNIFORM
QUANTIZER
DIGITAL
TRANSMISSION
ENCODER
DECODER
PREDICTOR
PREDICTOR
3.2 kHz
LOW-
PASS
FILTER
SPEECH
OUTPUT
e = x̂ − x = δ̂ − δ

FIG. 3

If quantizing is relatively fine and overload does not occur, the quantizing noise will have the same properties as in the PCM system. (The noise spectrum will be white and there will be a very small correlation between the noise and the input signal.)

If M = 1, then from Fig. 3 it is inferred that

$$\delta_i = X_i - a_1 X_{i-1} - a_1 e_{i-1} \tag{1.1.6}$$

The expected or average power in the signal to the quantizer is therefore

$$E(\delta_i^2) = E\left((X_i - a_1 X_{i-1} - a_1 e_{i-1})^2\right) \tag{1.1.7}$$

or

$$E(\delta_i^2) = E(X_i^2) - 2a_1 E(X_i X_{i-1}) + 2a_1^2 E(X_{i-1} e_{i-1}) - 2a_1 E(X_i e_{i-1}) + a_1^2 E(X_{i-1}^2) + a_1^2 E(e_{i-1}^2). \tag{1.1.8}$$

The only correlation between the input signal and the error occurs during periods of heavy load or overload (i.e., when $\hat{\delta}$ is often equal to the maximum quantizer level). Under these conditions the error signal correlates well with the derivative of the input signal, but not very well with the input itself. Hence, several terms may be dropped from Eq. 1.1.8.

$$E(\delta_i^2) \simeq E(X_i^2) - 2a_1 E(X_i X_{i-1}) + a_1^2 E(X_{i-1}^2) + a_1^2 E(e_{i-1}^2) \qquad (1.1.9)$$

The circuit between the quantizer output and the output of the predictor is the sampled data equivalent of an integrator in series with a delay. The cut-off frequency of the integrator is given by

$$f_c = \frac{1}{2\pi T} \ln\left(\frac{1}{a_1}\right), \qquad (1.1.10)$$

where T is the sampling interval.

There is an optimum a_1, and hence, an optimum f_c which is derived by setting the following derivative equal to zero:

$$\frac{d}{da_1} E(\delta_i^2) = 0 \qquad (1.1.11)$$

It may be assumed that the quantization is optimum; hence, the smaller the signal to the quantizer the smaller the error power will be. In Section 3 the design of an optimum "N" level quantizer is discussed. If Eq. (1.1.9) is differentiated with respect to a_1 and the result set equal to zero, the following value is obtained for the optimum predictor coefficient:

$$a_{1_{Opt}} = \frac{E(X_i X_{i-1})}{E\left(X_{i-1}^2\right) + E(e_{i-1}^2)} \tag{1.1.12}$$

$$a_{1_{Opt}} = \frac{E(X_i X_{i-1})/E\left(X_{i-1}^2\right)}{1 + \left(E(e_{i-1}^2)/E(X_{i-1}^2)\right)}$$

$$= \frac{\rho_1}{1 + \frac{1}{SNR}} \approx \rho_1 \ , \tag{1.1.13}$$

where ρ_1 is the normalized autocovariance for a one sample delay and SNR is the overall signal to noise power ratio.

$$\rho_1 = E(X_i X_{i-1})/E\left(X_{i-1}^2\right) \tag{1.1.14}$$

In the DPCM encoder there is an improvement in the signal to noise ratio over that obtained in a PCM system. The improvement is equal to the ratio of the power in the input signal, X, to the power in the difference signal, δ.

$$SNR = \frac{E\left(X_i^2\right)}{E\left(e_i^2\right)} = \frac{E\left(X_i^2\right)}{E\left(\delta_i^2\right)} \cdot \frac{E\left(\delta_i^2\right)}{E\left(e_i^2\right)} \tag{1.1.15}$$

If it is noted that $E\left(X_{i-1}^2\right) = E\left(X_i^2\right)$, the signal to noise improvement can easily be derived from Eq. (1.1.9).

$$\frac{E\left(X_i^2\right)}{E\left(\delta_i^2\right)} = SNR_{Improvement} = \frac{1}{1 - 2a_1\rho_1 + a_1^2\left(1 + \frac{1}{SNR}\right)} \tag{1.1.16}$$

For

$$a_1 = a_{1_{Opt}} = \rho_1 \Big/ \left(1 + \frac{1}{SNR}\right),$$

$$SNR_{Improvement} = \frac{1}{1 - \left(\rho_1^2/(1 + \frac{1}{SNR})\right)}$$

$$SNR_{Improvement} \approx \frac{1}{1 - \rho_1^2} \qquad (1.1.17)$$

For

$a_1 = 1$ (The case of an ideal integrator).

$$SNR_{Improvement} = \frac{1}{2(1-\rho_1) + \frac{1}{SNR}} \approx \frac{1}{2(1-\rho_1)}$$

(1.1.18)

Equation (1.1.13) does not give a subjective optimum for a_1 (i.e., it does not give a value for a_1 which is preferred by listeners). In most DPCM and delta modulation systems the integrator cut off is set below the band occupied by the input signal. Hence, a_1 is usually set equal to some value between unity and ρ_1. However, Eq. (1.1.18) shows that the SNR improvement is still near optimum so long as a_1 is greater than ρ_1.

At an 8 kHz sampling rate typical values for ρ_1 are 0.85 for speech low pass filtered at 3.2 kHz and 0.80 for speech limited to the 200 Hz to 3.2 kHz telephone band. Hence, signal to noise improvements of 4 to 6 db over PCM are realized with the integrator cut off set below 200 Hz.

Section 2 Higher Order Prediction

If $M > 1$, the optimum values for the predictor coefficients are obtained by setting the partial derivatives of the error power function, with respect to the $a_{k's}$ equal to zero.

$$E\left(\sigma_i^2\right) = E\left((X_i - \sum_{k=1}^{M} a_k \hat{X}_{i-k})^2\right) = E\left(\left[X_i - \sum_{k=1}^{M} a_k (X_{i-k} + e_{i-k})\right]^2\right) \tag{1.2.1}$$

$$E\left(\frac{d}{da_j}\sigma_i^2\right) = -2\left[E(X_i X_{i-j}) + E(X_i e_{i-j}) - \sum_{k=1}^{M} a_k \left\{E(X_{i-k}X_{i-j}) + E(X_{i-k}e_{i-j}) + E(X_{i-j}e_{i-k}) + E(e_{i-k}e_{i-j})\right\}\right] = 0 \tag{1.2.2}$$

Once again it can be safely assumed that the correlation between the input signal and the quantizing noise is very small; hence, the optimum condition is

$$E(X_i X_{i-j}) = \sum_{k=1}^{M} a_k \left[E(X_{i-k} X_{i-j}) + E(e_{i-k} e_{i-j}) \right] \tag{1.2.3}$$

Given the fact that $E(X_{i-k} X_{i-j})$ is equal to $E(X_{i-j} X_{i-k})$ and $j = 1, 2, 3, \ldots, M$, the above equation can be expressed in the matrix form shown below.

$$\begin{bmatrix} E(X_i X_{i-1}) \\ \cdot \\ \cdot \\ \cdot \\ E(X_i X_{i-j}) \\ \cdot \\ \cdot \\ \cdot \\ E(X_i X_{i-M}) \end{bmatrix} =$$

$$\begin{bmatrix} \{E(X_{i-1}X_{i-1})+E(e_{i-1}e_{i-1})\} & \cdots & \{E(X_{i-1}X_{i-M})+E(e_{i-1}e_{i-M})\} \\ \vdots & & \vdots \\ \{E(X_{i-1}X_{i-M})+E(e_{i-1}e_{i-M})\} & \cdots & \{E(X_{i-M}X_{i-M})+E(e_{i-M}e_{i-M})\} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

(1.2 4)

If it can be assumed that quantizer overload does not occur very often, then it is reasonable to suppose that the correlation between error samples separated in time is small. Hence, Eq. 1.2.4 may be written as follows:

$$\begin{bmatrix} E(X_i X_{i-1}) \\ \vdots \\ E(X_i X_{i-j}) \\ \vdots \\ E(X_i X_{i-M}) \end{bmatrix} =$$

$$\begin{bmatrix} \{E(X_{i-1}X_{i-1}).(1+\frac{1}{SNR})\} & \{E(X_{i-1}X_{i-2})\} & \cdots & \{E(X_{i-1}X_{i-M})\} \\ & \ddots & & \\ & & \{E(X_{i-j}X_{i-j})(1+\frac{1}{SNR})\} & \\ & & & \\ \{E(X_{i-1}X_{i-M})\} & \cdots & \cdots & \{E(X_{i-M}X_{i-M})(1+\frac{1}{SNR})\} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

(1.2.5)

If the system performs well, the overall signal to noise power ratio should be large and the equations defining the optimum predictor coefficients may be written as in 1.2.6.

$$\begin{bmatrix} E(X_iX_{i-1}) \\ \vdots \\ E(X_iX_{i-j}) \\ \vdots \\ E(X_iX_{i-M}) \end{bmatrix} = \begin{bmatrix} E(X_{i-1}X_{i-1})E(X_{i-1}X_{i-2})\cdots E(X_{i-1}X_{i-M}) \\ E(X_{i-1}X_{i-2})E(X_{i-2}X_{i-2}) \\ \\ \\ E(X_{i-1}X_{i-M}) \qquad\qquad E(X_{i-M}X_{i-M}) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

(1.2.6)

The MxM matrices in equation (1.2.4), (1.2.5) and (1.2.6) are symetrical and positive semidefinite; hence, there exists a recursive solution for $(a_1 a_2 \ldots a_M)$ which is computationally more efficient than the solution of a general set of M equations in M unknows[7]. The recursive solution of equation (1.2.6) proceeds as follows:

$$\begin{bmatrix} E(X_i X_{i-1}) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ E(X_i X_{i-M}) \end{bmatrix} = \begin{bmatrix} \mathcal{S} \end{bmatrix} \cdot \begin{bmatrix} \mathcal{S} \end{bmatrix}^t \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_M \end{bmatrix} \qquad (1.2.7)$$

where $\begin{bmatrix} \mathcal{S} \end{bmatrix}$ is a triangular matrix, and $\begin{bmatrix} \mathcal{S} \end{bmatrix}^t$ is its transpose.

$$\begin{bmatrix} \mathcal{S} \end{bmatrix} = \begin{bmatrix} S_{11} & \cdots & S_{1M} \\ & \ddots & \\ & \ddots & \\ 0 & \ddots & \\ & & S_{MM} \end{bmatrix} \qquad (1.2.8)$$

The following relationships (1.2.10 through 1.2.14) can be derived from the rules of matrix multiplication

and the fact that

$$\begin{bmatrix} \mathcal{S} \end{bmatrix} \cdot \begin{bmatrix} \mathcal{S} \end{bmatrix}^{t} = \begin{bmatrix} E(X_{i-1}X_{i-1}) \ldots E(X_{i-1}X_{i-M}) \\ \\ E(X_{i-1}X_{i-M}) \ldots E(X_{i-M}X_{i-M}) \end{bmatrix} . \tag{1.2.9}$$

$$S_{11} = \sqrt{E(X_{i-1}X_{i-1})} \tag{1.2.10}$$

$$S_{1j} = E(X_{i-1}X_{i-j})/S_{11} \tag{1.2.11}$$

$$S_{kk} = \sqrt{E(X_{i-k}X_{i-k}) - \sum_{\ell=1}^{k-1} S_{\ell k}^{2}} \qquad k > 1 \tag{1.2.12}$$

$$S_{kj} = E(X_{i-k}X_{i-j}) - \sum_{\ell=1}^{k-1} S_{\ell K}S_{\ell j} \qquad j > k \tag{1.2.13}$$

$$S_{kj} = 0 \qquad j < k \tag{1.2.14}$$

Solving Eq. (1.2.6) is equivalent to solving two equations.

$$\begin{bmatrix} \mathcal{S} \end{bmatrix} \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_M \end{bmatrix} = \begin{bmatrix} W_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ W_M \end{bmatrix} \quad \text{and} \tag{1.2.15}$$

$$\begin{bmatrix} \mathcal{S} \end{bmatrix}^t \begin{bmatrix} W_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ W_M \end{bmatrix} = \begin{bmatrix} E(X_i X_{i-1}) \\ \cdot \\ \cdot \\ \cdot \\ E(X_i X_{i-M}) \end{bmatrix} , \tag{1.2.16}$$

where

$$W_1 = \frac{E(X_i X_{i-1})}{S_{11}} \quad \text{and} \tag{1.2.17}$$

$$W_j = \frac{E(X_i X_{i-j}) - \sum_{\ell=1}^{j-1} S_{\ell j} W_\ell}{S_{jj}} \quad . \tag{1.2.18}$$

$$k > 1$$

The matrix $\begin{bmatrix} \mathcal{S} \end{bmatrix}$ is calculated from Eq. (1.2.10) through (1.2.13) and the vector $\{W_1 \ldots\ldots W_M\}$ is derived from Eq. (1.2.17) and (1.2.18). The number of operations required is approximately proportional to the square of

the order of prediction, M. The same is true for the final manipulations required in the solution, namely:

$$a_M = W_M / S_{MM} \text{ and} \tag{1.2.19}$$

$$a_j = W_j - \sum_{\ell=j+1}^{M} S_{j\ell}\, E(X_i X_{i-\ell}) \tag{1.2.20}$$

In the solution of M equations in M unknowns by straight forward use of determinants, the number of operations required is roughly proportional to M!; hence, the solution described here is much more efficient than conventional solutions when the order of prediction is large.

It should be noted that the autocovariance functions in Eq. (1.2.4), (1,2,5) and (1.2.6) need not be long term averages. Hence, the coefficients may be calculated periodically from overlapping blocks of input data and the predictor may be adapted to changing statistics.

Another attractive method for adapting the predictor coefficients is the steepest descent gradient search technique [7]. In this method a positive definite function of the error is selected. For continuous signals we have:

$$\delta(t)^2 = (X(t) - \sum_{k=1}^{M} a_k X(t - kT))^2 \qquad (1.2.21)$$

Next the gradient is taken with respect to the vector described by the predictor coefficients.

$$\text{Grad } (\delta(t)^2) = \begin{bmatrix} \frac{d}{da_1} & \delta(t)^2 \\ \frac{d}{da_j} & \delta(t)^2 \\ \frac{d}{da_M} & \delta(t)^2 \end{bmatrix} = - \begin{bmatrix} 2\delta(t)X(t-T) \\ 2\delta(t)X(t-jT) \\ 2\delta(t)X(t-MT) \end{bmatrix}$$

(1.2.22)

The predictor coefficients are then adapted in a direction opposite to the gradient, i.e.,

$$\frac{da_j}{dt} = \gamma' \, \delta(t)X(t-jT) \, , \qquad (1.2.23)$$

where γ' is a positive constant.

In a continuous system this is equivalent to setting the derivative of $\delta(t)^2$ equal to a negative value at all times;

$$\frac{d}{dt}\,\delta(t)^2 = \sum_{k=1}^{M}\left(\frac{d}{da_k}\,\delta(t)^2\right)\cdot\frac{da_k}{dt} \qquad (1.2.24)$$

From Eq. (1.2.21) and (1.2.23) it may be inferred that

$$\frac{d}{dt}\,\delta(t)^2 = -\,2\gamma'\delta(t)^2\cdot\sum_{k=1}^{M}\,[X(t-kT)]^2. \qquad (1.2.25)$$

This constitutes a proof that the error power will be reduced to a minimum (i.e. that the predictor coefficients will approach optimum values).

In a sampled data system where the a_j's are changed at each sampling instant, Eq. (1.2.23) is replaced by:

$$\Delta a_j\big|_{t=iT} = \gamma\delta_i X_{i-j}\ . \qquad (1.2.26)$$

Since δ_i and X_{i-j} are related to the overall signal level, γ is generally made inversely proportional to signal power so that the adaptation speed will be independent of signal level.

$$\Delta a_j\big|_{t=iT} = \frac{K\,\delta_i X_{i-j}}{\sum_{k=1}^{M}(X_{i-k})^2} \qquad (1.2.27)$$

where K is a constant.

In a sampled data system the coefficients can oscillate about the optimum values. If the discrete system operates on a relatively stationary speech sound there will exist a set of optimum predictor coefficients $\bar{\alpha}$ which may be obtained by solving Eq. (1.2.6). After i samples of the signal have been taken, the coefficient vector to be used for the (i + 1) th sample is

$$\bar{a}_{i+1} = \bar{a}_i + \frac{K \delta_i \bar{X}_i}{\|\bar{X}_i\|^2} , \qquad (1.2.28)$$

where

$$\bar{a}_i = \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_M \end{bmatrix} , \quad \bar{X}_i = \begin{bmatrix} X_{i-1} \\ X_{i-2} \\ . \\ . \\ . \\ X_{i-M} \end{bmatrix} ,$$

and

$$\|\bar{X}_i\|^2 = X_{i-1}^2 + X_{i-2}^2 + \ldots + X_{i-M}^2 .$$

If the solution vector, $\overline{\alpha}$ were used, an error, $\delta_{\alpha i}$, would occur. The difference between the error with optimum coefficients and the error at the ith sample with the coefficients obtained from the gradient search is

$$\delta_i = [\overline{\alpha} - \overline{a}_i]^t X_i + \delta_{\alpha i} . \qquad (1.2.29)$$

The difference between the coefficients after the ith sample and the optimum coefficients shall now be defined as the coefficient error vector.

$$\overline{\xi}_i \equiv \overline{a}_i - \overline{\alpha} \qquad (1.2.30)$$

Hence by Eq. (1.2.28) through (1.2.30)

$$\overline{\xi}_{i+1} = \overline{\xi}_i - \frac{k}{\|\overline{X}_i\|^2} \left((\overline{\xi}_i^t \overline{X}_i) \cdot \overline{X}_i - \delta_{\alpha i} \overline{X}_i \right) \qquad (1.2.31)$$

If we take the norm (i.e., the sum of the squares of all of the vector components) on both sides of the previous equation, the behavior of the system becomes more apparent.

$$\|\xi_{i+1}\|^2 = \|\xi_i\|^2 - \frac{2K}{\|\bar{X}_i\|^2}\left[(\bar{\xi}_i{}^t\bar{X}_i)^2 - \delta_{\alpha i}\bar{\xi}_i{}^t\bar{X}_i\right] + \frac{K^2}{\|\bar{X}_i\|^2}\left[(\bar{\xi}_i{}^t\bar{X}_i)^2 - 2\delta_{\alpha i}\bar{\xi}_i{}^t\bar{X}_i + \delta_{\alpha i}{}^2\right] \quad (1.2.32)$$

If K is sufficiently small, then

$$\|\xi_{i+1}\|^2 \simeq \|\xi_i\|^2 - \frac{2K\left[(\bar{\xi}_i{}^t\bar{X}_i)^2 - \delta_{\alpha i}\bar{\xi}_i{}^t\bar{X}_i\right]}{\|\bar{X}_i\|^2} \quad (1.2.33)$$

If $|\delta_{\alpha i}|$ is small compared to $|\bar{\xi}_i{}^t\bar{X}_i|$ (i.e., if the error is large),

$$\|\xi_{i+1}\|^2 \approx \|\xi_i\|^2 - \frac{2K(\bar{\xi}_i{}^t\bar{X}_i)^2}{\|\bar{X}_i\|^2} \quad (1.2.34)$$

By the Schwartz inequality and fact that $K << \frac{1}{2}$ it is obvious that $\|\xi_{i+1}\| < \|\xi_i\|^2$.

Hence, the coefficients adjust toward $\bar{\alpha}$ when the error is larger in magnitude than $\delta_{\alpha i}$. When $|\delta_{\alpha i}|$ is approached the process is slowed by the fact that $|\xi_{i+1}{}^2|$ may be larger than $|\xi_i{}^2|$ when $|\delta_{\alpha i}| > |\bar{\xi}_i{}^t\bar{X}_i|$ and sgn $(\delta_{\alpha i})$ = sgn $(\bar{\xi}_i{}^t\bar{X}_i)$. (See 1.2.33). When the sign of $\delta_{\alpha i}$ is the same as sgn $(\bar{\xi}_i{}^t\bar{X}_i)$ and $|\delta_{\alpha i}| > |\bar{\xi}_i{}^t\bar{X}_i|$, the prediction

error, δ_i is actually less in magnitude than $\delta_{\alpha i}$. (See Eqs. 1.2.29 and 1.2.30). Hence, it is reasonable to expect that if K is sufficiently small and the process remains stationary, the overall prediction error will approach that achieved by direct solution of Eq. (1.2.6).

A simpler gradient search is possible if the magnitude rather than the square of the error is minimized.

$$\frac{d}{da_j} \left| \delta_i \operatorname{sgn} (\delta_i) \right| \Big|_{t=iT} = - X_{i-j} \operatorname{sgn} (\delta_i) \quad (1.2.35)$$

$$\Delta a_j \Big|_{t=iT} = \frac{K X_{i-j} \operatorname{sgn} (\delta_i)}{\|\bar{X}_i\|} \quad (1.2.36)$$

where

$$\|\bar{X}_i\| = \sum_{k=1}^{M} |X_{i-k}| \quad (1.2.37)$$

Hence,

$$\bar{a}_{i+1} = \bar{a}_i + \frac{K \bar{X}_i \operatorname{sgn} (\delta_i)}{\|\bar{X}_i\|} \quad (1.2.38)$$

$$\bar{\xi}_{i+1} = \bar{\xi}_i + \frac{K \bar{X}_i \operatorname{sgn} (\delta_{\alpha i} - \bar{\xi}_i {}^t \bar{X}_i)}{\|\bar{X}_i\|} \quad (1.2.39)$$

$$\|\xi_{i+1}\|^2 = \|\xi_i\|^2 + \frac{2K\,\overline{\xi}_i{}^t\overline{X}_i\ \mathrm{sgn}\,(\delta_{\alpha i} - \overline{\xi}_i{}^t\overline{X}_i)}{\|X_i\|} + \frac{K^2|X_i{}^2|}{(\|X_i\|)^2} \qquad (1.2.40)$$

For small K

$$\|\xi_{i+1}\|^2 = \|\xi_i\|^2 + \frac{2K\,\overline{\xi}_i{}^t\overline{X}_i\ \mathrm{sgn}\,(\delta_{\alpha i} - \overline{\xi}_i{}^t\overline{X}_i)}{\|X_i\|} \qquad (1.2.41)$$

When $\delta_{\alpha i} < \xi_i{}^t X_i$ i.e., when the coefficients are not near $\overline{\alpha}$.

$$\|\xi_{i+1}\|^2 = \|\xi_i\|^2 - \frac{2K\,\|\xi_i{}^t X_i\|}{\|X_i\|} \qquad (1.2.42)$$

Hence, if K is sufficiently small $|\xi_{i+1}{}^2|$ will be less than $|\xi_i{}^2|$. When $|\overline{\xi}^t\overline{X}|$ approaches $|\delta_{\alpha i}|$ convergence is slowed to a halt. It is reasonable to expect that δ^2, the average error, will be approximately equal to $\delta_\alpha{}^2$ if the process remains stationary.

Once again the proof of convergence is dependent upon K being sufficiently small. Unfortunately, the rate at which the coefficients can change is proportional to K.

Hence, selection of the best value for K involves a compromise between the ability to track changes in speech statistics and the ability to converge to the vicinity of $\overline{\alpha}$.

The simulations (Part II Section 3) have shown that the two search techniques described above perform equally well on speech. The signal to noise improvement realized using gradient adaptation of the predictor coefficients at an 8 kHz sampling rate was 3 to 6 db less than that reported by other researchers using optimum coefficients [2, 17]. The sampling rate was increased from 8 kHz to 16 kHz and the predictor was left unchanged. (i.e., the taps on the delay line in the predictor were spaced two sampling intervals apart). With K reduced to half of its previous value, the search was carried out at a 16 kHz rate as indicated by Eq. (1.2.43).

$$\Delta a_j \Big|_{t=iT} = \frac{K \delta_i X_{i-2j}}{\sum_{k=1}^{M} |X_{i-2k}|} \qquad (1.2.43)$$

A SNR improvement approximately equal to that obtained by B. S. Atal [2] and P. Noll [17] was thus achieved. The most interesting conclusion to be drawn from this experiment

is that when the rate of computation required by the gradient search approaches that needed to calculate the autocovariance functions and to solve Eq. (1.2.6) once every 5 to 10 millisecond, the results are the same. The steepest-descent gradient search involves only one simple algorithm ; therefore, it may be a more efficient approach than solutions involving Eq. (1.2.6).

Given an adequate solution to Eq. (1.2.6), an approximate formula for the signal to noise improvement gained by the predictor can be derived as follows: First, Eq. (1.2.1) is rewritten as:

$$E(\delta_i^2) = E(X_i^2) - 2\sum_{k=1}^{M} a_k[E(X_i X_{i-k}) + E(X_i e_{i-k})] + E\left(\sum_{k=1}^{M} (a_k(X_{i-k} + e_{i-k}))^2\right) \qquad (1.2.44)$$

If the overall signal to noise ratio is large enough and if the correlation between the signal and the noise is small enough, the noise terms may be dropped from (1.2.44).

$$E(\delta_i^2) \simeq E(X_i^2) - 2\sum_{k=1}^{M} a_k\, E(X_i X_{i-k}) + E\left(\left(\sum_{k=1}^{M} a_k\, X_{i-k}\right)^2\right) \qquad (1.2.45)$$

where

$$E\Big(\Big(\sum_{k=1}^{M} a_k \, X_{i-k} \Big)^2 \Big) = E\Big(\sum_{k=1}^{M} \Big(\sum_{j=1}^{M} a_j \, a_k X_{i-k} X_{i-j} \Big) \Big) \quad \text{or}$$

$$E\Big(\Big(\sum_{k=1}^{M} a_k X_{i-k} \Big)^2 \Big) = \sum_{k=1}^{M} a_k \Big(\sum_{j=1}^{M} a_j \, E(X_{i-k} X_{i-j}) \Big) \, . \quad (1.2.46)$$

If the noise terms are dropped, Eq. (1.2.3) may be written as:

$$E(X_i X_{i-k}) = \sum_{j=1}^{M} a_j \, E(X_{i-k} X_{i-j}) \quad (1.2.47)$$

which is of course equivalent to (1.2.6). Hence, (1.2.46) may be rewritten as:

$$E\Big(\Big(\sum_{k=1}^{M} a_k \, X_{i-k} \Big)^2 \Big) = \sum_{k=1}^{M} a_k \, E(X_i X_{i-k}) \quad (1.2.48)$$

and (1.2.45) as:

$$E(\delta_i^{\,2}) = E(X_i^{\,2}) - \sum_{k=1}^{M} a_k \, E(X_i X_{i-k}) \quad (1.2.49)$$

The signal to noise improvement factor for a DPCM with a higher order predictor is therefore given as:

$$SNR_{Improvement} = \frac{E\left(X_i^2\right)}{E\left(\delta_i^2\right)} = \frac{1}{1 - \sum_{k=1}^{M} a_k \frac{E(X_i X_{i-k})}{E\left(X_i^2\right)}} \qquad (1.2.50)$$

$$SNR_{Improvement} = \frac{1}{\left(1 - \sum_{k=1}^{M} a_k \rho_k\right)} \qquad (1.2.51)$$

Section 3 The Optimum Fixed Quantizer

If the quantizer in Fig. 3 is replaced by a short circuit, then

$$\hat{X} = X \text{ and} \tag{1.3.1}$$

$$\delta = \left[1 - \sum_{k=1}^{M} a_K Z^{-k} \right] X \ . \tag{1.3.2}$$

Hence, a signal like that expected at the quantizer can be generated by linearly filtering the input signal. In the case where $M = 1$ and $a_1 = 1$,

$$\delta = (1 - Z^{-1})X \quad \text{or} \tag{1.3.3}$$

$$\delta_i = X_i - X_{i-1} \tag{1.3.4}$$

Another way to describe this relationship is the following:

$$\delta = T\left(\frac{dX}{dt}\right) \tag{1.3.5}$$

where T is the sampling interval.

In order to design an optimum quantizer, the probability distribution of δ, $P(\delta)$, must be obtained. This may be done by simulating the DPCM coder with a uniform quantizer and making a histogram of the percent of the time that the various quantizer levels are used,

or it may be done by directly processing the difference signal, δ, obtain by linear filtering.

An analysis similar to that presented next was published by Joel Max [13] in 1960. The error power of the quantizer in Fig. 4 may be described as follows:

$$E(e^2) = E\left((\delta-\hat{\delta})^2\right) = \sum_{j=1}^{N} \int_{\delta_j}^{\delta_{j+1}} (\delta-\hat{\delta}_j)^2 P(\delta)d\delta$$

where $-\delta_1 = \delta_{N+1} = \infty$. (1.3.6)

The minimum conditions on Eq. (1.3.6) occur when the derivatives with respect to δ_j and $\hat{\delta}_j$ are equal to zero.

$$\frac{dE(e^2)}{d\delta_j} = \frac{d}{d\delta_j}\int_{\delta_{j-1}}^{\delta_j} (\delta-\hat{\delta}_{j-1})^2 P(\delta)d\delta + \frac{d}{d\delta_j}\int_{\delta_j}^{\delta_{j+1}} (\delta-\hat{\delta}_j)^2 P(\delta)d\delta \tag{1.3.7}$$

By the fundamental theorem of calculus,

$$\frac{dE(e^2)}{d\delta_j} = (\delta_j-\hat{\delta}_{j-1})^2 P(\delta_j) - (\delta_j-\hat{\delta}_j)^2 P(\delta_j) = 0 \tag{1.3.8}$$

Therefore if $P(\delta_j)\neq 0$, it can easily be shown that the first of the two minimum conditions is given simply as:

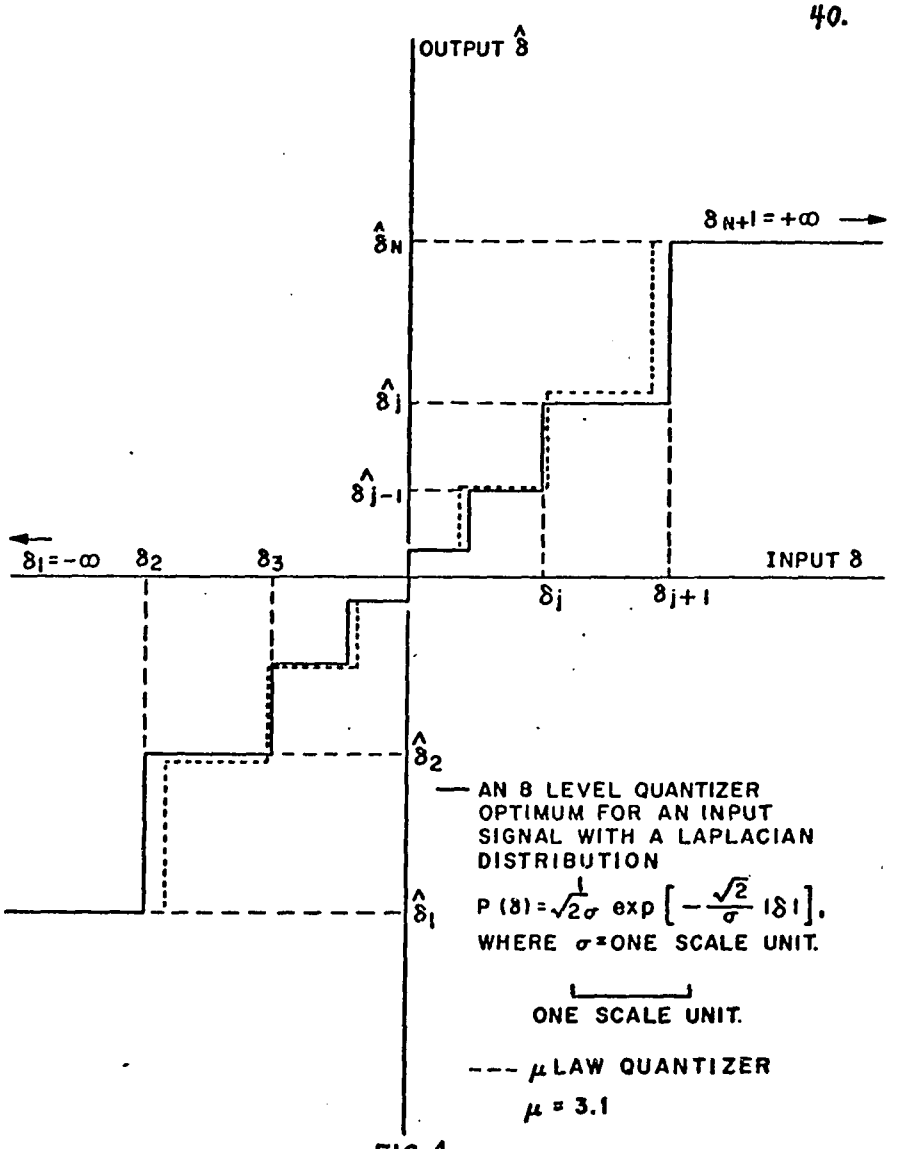

OUTPUT $\hat{\delta}$
$\delta_{N+1} = +\infty$
$\hat{\delta}_N$
$\hat{\delta}_j$
$\hat{\delta}_{j-1}$
$\delta_1 = -\infty$
δ_2
δ_3
INPUT δ
δ_j
δ_{j+1}
$\hat{\delta}_2$
$\hat{\delta}_1$
— AN 8 LEVEL QUANTIZER OPTIMUM FOR AN INPUT SIGNAL WITH A LAPLACIAN DISTRIBUTION
$P(\delta) = \frac{1}{\sqrt{2}\sigma} \exp\left[-\frac{\sqrt{2}}{\sigma}|\delta|\right]$,
WHERE σ = ONE SCALE UNIT.
ONE SCALE UNIT.
--- μ LAW QUANTIZER
$\mu = 3.1$

FIG. 4

$$\delta_j = \frac{\hat{\delta}_j + \hat{\delta}_{j-1}}{2} \tag{1.3.9}$$

Hence, each quantizer step occurs at an input level, δ_j, which is equal to one half the sum of the adjacent output levels, $\hat{\delta}_{j-1}$ and $\hat{\delta}_j$.

$$\frac{dE(e^2)}{d\hat{\delta}_j} = \frac{d}{d\hat{\delta}_j} \int_{\delta_j}^{\delta_{j+1}} (\delta - \hat{\delta}_j)^2 P(\delta) d\delta = 0 \tag{1.3.10}$$

Hence,

$$\int_{\delta_j}^{\delta_{j+1}} (\delta - \hat{\delta}_j) P(\delta) d\delta = 0 \tag{1.3.11}$$

The other requirement (see Eqs. 1.3.10 and 1.3.11) is that each output level $\hat{\delta}_j$ be located at the centroid of the probability distribution in the interval δ_j to δ_{j+1}. Equations (1.3.9) and (1.3.11) describe the overall geometry of the optimum quantizer. To obtain its absolute dimensions, one interval δ_j to δ_{j+1} must be obtained by a linear search involving Eqs. (1.3.9) and (1.3.11). For example, we may start with $\delta_1 = -\infty$ and arbitrarily choose a value for δ_2. Then by satisfying Eq. (1.3.11), we obtain a value for $\hat{\delta}_1$, the most negative quantizer output level. Then, by setting $j = 2$ and

applying Eq. (1.3.9) the next output level $\hat{\delta}_2$ may be calculated. Then the next input interval, δ_2 to δ_3, is calculated from Eq. (1.3.11). The process is repeated until the Nth step is reached and the integral in Eq. (1.3.11) is taken from δ_N to $+\infty$. If this last integration results in a positive answer, $|\delta_2|$ is decreased and the procedure is repeated. If the final integration gives a negative result $|\delta_2|$ is increased. Since most of the probability distributions encountered in speech processing decrease monotonically as $|\delta|$ increases, only one optimum value exists for δ_2, and this can easily be found by applying the search procedure outlined here.

The quantizer shown in Fig. 4 (solid lines) was derived with the aid of a computer program by M. D. Paez and T. H. Glisson [18] at the Electrical Engineering Department of North Carolina State University. Paez and Glisson also derived optimum quantizers for signals with Gamma distributions. Joel Max derived a set of optimum quantizers for signals with Gaussian probability distributions. In addition optimum uniform quantizers (i.e., quantizers where $(\delta_3-\delta_2) = (\delta_4-\delta_3)$ .... $= (\delta_N-\delta_{N-1}) = (\hat{\delta}_2-\hat{\delta}_1) = \ldots = (\hat{\delta}_N-\hat{\delta}_{N-1})$ have been derived for uniformly, Gaussian, Laplacian, and Gamma distributed signals.

In Fig. 5 the maximum signal to noise ratios attainable with various quantizers are plotted as a function of the number of quantization steps, N. In this case the uniform quantizer is processing a uniformly distributed signal, the optimum Gaussian a Gaussian distributed signal, etc. The optimum Gamma distribution apparently matches the amplitude probability statistics of speech best. (See Fig. 6.)

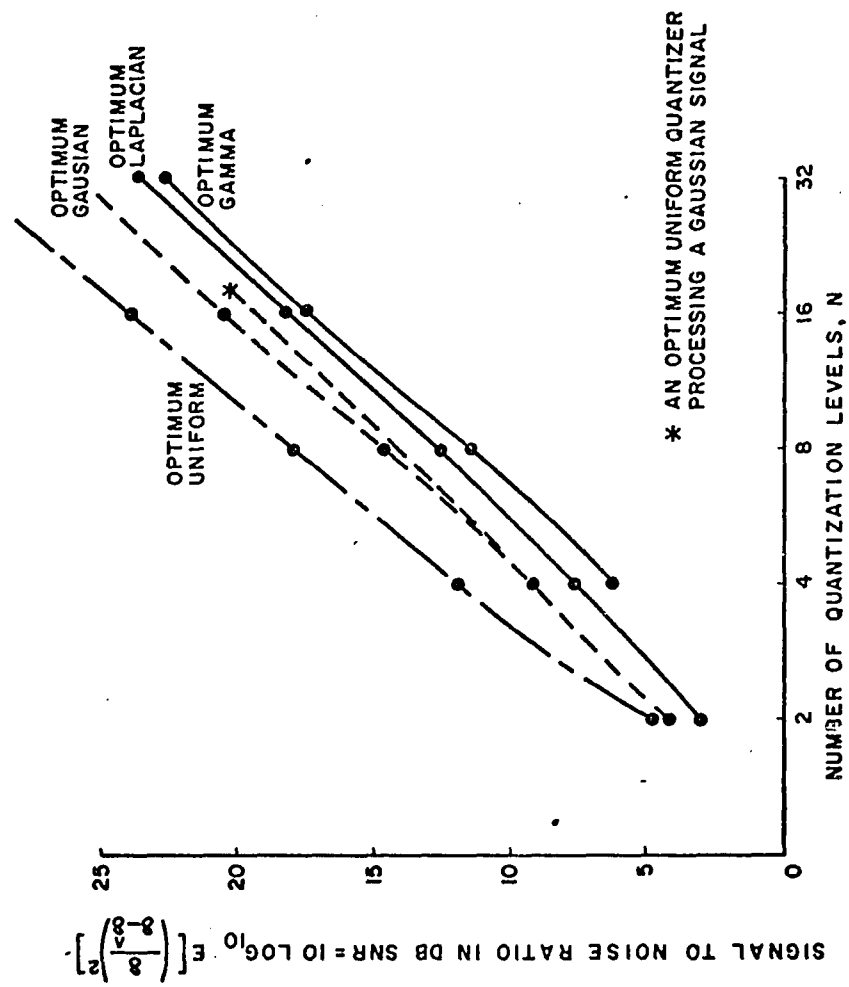

OPTIMUM GAUSIAN
OPTIMUM LAPLACIAN
OPTIMUM GAMMA
OPTIMUM UNIFORM
* AN OPTIMUM UNIFORM QUANTIZER PROCESSING A GAUSSIAN SIGNAL
NUMBER OF QUANTIZATION LEVELS, N
SIGNAL TO NOISE RATIO IN DB SNR = 10 LOG10 E[(σ/(σ−σ̂))²]
0
5
10
15
20
25
2
4
8
16
32

FIG. 5

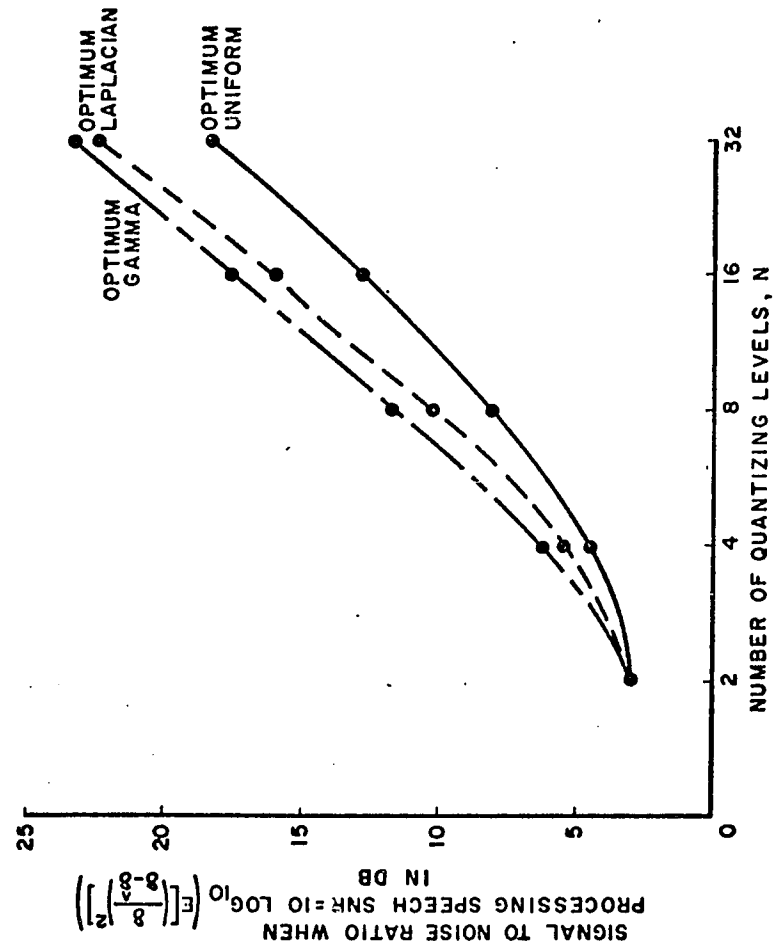

OPTIMUM LAPLACIAN
OPTIMUM UNIFORM
OPTIMUM GAMMA
25
20
15
10
5
0
2
4
8
16
32
NUMBER OF QUANTIZING LEVELS, N
SIGNAL TO NOISE RATIO WHEN PROCESSING SPEECH SNR = 10 LOG10 (E[(σ/σ-σ>)^2])
IN DB

FIG. 6

The distributions encountered in speech processing are generally not quite as favorable as the Gaussian. Distributions that are more uniform than the Gaussian are not likely to be encountered.

The most widely used form of quantization, μ law quantization, is nearly optimum for Laplacian distributed signals. This form of quantization may be realized using a uniform quantizer operating in series with matched non-linear elements, namely; a compressor and an expandor (see Fig. 7). The first assumption used in deriving the μ law quantizer or rather μ law compressor and expandor characteristics is that the probability distribution, $P(\delta)$, is constant across the quantizer intervals, δ_j to δ_{j+1}. Despite the fact that this assumption does not appear to be valid when N is small, it holds up rather well provided that an optimum value is calculated for μ. When the above assumption is applied to Eq. 1.3.11, the following relationships are obtained.

$$\int_{\delta_j}^{\delta_{j+1}} (\delta - \hat{\delta}_j) P(\delta_j^*) d\delta = 0 \ , \qquad (1.3.12)$$

where $$\delta_{j+1} > \delta_j^* > \delta_j .$$

Evaluating the integral we have,

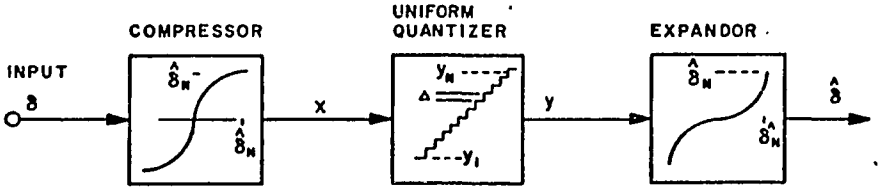


FIG. 7

$$\frac{P(\delta_j^*)}{2}\left[(\delta_{j+1} - \hat{\delta}_j)^2 - (\delta_j - \hat{\delta}_j)^2\right] = 0 \qquad (1.3.13)$$

If $P(\hat{\delta}_j) \neq 0$ then

$$\hat{\delta}_j = \frac{\delta_{j+1} + \delta_j}{2} \qquad (1.3.14)$$

and (1.3.9) is no longer retained, since (1.3.14) and (1.3.9) define a uniform quantizer. It should be noted that Eq. (1.3.14) does not apply to the end steps where $\delta_{j+1} = +\infty$ or $\delta_j = -\infty$.

At this point in the discussion it is convenient to introduce a new variable,

$$\Delta_j = \delta_{j+1} - \delta_j \qquad (1.3.15)$$

and to set

$$\delta_j^* = \hat{\delta}_j = \delta_j + \Delta_j/2 \qquad (1.3.16)$$

The total quantizer noise power is given by Panter [19] as

$$E(e^2) = \sum_{j=1}^{N} \frac{\Delta_j^3}{12} P(\hat{\delta}_j) \qquad (1.3.17)$$

Panter also derives an optimum condition for the quantizer as follows. First he defines

$$M(\hat{\delta}_j) \equiv \left[P(\hat{\delta}_j)\right]^{\frac{1}{3}} \cdot \Delta_j \, . \qquad (1.3.18)$$

Therefore,

$$E(e^2) = \frac{1}{12} \sum_{j=1}^{N} M(\hat{\delta}_j) \qquad (1.3.19)$$

Using Lagrange's method of undetermined multipliers it is possible to show that the error is minimum when

$$M(\hat{\delta}_1) = M(\hat{\delta}_2) = \ldots = M(\hat{\delta}_j) = M(\hat{\delta}_{j+1}) = \ldots = M(\hat{\delta}_N) = \frac{K}{N} \qquad (1.3.20)$$

where $\underline{K}$ is a constant equal to the following.

$$K = \sum_{j=1}^{N} M(\hat{\delta}_j) \qquad (1.3.21)$$

Hence,

$$E(e^2)_{Min} = \frac{1}{12} \frac{K^3}{N^2} = \frac{1}{12N^2} \left[\sum_{j=1}^{N} M(\hat{\delta}_j) \right]^3 \qquad (1.3.22)$$

$$E(e^2)_{Min} = \frac{1}{12N^2} \left[\sum_{j=1}^{N} P(\hat{\delta}_j)^{\frac{1}{3}} \Delta_j \right]^3 , \qquad (1.3.23)$$

and $$E(e^2)_{Min} \approx \frac{1}{12N^2} \left[\int_{-\infty}^{\infty} \left[P(\delta) \right]^{\frac{1}{3}} d\delta \right]^3 \approx \frac{1}{12N^2} \left[\int_{-\infty}^{\infty} M(\delta) d\delta \right]^3 \qquad (1.3.24)$$

where

$$M(\delta) = P(\delta)^{\frac{1}{3}}$$

If $\underline{N}$ is assumed to be even, and if $P(\delta)$ is symmetrical about $\delta = 0$, then

$$\hat{\delta}_j = \Delta_{\frac{N}{2}} + \Delta_{\frac{N}{2}+1} + \cdots \Delta_j \Big/ 2 \qquad (1.3.25)$$

where $\hat{\delta}_j > 0$, and the negative quantizer steps are the same as the positive steps. Next Panter recalls that Eq. (1.3.18) is equal to K/N. Hence,

$$\hat{\delta}_j = \frac{K}{N}\left[\frac{1}{\left[P(\hat{\delta}_{N/2})\right]^{\frac{1}{3}}} + \frac{1}{\left[\left(P\ \hat{\delta}_{\frac{N}{2}} + 1\right)\right]^{\frac{1}{3}}} + \cdots \frac{1}{2\left[P(\hat{\delta}_j)\right]^{\frac{1}{3}}}\right] \qquad (1.3.26)$$

At this point a plausible relationship between Panter's equation above and μ law quantization will be shown. First, the statistics of the input signal to the quantizer are approximated by a Laplacian distribution.

$$P(\delta) = \frac{\alpha}{2} e^{-\alpha |\delta|} \qquad (1.3.27)$$

Since the μ law is normally associated with compressor and expandor characteristics, the expandor required in conjunction with a uniform quantizer is derived first.

We start with the assumption that the compressor acts to convert the input signal δ into a signal, x, with a flat distribution of the quantity $M(\delta)$ (See Fig. 7). In order to return to the statistical properties of the input signal, the uniform quantizer output steps, y_j, must be converted to levels spaced in agreement with equation (1.3.26). This is accomplished by substituting

$$y_{\frac{N}{2}+1},\ y_{\frac{N}{2}+2},\ \ldots,\ y_j \quad \text{for} \quad \hat{\delta}_{\frac{N}{2}+1},\ \ldots\ldots,\ \hat{\delta}_j$$

Then the summation is easily replaced by an integral expression where the $\left(y_{\frac{N}{2}+1} \ \ldots\ y_j\right)$ are replaced by the continuous variable, $\underline{y}$.

$$\hat{\delta} = A\int_0^{y} \frac{d\delta}{\left[P(\delta)\right]^{\frac{1}{3}}} \tag{1.3.28}$$

The constant A is selected so that

$$\hat{\delta}_N = y_N \tag{1.3.29}$$

If equation (1.3.27) is substituted for $P(\delta)$ in (1.3.28) and if the integration is performed, the following expandor characteristic is obtained.

$$\hat{\delta} = \frac{3A}{\alpha}\left(\frac{2}{\alpha}\right)^{\frac{1}{3}}\left[e^{\frac{\alpha y}{3}} - 1\right], \tag{1.3.30}$$

or

$$\hat{\delta} = \frac{\hat{\delta}_N}{\left(e^{\frac{\alpha\hat{\delta}_N}{3}} - 1\right)}\left[e^{\frac{\alpha y}{3}} - 1\right], \qquad (1.3.31)$$

where $\hat{\delta} > 0$ and $y \geq 0$.

The input is limited at $\pm(\delta_N + \Delta_N) = \pm\left(\hat{\delta}_N + \frac{\Delta_N}{2}\right) \approx \pm\hat{\delta}_N$ the values at which the output is limited. When the signal to the expandor is equal to $\pm\hat{\delta}_N$ the output is $\pm\hat{\delta}_N$; the same is true of the compressor. If $\mathbf{x}$ is substituted for y and δ is substituted for $\hat{\delta}$ in equation (1.3.31), the compressor characteristic can then be obtained by solving for $\mathbf{x}$.

$$\mathbf{x} = \frac{3}{\alpha} \ln\left[\left(e^{\frac{\alpha\hat{\delta}_N}{3}} - 1\right)\frac{\delta}{\hat{\delta}_N} + 1\right] \qquad (1.3.32)$$

$$X, \delta > 0$$

The optimum value for the conventional compressor parameter μ is given by

$$\mu_{Opt} = \left(e^{\frac{\alpha\hat{\delta}_N}{3}} - 1\right) \approx \left(e^{\frac{\sqrt{2}\,\hat{\delta}_N}{3\,\sigma}} - 1\right) \qquad (1.3.33)$$

where $\underline{\sigma}$ is the standard deviation of the input signal. If μ is set equal to μ_{Opt} and Eq. (1.3.33) is substituted into (1.3.32), the classic μ law compressor formulas are obtained.

$$x = \hat{\delta}_N \left[\ln\left(\mu \frac{\delta}{\hat{\delta}_N} + 1\right) / \ln(\mu + 1) \right]$$

$$\delta_N > \delta > 0 \qquad (1.3.34)$$

and for $\delta < 0$

$$x = -\hat{\delta}_N \left[\ln\left(-\mu \frac{\delta}{\hat{\delta}_N} + 1\right) / \ln(\mu + 1) \right] \qquad (1.3.35)$$

The design of an optimum μ law quantizer proceeds as follows. First, a trial value is selected for $\hat{\delta}_N$ and y_N is set equal to $\hat{\delta}_N$ (See Eq. 1.3.29). Then for uniformly spaced y_j the output levels, $\hat{\delta}_j$, are computed using Eq. (1.3.31). Likewise, the uniformly spaced decision levels, x_j, (where $x_j = \frac{y_j + y_{j-1}}{2}$) are used to compute the input decision levels, δ_j. These may also be calculated from (Eq. 1.3.29) when x_j is substituted for y and δ_j for $\hat{\delta}$. The noise power is then computed from Eq. (1.3.6) and the process is repeated with a new value for $\hat{\delta}_N$. By a trial and error search an optimum for $\hat{\delta}_N$ can easily be obtained.

An 8 level μ law quantizer optimized for a Laplacian distributed signal is shown in dotted lines in

Fig. 4. Its performance is only slightly inferior to the optimum Max quantizer. The same is true of other quantizers that are more easily realized than Max's.

In PCM systems with 64 or 128 levels, μ is usually set equal to a value greater than or equal to 100. The values used are greater than the optimum value given by Eq. (1.3.33). The reason for using larger values of μ is that the PCM systems become less sensitive to fluctuations in overall signal level when μ is large. When μ is increased above optimum the average signal to noise ratio is lowered by a few db and the dynamic range[1] is improved. This points out a weakness in systems with fixed optimum quantizers. Such systems are quite sensitive to fluctuations in overall signal level and to a lesser degree to variations in speech statistics among talkers.

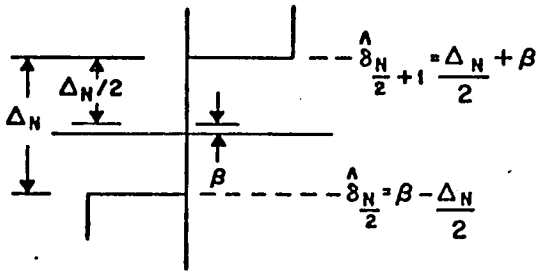
Another problem encountered in DPCM coders with fixed quantizers is that the quantizer output levels nearest the zero level must be accurately specified. If $\left| \hat{\delta}_{\frac{N}{2}} \right|$ and $\left| \hat{\delta}_{\frac{N}{2}+1} \right|$ are not equal, oscillations like those

---

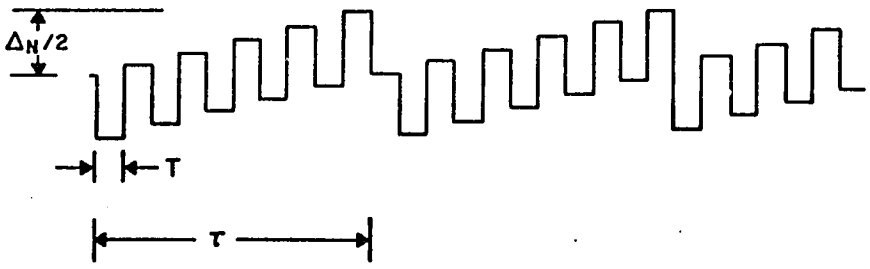
(1) The dynamic range of a system is generally defined as the range of input signal levels for which the signal to noise ratio remains within 3 db of the maximum.

shown in Fig. 8 will occur at the output of the integrator in the absence of an input signal. As the difference between $\left|\hat{\delta}_{\frac{N}{2}}\right|$ and $\left|\hat{\delta}_{\frac{N}{2}+1}\right|$ is reduced, the frequency of the idle channel oscillations is lowered and the chances for eliminating the idling noise with a high pass filter are improved. If the idle channel oscillations should occur in the telephone band, the RMS signal to idle noise ratio of the DPCM coder with a fixed, optimum, Laplacian quantizer is only 29 db when N = 16 and 23 db when N = 8. In view of the fact that the idle channel noise is a quasi-periodic phenomena with most of its power concentrated at a few frequencies, the signal to idle channel noise ratio must be increased to about 60 db if telephone quality transmission is to be approached.

IMPERFECT QUANTIZER



THE IDLE CHANNEL OSCILLATION RESULTING FROM THE IMPERFECTION, β



$$\tau \text{ AVERAGE} = \frac{\Delta_N}{2\beta} T$$

FIG. 6

Section 4 Delta Modulation

Both of the problems cited in the last section can be eliminated by the application of adaptive quantization. Adaptive quantization is in wide use in only the simplest of DPCM systems, the delta modulator. In delta encoding, the difference between the input signal and the voltage on a predictor is sampled at a rate well above the Nyquist frequency: hence, the autocovariance at one sample delay, ρ_1, is much closer to unity than for a delay of one Nyquist interval. The signal to noise improvement given by Eq. (1.1.17) is therefore much larger, permitting the use of a simple two level quantizer.

If the linear delta modulator shown in Fig. 9a is compared with the DPCM system shown in Fig. 3, it becomes obvious that the delta encoder is simply a DPCM encoder with a two level quantizer. For first order prediction (M = 1), the transfer function of the predictor and the positive feedback loop associated with it in Fig. 3 is given by

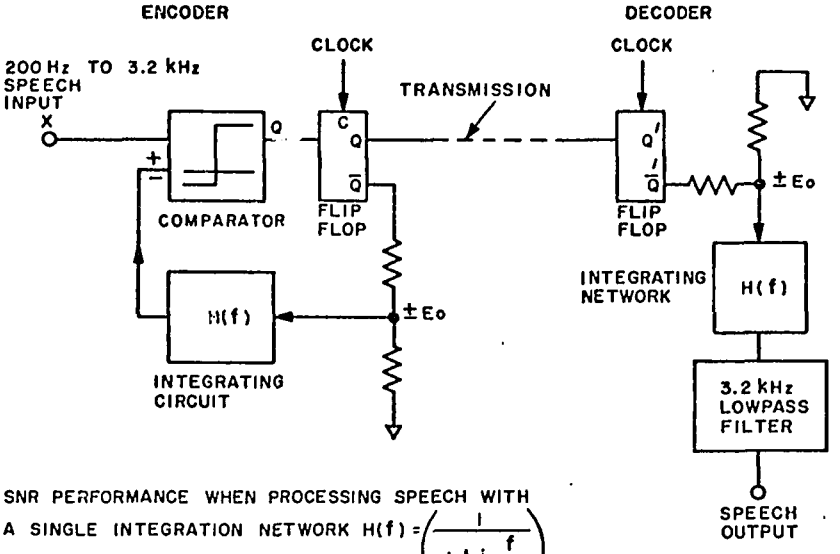
$$\frac{\underset{\delta}{\hat{Y}}}{}(Z) = \frac{a_1 Z^{-1}}{1 - a_1 Z^{-1}} \quad . \qquad (1.4.1)$$

The corresponding Laplace transform is given next.

$$\frac{\underset{\delta}{\hat{Y}}}{}(S) = \frac{a_1 e^{-ST}}{S + \frac{1}{T}\ell_N(a_1)} \quad , \qquad (1.4.2)$$

A LINEAR DELTA MODULATOR

(a)



SNR PERFORMANCE WHEN PROCESSING SPEECH WITH A SINGLE INTEGRATION NETWORK $H(f) = \left(\frac{1}{1 + j\frac{f}{fc}}\right)$

(b)

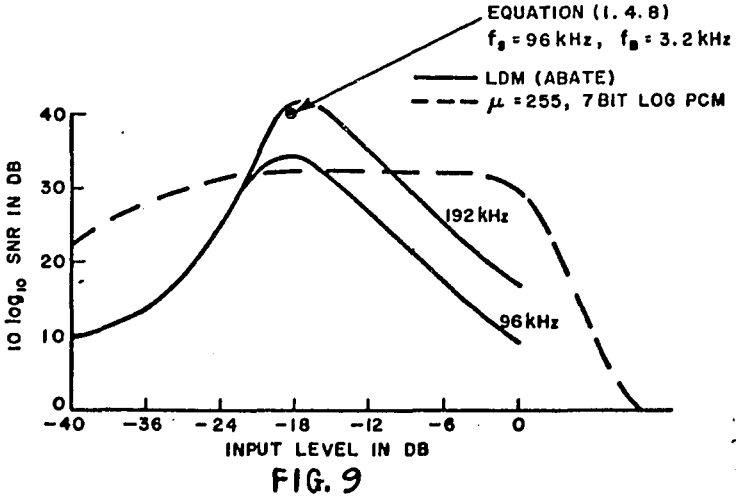


FIG. 9

or

$$\frac{y}{\hat{\delta}}(S) = \frac{a_1 e^{-ST}}{S + 2\pi f_c} \quad . \tag{1.4.3}$$

Equations (1.4.2) and (1.4.3) indicate that the predictor and the positive feedback loop associated with it may be replaced by an integrator.

$$H_1(S) \simeq \frac{a_1}{S + 2\pi f_c} \tag{1.4.4}$$

The delay in equation (1.4.3) implies that the output of the integrator will equal $a_1\hat{X}_i$ at the i + 1th sampling instant. Since a_1 is always near unity, the signal at the integrator is a delayed approximation to the input. The encoder is replaced by a differential comparator and a flip-flop. The comparator is connected so that the sign of the difference signal, sgn(δ), is stored on the flip-flop at the sampling instant (i.e., when the flip-flop is clocked). The equivalent quantizer step is determined by the network at the input to the integrator.

$$\Delta = \hat{\delta}_2 - \hat{\delta}_1 = 2E_0 \tag{1.4.5}$$

A comparison of the signal to noise characteristics of 7 bit log PCM and linear delta modulation with a first order predictor (i.e., a single integrator) is

shown in Fig. 9b. The signal to noise curves for the delta modulator were calculated from J. Abate's results,[1] assuming that speech can be modeled as a random process with a Laplacian amplitude probability distribution and a spectrum which is flat below 400 Hz, integrated from 400 Hz to 3.2 kHz and sharply low pass filtered at 3.2 kHz.

Although the model applied to J. Abate's formulas is a farily good approximation of speech, it is more accurate to assume that the speech power spectrum is integrated twice above 1 kHz [9]. Abate's results are therefore slightly pessimistic.

The signal to noise ratio at the optimum signal level or optimum quantization step size can be calculated more accurately by applying the theory developed in sections one and three. The signal to noise ratio as defined in section one can be calculated from the signal to noise improvement formula (1.1.17) and the fact that a two level optimum Max quantizer operating on the derivative of the speech signal offers a 3 db signal to noise ratio. However, much of the quantizing noise generated in delta modulation is rejected by the desampling filter. The noise power spectrum of an optimum linear delta modulator is relatively flat in the band from 200 Hz to one half the sampling frequency. Hence,

the fraction of the noise in the pass-band is equal to the cut-off frequency of the filter divided by one half of the sampling frequency. The formula for signal to noise for a delta modulator is therefore

$$SNR = \frac{2f_S}{2f_B(1 - \rho_1^2)} \quad , \tag{1.4.6}$$

where f_S is the sampling frequency and f_B is the cut-off frequency of the de-sampling filter. If the sampling frequency is greater than six times the signal bandwidth, the following relationships hold.

$$\frac{1}{1 - \rho_1^2} \approx 1.5\left(\frac{f_S}{2f_B}\right)^2 \tag{1.4.7}$$

$$SNR = \frac{3}{\left(\frac{f_S}{2f_B}\right)^3} \tag{1.4.8}$$

F. DeJager [21] has shown that a linear delta modulator can reproduce speech with a quality equal to that of 7 bit log PCM while operating at a sampling rate of about 120 kHz.

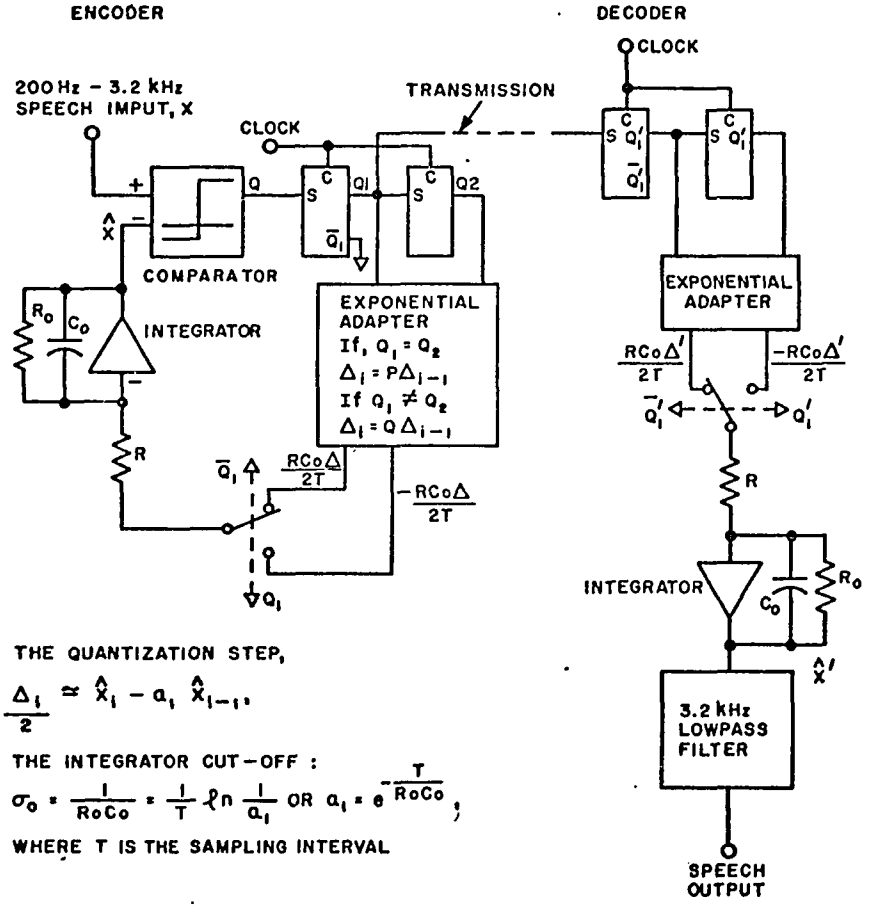
By adapting the step size to match changing signal levels the optimum SNR performance of a linear delta modulator can be extended over a range of input levels. Hence, adaptation improves the performance of delta

modulators by extending the dynamic range and by squelching idle channel noise. Many logics have been devised for adapting the quantization step; these differ primarily in the rate of adaptation.

An example of fast or instantaneous adaptation is N. S. Jayant's [11] exponentially adaptive delta modulator with a one bit memory (see Fig. 10). In this delta modulator, the quantization step is multiplied by $P > 1$ if the present and previous bits are of the same sign and by $Q < 1$ if they are different. Due to the fact that adaptation occurs at every sampling instant, only a single integrator can be used for reasons of stability. Jayant has shown by mathematical derivations and by simulations that the quantizing error is minimized when $P = \frac{1}{Q} = 1.5$. The signal to noise measurements obtained in Jayant's simulations while processing speech compare favorable with J. Abate's results. This indicates that the adaptation keeps the step size very close to optimum and that Abate's stochastic model for the speech signal is a bit pessimistic (see Fig. 11). On the other hand, equation (1.4.8) yields results that are within 1 db of Jayant's.

A careful study of a hardware model built by the author [4] has shown that the exponentially adaptive

AN EXPONENTIALLY ADAPTIVE DELTA MODULATOR WITH A ONE BIT MEMORY



THE QUANTIZATION STEP,

$$\frac{\Delta_i}{2} \simeq \hat{X}_i - a_1 \hat{X}_{i-1},$$

THE INTEGRATOR CUT-OFF :

$$\sigma_0 = \frac{1}{RoCo} = \frac{1}{T} \ell n \frac{1}{a_1} \text{ OR } a_1 = e^{-\frac{T}{RoCo}};$$

WHERE T IS THE SAMPLING INTERVAL

FIG. 10

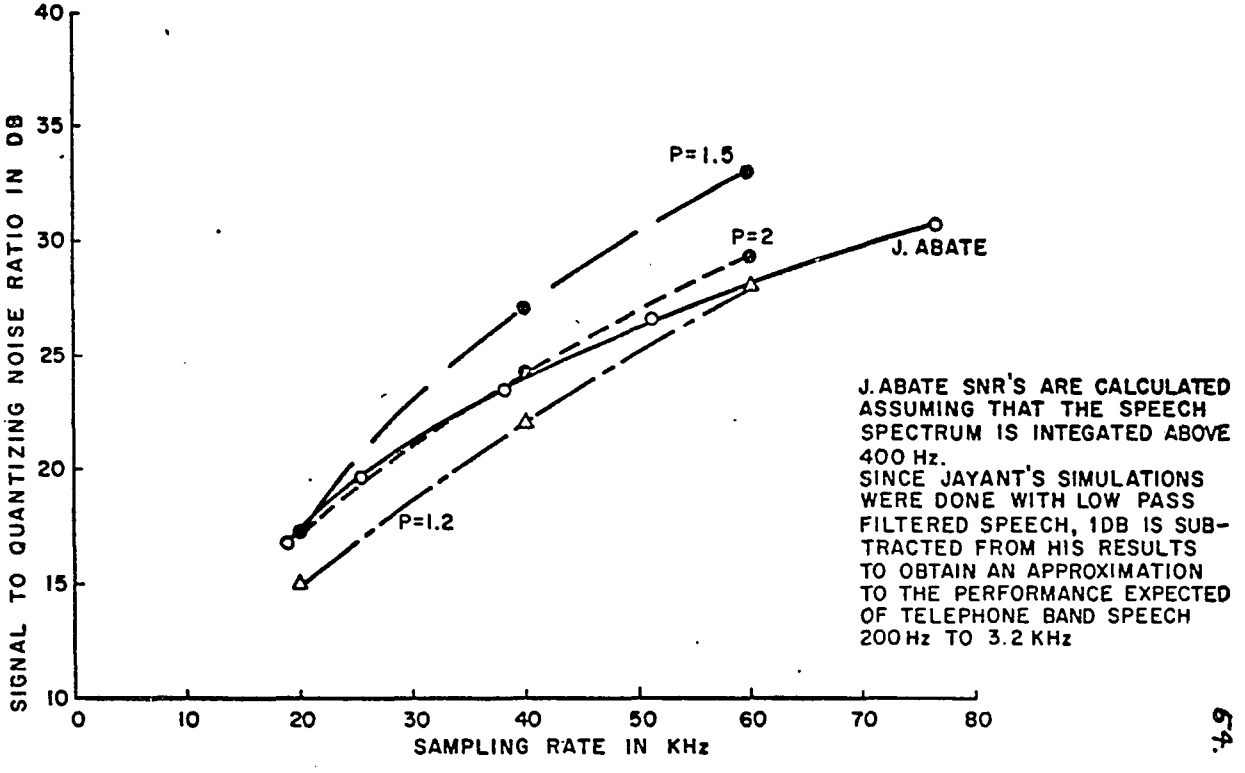

SIGNAL TO QUANTIZING NOISE RATIO IN DB
SAMPLING RATE IN KHz
P=1.5
P=2
P=1.2
J. ABATE
J. ABATE SNR'S ARE CALCULATED ASSUMING THAT THE SPEECH SPECTRUM IS INTEGATED ABOVE 400 Hz.
SINCE JAYANT'S SIMULATIONS WERE DONE WITH LOW PASS FILTERED SPEECH, 1DB IS SUBTRACTED FROM HIS RESULTS TO OBTAIN AN APPROXIMATION TO THE PERFORMANCE EXPECTED OF TELEPHONE BAND SPEECH 200Hz TO 3.2 KHz

FIG. 11

delta modulator is an easily realizable device that does not require precise balancing. The author has also shown that if all corresponding positive steps, Δ_{i+}, are made slightly smaller than the negative steps Δ_{i-}, then the exponentially adaptive delta modulator is asymptotically stable, and idle channel patterns will involve only the two smallest step sizes.

The quality of speech regenerated by N. S. Jayant's instantaneously adaptive delta modulator is quite good at sampling rates above 28 kHz. It is interesting to note that a sub-optimal setting of $\underline{P}$ favoring slope overload noise was preferred by most listeners. [12] This indicates that the slower companding used in the continuously adaptive delta modulators is adequate.

In other instantaneously adaptive delta modulators as well as in the continuously adaptive delta modulators, the integrating networks are more complex than in Jayant's scheme. Generally double integration is employed somewhere above 1 kHz to one sixth of the sampling frequency. At sampling rates of 20 to 40 kHz, the subjective quality of speech regenerated using double integration is somewhat better than that obtained with single integration. No measurable improvement in the signal to noise ratio when processing speech or in

J. A. GREEFKES AND F. DE JAGER'S CONTINOUS DELTA MODULATION SYSTEM

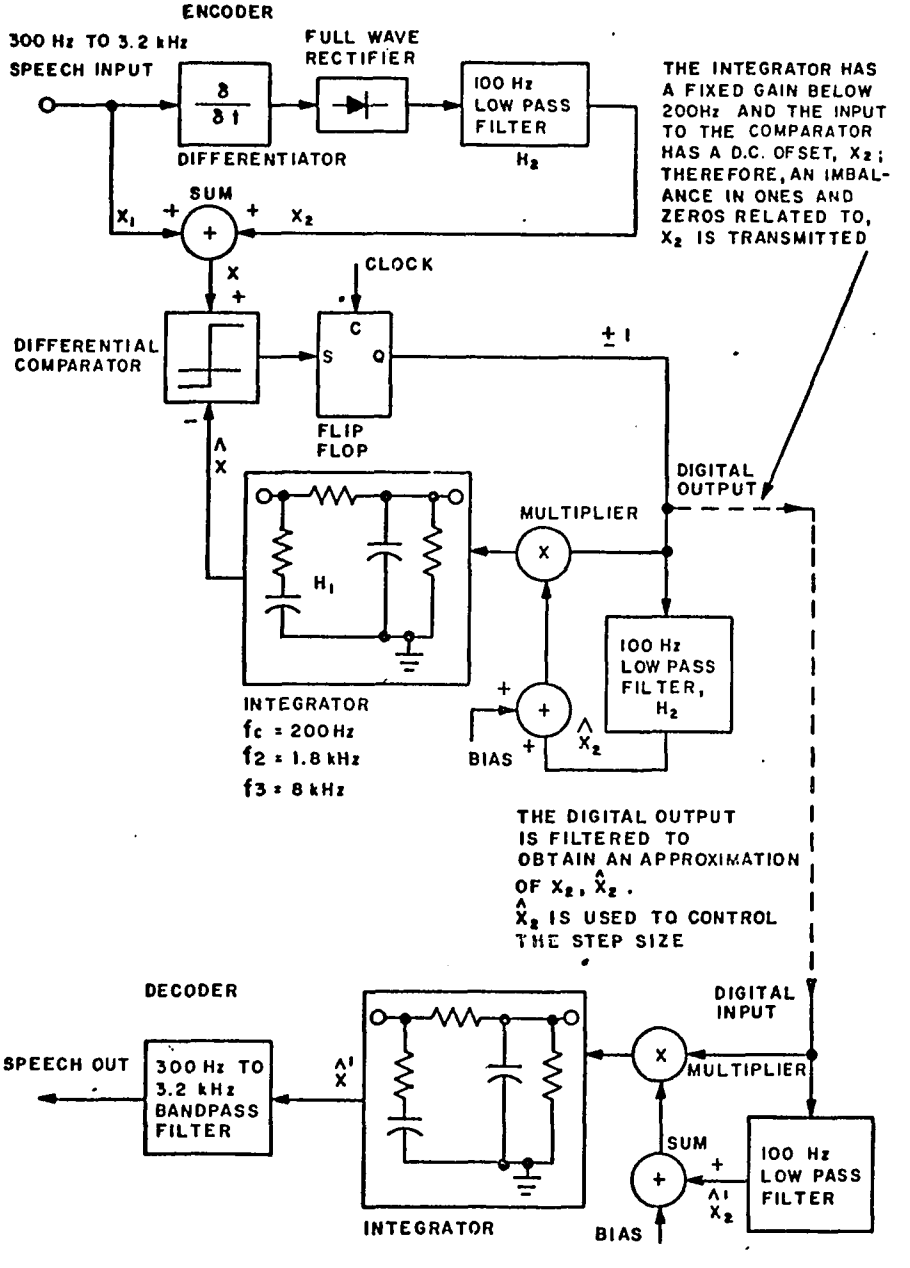


FIG. 12

intelligibility are obtained from double integration, however. The lack of improvement in signal to noise ratios is to be expected, since the design of the double integration networks is based on subjective considerations having no relation to second order prediction.

A slow or continuously adaptive delta modulator was first proposed by J. A. Greefkes and F. DeJager [5]. In this delta modulator the magnitude of the derivative of the input signal is low-pass filtered at 100 Hz and coded by a pulse frequency modulator incorporated in the delta encoder. The signal recovered from the pulse frequency modulation channel is used to control the step size. In order to accommodate two channels in one coder (one channel for the step size companding information and one for speech), the speech signal must be limited to the 300 Hz to 3.2 kHz band. (See Fig. 12).

In Tomozawa and Kaneko's [23] adaptive delta modulator, the output code is low-pass filtered at 100 Hz, rectified, and low-pass filtered again to obtain the step-size, adaptation signal. If at a given time the quantization step is too small, series of consecutive ones and zeros are generated at the output of the encoder. Hence, the bit stream acquires more low frequency energy which will pass through the H_2 filter (see Fig. 13), causing the companding signal and the step size to grow

A. TOMOZAWA AND H. KANEKO'S COMPANDED DELTA MODULATOR

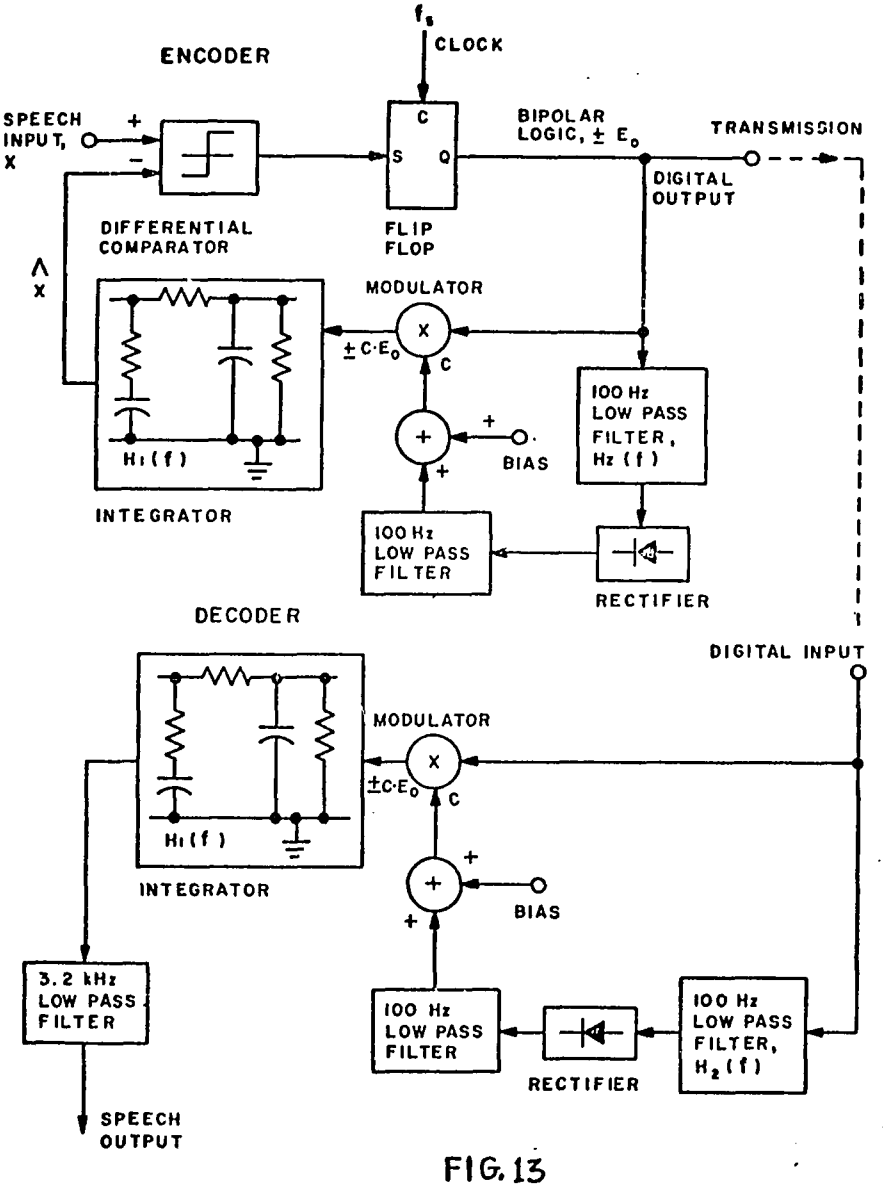


FIG. 13

larger. When the step size is too large the output oscillates between zero and one most of the time, causing energy in the bit stream to shift out of the passband of H_2. This causes a drop in the adaptation voltage and a subsequent decrease in the quantizing step.

All of the adaptive delta modulators mentioned here are capable of regenerating highly intelligible speech at sampling rates as low as 20 kHz, and transmission error rates as high as 1 in 100. If the cut-off frequency on the low-pass filters used in the continuously adaptive delta modulators is reduced, error rates as high as 1 in 10 can be tolerated; however, performance in the absence of errors is partially impaired.

The important conclusions to be drawn from this section are the following:

(1) Adaptation of the quantizer extends the dynamic range of a coder and there by improves performance.

(2) Adaptative quantization also squelches idle channel noise.

(3) Slow adaptation offers more resistance to channel errors.

(4) The signal to noise performance of the exponentially adaptive delta modulator with a one bit memory is nearly the best that can be obtained with a fixed integrator network. Hence, the signal to noise performance curves taken on this delta modulator can be used to compare delta modulation performance to other coding schemes.

PART II

THE NEW RESEARCH

Although some of the theory in Part I was formulated by the author, most of it is a review of the literature. Part II, on the other hand, contains only new research and new results. In the first section, the simulation of a DPCM coder with an adaptive quantizer is described, signal-to-noise measurements are presented, and subjective tests are discussed. In the second section, a hardware realization of the ADPCM coder is shown, and an application to automatic voice answer back systems is pointed out. In the third section, an ADPCM coder is described where an eighth-order predictor is adapted to fit the changing statistics of the input signal. Improvements gained by the application of adaptive, higher-order prediction are discussed in this section.

Section 1 A DPCM System With An Adaptive Quantizer

Adaptive DPCM (ADPCM) coding was investigated by the author for the following reasons:

(1) Differential or predictive coding offers an advantage over direct coding of the input signal as in PCM systems. (See equations 1.1.17 and 1.2.51)

(2) Adaptive Quantization provides a wide dynamic range and thus eliminates any need to compromise on the quantizer design.

(3) The signal to noise ratio of a delta modulator is at best proportional to the cube of the information rate; where as the SNR of DPCM and PCM coders increases exponentially with respect to the bit rate. (For each bit added to the code word, the number of quantization levels is doubled and the SNR can be increased by 6db.) Hence it appeared likely that the performance of ADPCM might be superior to Delta Modulation even at relatively low bit rates.

(4) It was also logical to assume that the problems of limited dynamic range and idle channel oscillations associated with a conventional

DPCM system having a fixed quantizer would also be eliminated in an adaptive system.

The on line computing facility of the Acoustics Research Department at Bell Telephone Laboratories was used to simulate a variety of Adaptive DPCM systems and to test the theories presented earlier. A description of this facility is given in Appendix No. 1.

The First Simulation

The first adaptive DPCM coder simulated on the computer (see Fig. 14) was designed on the premise that the best adaptive quantization could be obtained from some average of the absolute value of previous quantizer outputs. In Jayant's and in Kaneko and Tomozawa's delta modulators a measurement of slope overload (i.e., the occurance of sequences of like symbols in the output code) is used to compand the quantization step because the delta modulator code words contain no amplitude information. In a DPCM coder, however, such information is readily available in the code words and at the output of the quantizer.

In the first round of simulations a uniform, 8 level quantizer was used with a 3 bit code. The sampling rate was set at 8kHz and the prediction

ADAPTIVE DPCM SYSTEM NO. 1

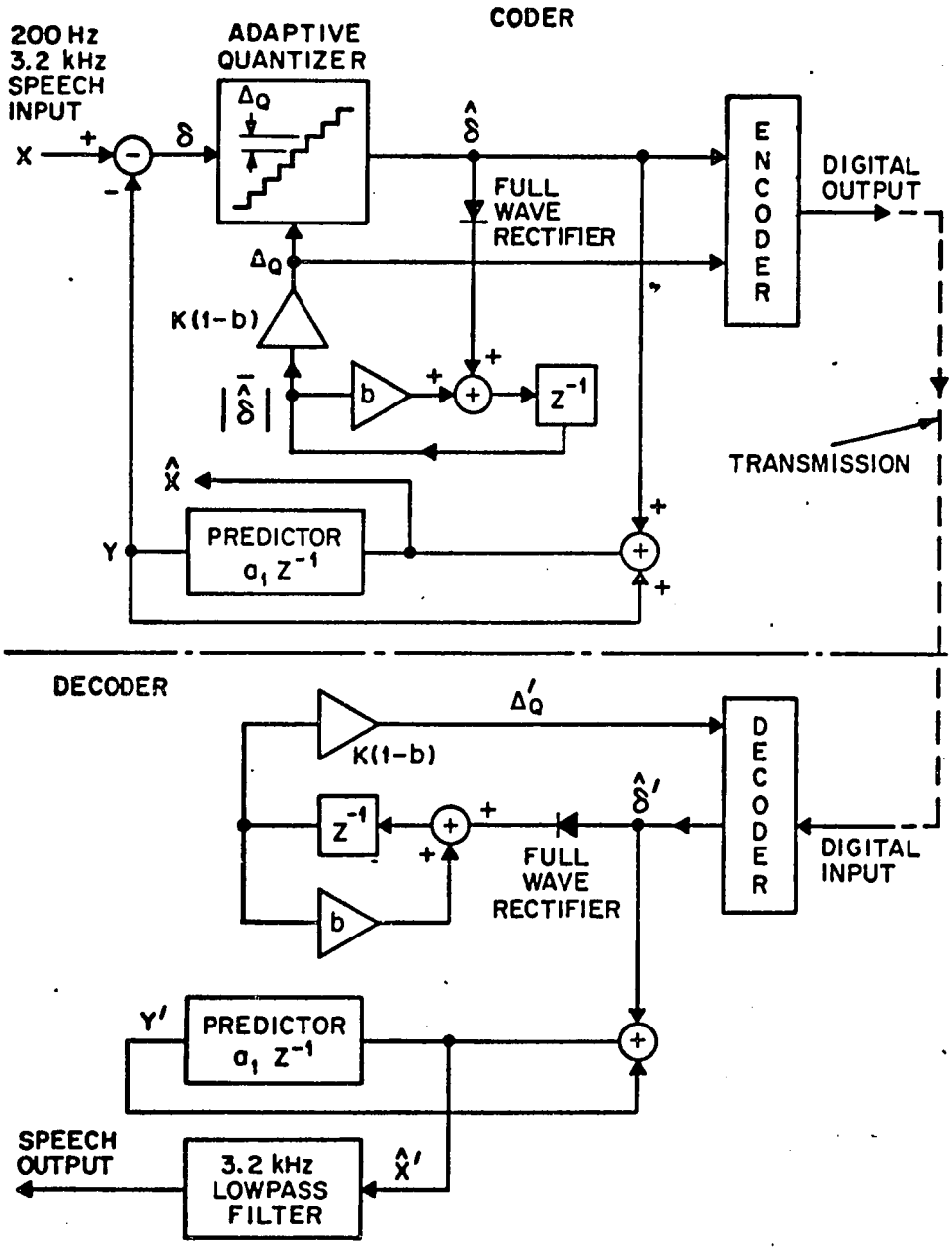


FIG. 14

coefficient, a_1, was set equal to one. The quantized error signal, $\hat{\delta}$, was full-wave rectified and low pass filtered to obtain the quantization step, Δ_Q, to be used to encode the next sample. Two filter parameters (in-band gain, K, and the feedback coefficient, b) were adjusted on line until the best possible results were obtained. The best signal to noise measurement was achieved with an inband gain of K = 0.5 and a cut-off frequency of 220 Hz (i.e., with b = .84).

The Second Simulation

A comparison of time waveforms (input x, output $\hat{x}$ and the quantization step Δ_Q) revealed that severe overload was occurring during times when Δ_Q was increasing. Hence, it was decided that the program should be changed to permit the quantizer to expand rapidly and contract slowly. A circuit which performs the same functions as the program is shown in Fig. 15.

If the magnitude of the quantizer output in Fig. 15 is less than the voltage stored on the capacitor, Δ_Q is reduced by the following factor.

$$\Delta_{Q\ i+1} = \Delta_{Q\ i}\ e^{-T/RC} \qquad (2.1.1)$$

On the other hand if $|\hat{\delta}|$ is greater than the voltage on the capacitor, then

ADAPTIVE DPCM SYSTEM NO. 2

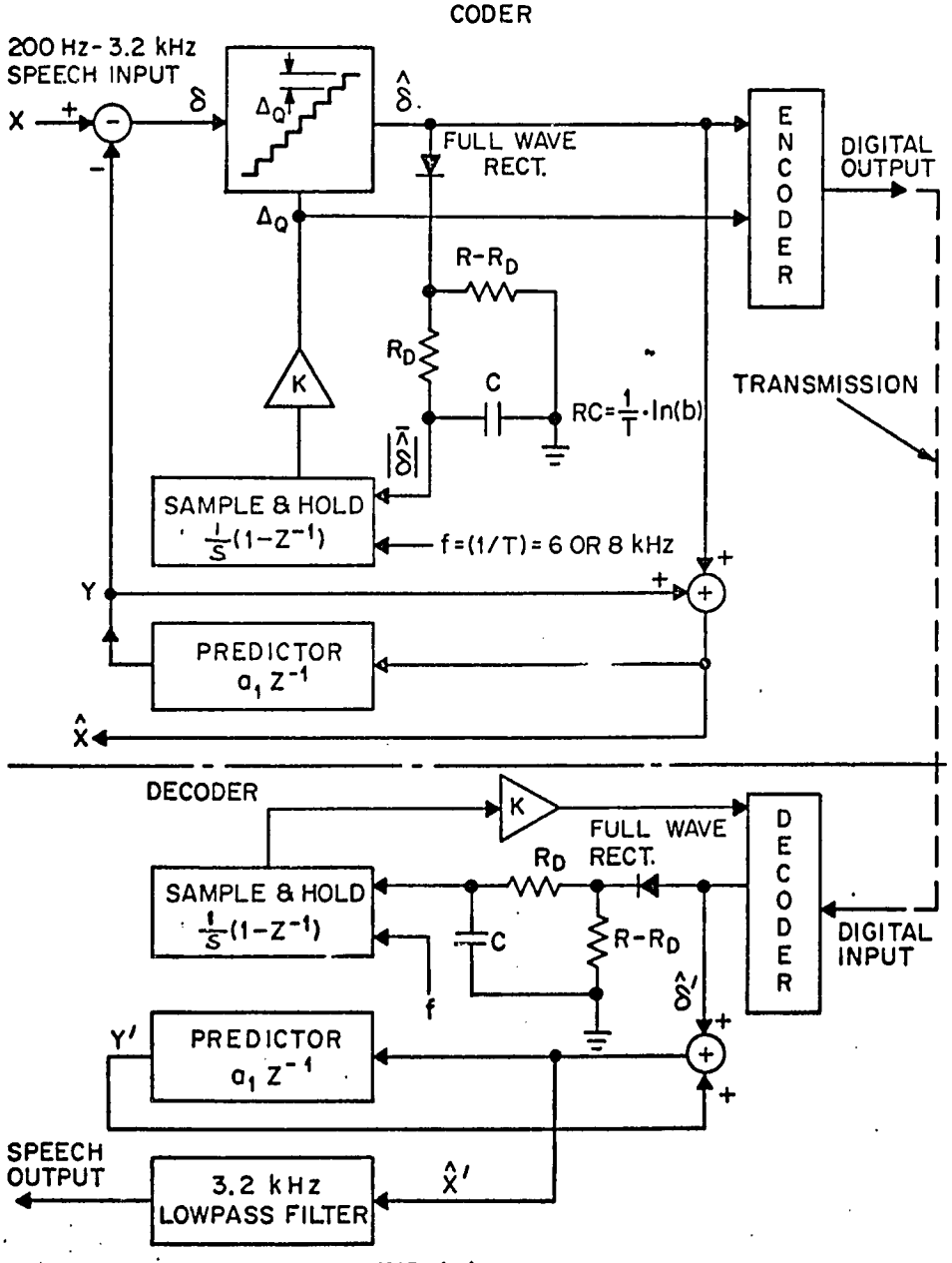


FIG. 15

$$\Delta_{Q_{i+1}} = (K \mid \hat{\delta}_i \mid - \Delta_{Q_i})(1-e^{-T/R_dC}) + \Delta_{Q_i} \qquad (2.1.2)$$

$$\Delta_{Q_{i+1}} = (K \mid Q_i \mid - 1)\ \Delta_{Q_i}\ (1-e^{-T/R_dC}) + \Delta_{Q_i} \qquad (2.1.3)$$

where Q_i is a number such that

$$Q_i \Delta_{Q_i} = \hat{\delta}_i \qquad (2.1.4)$$

Subsequent simulations revealed that the best performance is achieved when $R_d \simeq 0$; therefore, when $\mid \hat{\delta}_i \mid$ is greater than the voltage on the capacitor (i.e., when $Q_i > 2.0$),

$$\Delta_{Q_{i+1}} = K \mid Q_i \mid \ \Delta_{Q_i} \ . \qquad (2.1.5)$$

A signal to noise ratio of 15db was achieved with Rd = 0, K = 0.5 and the cut-off frequency of the remaining RC filter set at 170 Hz (b=.875).

The signal to noise ratios given here refer to the ratio of average power in the input signal to average power in the error signal after it is low-pass filtered at 3.2 kHz.

$$SNR = \frac{E(x^2)}{E(e_{LP}^2)} = \frac{E(x^2)}{E((x-\hat{x})_{LP}^2)} = \frac{E(x^2)}{E((\delta-\hat{\delta})_{LP}^2)} \qquad (2.1.6)$$

The low-pass filter used was a nonrecursive, transversal filter with a 128 sample impulse response. The filter was designed and programmed by L. R. Rabiner using a frequency sampling technique.[20] The filter has unity gain from D.C. to 3.2 kHz and an attenuation of 40 db or more from 3.3 kHz to 4 kHz. As is the case with sample data filters, the amplitude response repeats itself in the frequency domain in the same manner as the spectra of sampled signals. Although power in the aliasing bands (4.8 kHz to 11.2 kHz, 12.8 kHz to 19.2 kHz, etc.) is passed by the filter, noise power in these bands is compared with power in the sampled input speech spectrum.

The Third Simulation (Instantaneous Companding)

After considering some schemes for realizing the adaptive DPCM system in hardware, it was decided that it would be easier to adapt the quantizer from logic controlled from the encoder as in Fig. 16. At first a logic exactly equivalent to the above was simulated and identical results were achieved. When a code word

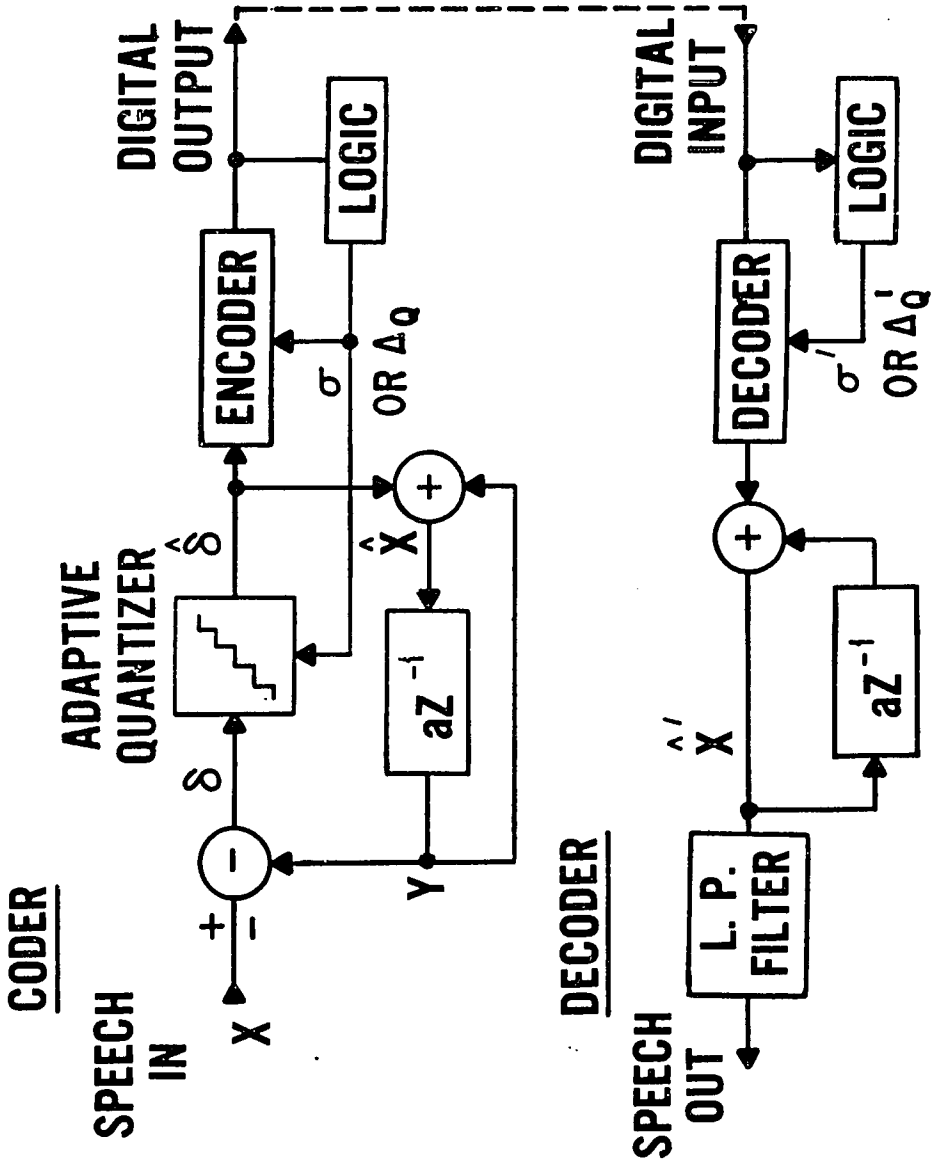

CODER
SPEECH IN
X
+
−
δ
ADAPTIVE QUANTIZER
δ̂
ENCODER
DIGITAL OUTPUT
LOGIC
σ
OR Δ_Q
X̂
aZ^-1
Y
DECODER
SPEECH OUT
L. P. FILTER
X̂'
DECODER
DIGITAL INPUT
LOGIC
σ'
OR Δ_Q'
aZ^-1

FIG.16

indicated that the largest possible quantizer output was used, ($K \mid Q_i \mid = 1.75$) the quantizer size was multiplied by 1.75 as indicated by Eq. (2.1.5). For $\mid \hat{\delta} \mid$ of the next smallest size, the quantizer was enlarged by 25 percent ($K \mid Q_i \mid = 1.25$), and for codes indicating smaller values of $\mid \hat{\delta} \mid$ the quantizer was reduced to .875 of its previous size. An attempt was made to find better multipliers, and it was found that those given above yielded the highest signal to noise ratios.

In order to make the description of the coder perfectly clear, the uniform, 8 level quantizer is shown in Fig. 17 with the code words assigned to each output level, and the companding rule is stated in Table 1. The coder operates in the following manner. The difference, δ, between a band limited input signal, X, and a predicted value, Y, is quantized and coded. Then the quantized difference is summed with Y to obtain a corrected estimate of the input signal, $\hat{X}$. The code depicting the quantizer level used to estimate δ is transmitted to the decoder. It is also processed by a logic (see Table 1) which expands or contracts the quantizer before the next input sample is encoded. The quantizer size reference, Δ_Q' at the decoder is derived from a logic identical to that which compands

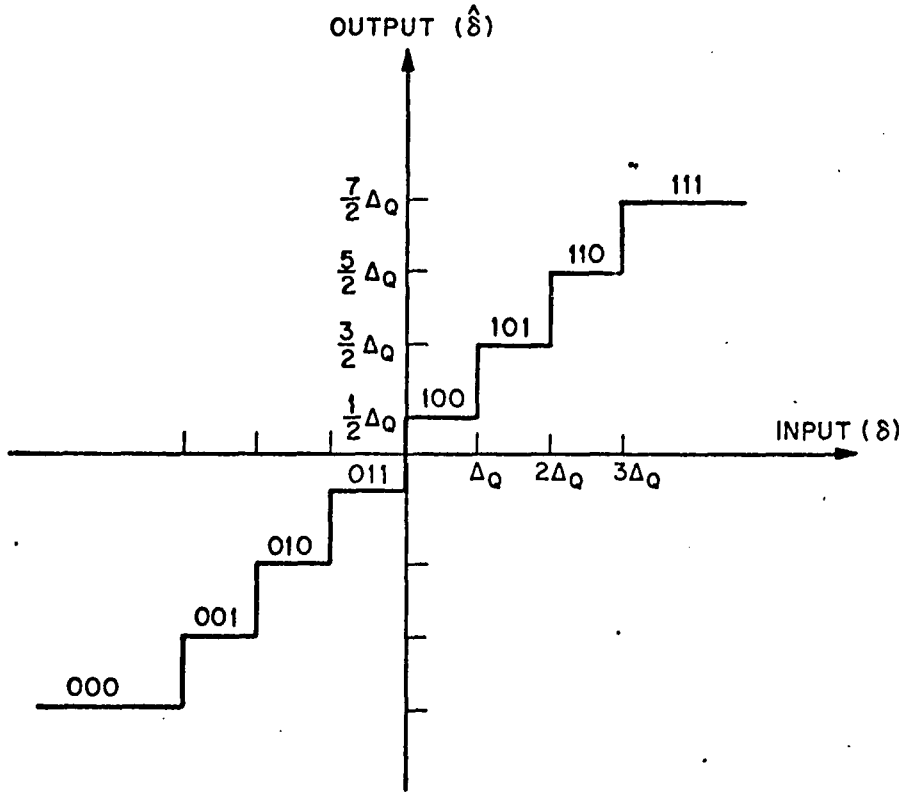

OUTPUT ($\hat{\delta}$)
$\frac{7}{2}\Delta_Q$
$\frac{5}{2}\Delta_Q$
$\frac{3}{2}\Delta_Q$
$\frac{1}{2}\Delta_Q$
111
110
101
100
011
010
001
000
INPUT (δ)
Δ_Q
$2\Delta_Q$
$3\Delta_Q$

FIG.17

the quantizer in the encoder; hence Δ_Q' is identical Δ_Q and $\hat{x}'$ is identical to $\hat{x}$ in the absence of transmission errors.

Companding Logic for an ADPCM coder with a Fixed Predictor ($a_1 = .984$) and an 8 level, uniform quantizer. SNR = 15db (Sampling Rate 8 kHz, Bandwidth: 200 Hz to 3.2 kHz)

Code Word	Quantizer Multiplier
111 or 000	1.75
110 or 011	1.25
101, 100, 011 or 010	.875

TABLE NO. 1

A DPCM coder with a 16 level, uniform, adaptive quantizer was simulated next and a signal to noise ratio of 20.5 db was obtained with the companding logic shown in Table 2. The digital codes corresponding to the sixteen levels were 0000, 0001, 0010, 0011, 0100, 1000, 1001, 1010, 1100, 1101 and 1111. The code words

are listed in order; the first depicts the most negative quantization level and the last depicts the most positive.

Companding Logic for an ADPCM coder with a fixed predictor (a_1 = .984) and a 16 level, Uniform Quantizer (SNR = 20.5 db Sampling Rate = 8 kHz Bandwidth: 200 Hz to 3.2 kHz.)

Code Word	Quantizer Multiplier
1111 or 0000	3.00
1110 or 0001	1.75
1101, 1100, 0011 or 0010	1.25
One of the other 8 codes	.875

TABLE No. 2

At this point several measurements were taken to see if the performance of the coder might not be improved. Among these there were measurements of the probability of occurance of the various code words. From this data a probability histogram of the prediction error, P(δ), (See Fig. 18) was drawn where δ is measured in companded quantizer step

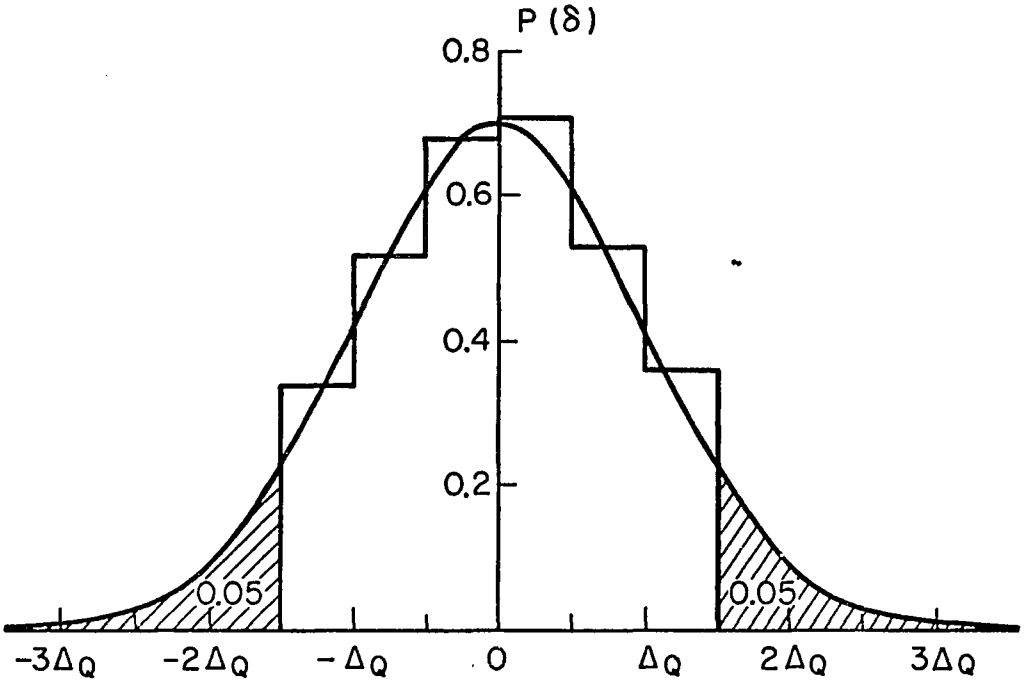


FIG. 18

units. A study of the histogram revealed that the probability was distributed almost exactly as in a Gaussian distribution with a zero mean and a standard deviation of one quantizer step, Δ_Q.

It was decided that another simulation should be done with an optimum Gaussian quantizer. The input decision levels and the output quantization levels of optimum quantizers like those designed by J. Max and others [13,18] are measured in standard deviations of the signal for which they are optimum. Therefore the quantizer size will henceforth be referred to as the standard deviation, σ.

An 8 level quantizer, optimum for Gaussian distributed signals was incorporated in the coder and a signal to noise ratio of 16 db was recorded with the multipliers given in Table No. 3.

Companding Logic for an Adaptive DPCM Coder with a Fixed Predictor ($a_1 = .984$) and an 8 level, Optimum Gaussian Quantizer (SNR = 16db Sampling Rate = 8 kHz Bandwidth: 200 Hz to 3.2 kHz)

Code Word	Quantizer Multiplier
111 or 000	2.00
110 or 001	1.25
101, 100, 011 or 100	.875

TABLE NO. 3

With a 16 level optimum Gaussian Quantizer the following results were recorded.

Companding Logic for an Adaptive DPCM Coder with a Fixed Predictor (a_1 = .984) and a 16 level, Optimum Gaussian Quantizer (SNR = 22 db Sampling Rate = 8 kHz Bandwidth: 200 Hz to 3.2 kHz)

Code Word	
1111 or 0000	3.75
1110 or 0001	1.75
1101, 1100, 0011 or 0010	1.25
One of the other 8 code words	.875

TABLE NO. 4

The code words were assigned in the same sequences used with the uniform quantizers.

The signal to noise ratios obtained with the ADPCM coders are approximately equal to the best that can be expected from a DPCM coder with an optimum fixed quantizer. However, the adaptive coders have the advantage of an extended dynamic range and quiet idle channel performance. The normalized autocorrelation of the speech signal at a delay of one sampling interval was 0.79; hence, by Eq. 1.1.18 the signal to noise improvement is 3.8 db. The error signal is diminished by 0.6 db when it is limited to the 200 Hz to 3.2 kHz band. The probability distribution of the error signal, δ, is quite similar to that of the speech signal; therefore, the signal tc noise performance of a fixed quantizer should be about the same as that of the optimum Gamma, Max quantizer in Fig. 6. A formula for the optimum performance expected of a DPCM coder with a fixed quantizer is given below.

$$\text{SNR} = 10 \log_{10} E\left(\frac{e^2}{e_{LP}^2}\right) + 10 \log_{10} E\left(\frac{\delta^2}{e^2}\right)$$

$$+ 10 \log_{10} \frac{1}{2(1-\rho_1)} \qquad (2.1.7)$$

The optimum SNR is 16 db when using an 8 level quantizer and 21.9 db when using a 16 level quantizer (See Fig. 6 and Eq. 2.1.7). Two more simulations were run with the sampling rate reduced to 6 kHz and the bandwidth reduced to 2.8 kHz. The results of these simulations are given in Tables 5 and 6.

Companding Logic for an Adaptive DPCM Coder with a Fixed Predictor (a_1 = .988) and an 8 level, Optimum Gaussian Quantizer (SNR = 14.8 db Sampling Rate = 6 kHz Bandwidth: 200 Hz to 2.8 kHz)

Code Word	Quantizer Multiplier
111 or 000	2.00
110 or 001	1.25
101, 100, 011 or 010	0.80

TABLE NO. 5

Companding Logic for an Adaptive DPCM Coder with a Fixed Predictor (a_1 = .988) and a 16 level, Optimum Gaussian Quantizer (SNR = 20.7 db Sampling Rate = 6 kHz Bandwidth: 200 Hz to 2.8 kHz)

Code Word	Quantizer Multiplier
1111 or 0000	3.75
1110 or 0001	1.75
1101, 1100, 0011 or 0010	1.25
One of the other 8 codes	0.8

TABLE NO. 6

Although sampling so near to the Nyquist rate increases the burden placed on the desampling filter, a relatively small commercially available 7th order Tschebychev filter with a 2 db in-band ripple does the job adequately.

A Comparison of ADPCM, LOG PCM and DELTA Modulation (Objective Measures)

In Fig. 19 the signal to noise performance of Jayant's adaptive delta modulator is compared with the ADPCM system operating at a 6 kHz sampling rate.

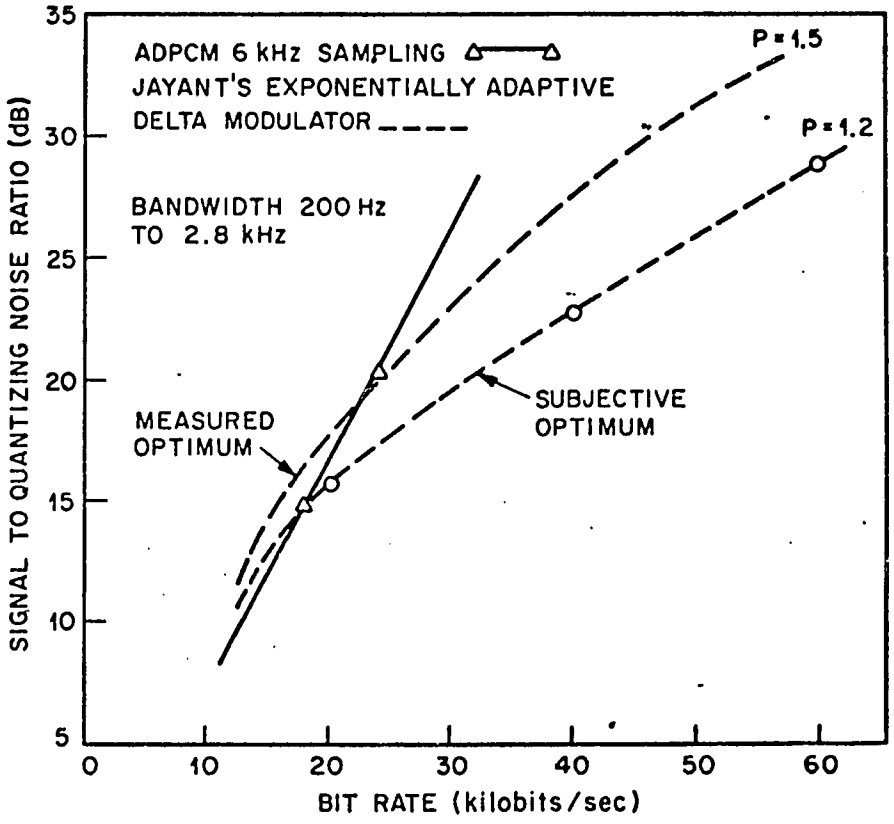

ADPCM 6 kHz SAMPLING
JAYANT'S EXPONENTIALLY ADAPTIVE
DELTA MODULATOR
BANDWIDTH 200 Hz
TO 2.8 kHz
P = 1.5
P = 1.2
MEASURED
OPTIMUM
SUBJECTIVE
OPTIMUM
SIGNAL TO QUANTIZING NOISE RATIO (dB)
BIT RATE (kilobits/sec)
35
30
25
20
15
10
5
0
10
20
30
40
50
60

FIG. 17

Here the oversampling ratio[1] of the adaptive delta modulator is approximately equal to the word size in the ADPCM system. In Fig. 20 the ADPCM systems operating at 8 kHz sampling are compared with Jayant's delta modulator and with log PCM (μ = 100, 8 kHz sampling).

The eight decibel advantage held by the ADPCM coder over PCM is due to the signal to noise improvement factor associated with a DPCM system and to the use of optimum quantization. The μ law quantization used in PCM systems sacrifices about 4 db in SNR to achieve a wide dynamic range. No such compromise need be made in the ADPCM system because the adaptive feature provide for a wide dynamic range.

The comparisons made in Figs. 19 and 20 are cross comparison between published results, and the results of the simulations. Recent simulations run by A. E. Rosenberg indicate that ADPCM coders perform as well or better than the adaptive delta modulator at all bit rates, when both are processing the same signal. It appears that a measured difference less than 2 db exists between ADPCM and adaptive delta modulation at 24 kilobits per second or less. When an optimum

---

(1) The oversampling ratio is defined as, $B \equiv (f_s/2f_B)$ where f_s is sampling rate of the delta modulator and f_B is the bandwidth of the signal.

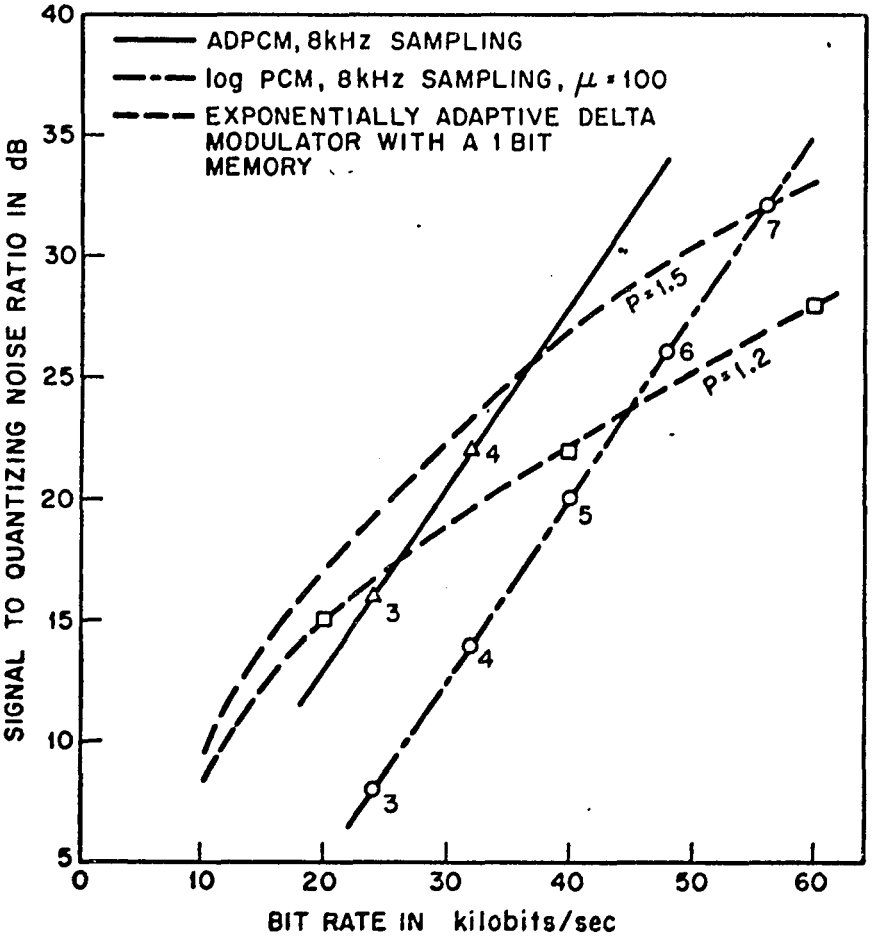

ADPCM, 8kHz SAMPLING
log PCM, 8kHz SAMPLING, μ = 100
EXPONENTIALLY ADAPTIVE DELTA MODULATOR WITH A 1 BIT MEMORY
P=1.5
P=1.2
3
4
7
6
5
3
4
SIGNAL TO QUANTIZING NOISE RATIO IN dB
BIT RATE IN kilobits/sec
40
35
30
25
20
15
10
5
0
10
20
30
40
50
60

FIG. 20

quantizer of 16 or more levels is used, the performance of the ADPCM systems is superior. The extrapolation of the signal to noise curves for ADPCM is justified because published results indicate that the signal to noise ratio of an optimum quantizer will rise by 6 db when the number of quantization levels is doubled and a bit is added to the code word (see Figs. 5 and 6).

A Theory For Optimum Quantizer Multipliers

Thus far in Section 1 of Part II an experiment has been devised and the results have been recorded. Basically, what was finally done is that an ADPCM coder program was written where for each of the possible output codes the quantizer was adapted by a fixed multiplier. The multipliers were adjusted on line by the programmer until he felt that he could no longer improve the performance.

Although a precise formula for optimum quantizer multipliers has not been found, some theory has been generated which explains why the optimum multipliers dictate that the quantizer should be expanded rapidly and contracted slowly. Jayant simulated an adaptive uniform quantizer which utilizes a one word memory. After an exhaustive search he found that rules similar to those stated above were optimum while processing a 10,000 sample Gauss-Markov sequences, δ_i.

$$\delta_i = \rho_1 \, \delta_{i-1} + (\sqrt{1-\rho_1^2}) \quad N_i \qquad (1.2.8)$$

where ρ_1 is the correlation between noise samples and N_i is a white Gaussian sequence. Jayant found that fast increases and slow decreases in quantizer size constitute the best adaptation and that the signal to noise ratio is not increasedas compared with an optimum, fixed, uniform quantizer unless the correlation between adjacent steps is quite large.

Jayant then developed a theory which partially explains our results. The best estimate of the standard deviation of the next sample of a signal given the present sample is

$$E_1 \, (\sigma_{i+1}^2) = \delta_i^2 \quad . \qquad (2.1.9)$$

The best estimate based on the present and previous samples is

$$E_2 \, (\sigma_{i+1}^2) = \frac{\delta_i^2 + \delta_{i-1}^2}{2} \quad , \qquad (2.1.10)$$

and the difference between the two estimates is

$$\Delta E(\sigma_{i+1}^2) = E_2(\sigma_{i+1}^2) - E_1(\sigma_{i+1}^2)$$

$$= \frac{\delta_{i-1}^2 - \delta_i^2}{2} \qquad (2.1.11)$$

If the present quantizer size, σ_i^2, is set equal to $E_2(\sigma_i^2)$, then

$$\sigma_i^2 = \delta_{i-1}^2 + \Delta E(\sigma_i^2) \qquad (2.1.12)$$

If δ_{i-1}^2 is eliminated by combining (2.1.10) and (2.1.12), the following relationship is obtained

$$\frac{E_2(\sigma_{i+1}^2)}{\sigma_i^2} \equiv \frac{\sigma_{i+1}^2}{\sigma_i^2} = \frac{1}{2}\left(\frac{\delta_i^2}{\sigma_i^2} + 1 - \frac{\Delta E(\sigma_{i+1}^2)}{\sigma_i^2}\right) \qquad (2.1.13)$$

Of the quantities required to solve Eq. (2.1.13) only δ_i^2 is known correct to within the quantization error.

$$\delta_i = Q_i\sigma_i + e_i \qquad (2.1.14)$$

where Q_i is a number such that $\hat{\delta}_i = Q_i\sigma_i$. Hence, Eq. (2.1.13) may be written as follows:

$$\frac{\sigma_{i+1}^2}{\sigma_i^2} = \frac{1}{2}\left(Q_i^2 + 1 + \frac{e_i^2}{\sigma_i^2} + \frac{2Q_ie_i}{\sigma_i} - \frac{\Delta E(\sigma_i^2)}{\sigma_i^2}\right) \qquad (2.1.15)$$

The last term in Eq. (2.1.15) is an unknown random variable with a mean value of zero; therefore, the best adaptation multiplier is given by the following:

$$\frac{\sigma_{i+1}}{\sigma_i} = \sqrt{\frac{Q_i^2 + 1 + (e_i^2/\sigma_i^2) + (2Q_i e_i/\sigma_i)}{2}} \qquad (2.1.16)$$

Given an eight level uniform quantizer as in the first simulations, Max found that the optimum values for the output levels ($\hat{\delta}_i = Q_i \sigma_i$) were $2.03\sigma_i$, $1.45\sigma_i$, $0.87\sigma_i$ and $.29\sigma_i$. With Q_i = 2.03, 1.45, etc. and the assumption that the histogram shown in Fig. 18 is the optimum distribution of output level usage, the following multipliers are obtained from Eq. (2.1.16).

Code Nδ	Qi	σ_{i+1}/σ_i By Eq. (2.1.16)	σ_{i+1}/σ_i Experimental Optimum
111 or 000	2.03	1.61	1.75
110 or 011	1.45	1.21	1.25
101 or 010	0.87	0.93	.875 (8 kHz sampling) or .80 (6 kHz sampling)
100 or 011	0.29	0.74	.875 (8 kHz sampling) or .80 (6 kHz sampling)

TABLE NO. 7

Considering the simplicity of the derivation, the optimum multipliers given by Eq. (2.1.16) agree rather well with the experimental values obtained in the ADPCM simulations with speech inputs. In his simulations with Gauss-Markov sequences Jayant found that the best multipliers were nearly equal to the experimental values given for processing speech at 8 kHz sampling. In the simulation of the ADPCM system with a speech input and in Jayant's simulations with Gauss-Markov sequences the best performance was achieved when the two attenuating multipliers were equal. When the sampling rate was changed from 8 kHz to 6 kHz, the lower multipliers change when processing speech. On the other hand, no such change occurred in Jayant's data when ρ_1 was changed. The discrepencies in the lower multipliers and a reduction of 1 to 2 db in the signal to noise ratio obtained at the quantizer as compared to Jayant's results is probably related to the fact that the error signal, δ, is not Gaussian. The discrepencies may also be related to the quantity $\Delta E(\sigma_i^2)/\sigma_i^2$. The estimates of the effects of the noise terms, $2Q_i e_i/\sigma_i$ and e_i^2/σ_i^2 were obtained from the histogram shown in Fig. 18. The effects of these terms are too small for errors in them to be significant.

Subjective Tests

If the measurements shown in Fig. 19 are used to judge the ADPCM system, it might be concluded that at 18 kilobits/sec. the adaptive delta modulator with P = 1.5 is superior to the 3 bit ADPCM system. But the adaptive delta modulator with P = 1.2 was preferred by listeners over the optimum delta modulator (P = 1.5). Likewise it might be concluded from Fig. 20 that 5 bit log PCM (μ = 100) is superior to the 3 bit ADPCM system, and that 6 bit log PCM is superior to the 4 bit ADPCM system. It was decided that a subjective comparison should be made between the ADPCM systems and μ = 100, log PCM. At a sampling rate of 8 kHz the following speech sample was processed by the 3 bit and 4 bit ADPCM systems, and by 4 bit, 5 bit, 6 bit and 7 bit (μ = 100) log PCM systems.

"The circuit operates on the same principle as
N. S. Jayant's simulation."

Several copies of each of the processed speech samples were made. A sentence processed by the 3 bit ADPCM system was spliced ahead of a sentence processed by each of the log PCM systems with a half second of silence between them. Then a sentence processed by the 3 bit ADPCM system was spliced behind a sentence processed

by each of the log PCM systems. The same splicing procedure was carried out with sentences processed by the 4 bit ADPCM system to obtain a total of sixteen paired comparisons. Each pair was assigned a number and the numbers were written down in the order in which they appeared in the random numbers table published in the CRC Mathematical Tables (12th edition). Finally the pairs were spliced together in the random order obtained from the table, with three seconds of silent leader placed between them.

Each of 22 subjects was asked to listen to one of the sentences processed by an ADPCM system, and to attempt to repeat back what was said. Almost all of the listeners were able to repeat back 90% or more of the test sentence on the first try, and to get the entire sentence correct on the second. Some difficulty was encountered with the unusual name "N. S. Jayant"; however, all of the listeners mouthed some sounds which were phonetically close to "N. S. Jayant". Hence, it was concluded that the intelligibility of the ADPCM systems approached that of natural speech.[8] After this brief intelligibility test, each listener was permitted to listen to part of the tape so as to become accustomed to it. Then he was asked to listen to the entire tape and to indicate which member of each

pair he preferred (i.e., which one, in his opinion, was of better quality). The decision was forced upon him; he was required to listen repeatedly to a pair until he made a decision. If, however, a subject required a second try or if he thought the decision was difficult he was asked to indicate the difficulty by checking a box marked "same".

The listeners were a mixture of Bell Laboratories employees, N.C.E. faculty and students, the author's wife and two of his neighbors. Eighteen of the listeners were men and four were women. The test lasted about 5 minutes; hence, fatigue and discomfort due to wearing earphones, etc. was minimized.

The results were scored as follows: If a comparison resulted in a clear preferrence, the preferred system was given one point. If the box marked "same" was checked, the preferred system was given two thirds of a point and the other was given one third of a point. This scoring system allots one point to a comparison; therefore, a confidence measurement on the outcome can be obtained by comparison with a simple coin flipping experiment. Each pair of systems was compared twice (once with the order of presentation reversed) by each of the twenty-two listeners; hence, 44 comparisons were

made. If an honest coin is flipped 44 times, the mean number of heads will be 22 and the standard deviation will be 3.32.

$$m = PN \text{ and} \qquad (2.1.17)$$

$$\sigma = \sqrt{NPq} \quad , \qquad (2.1.18)$$

where P is the probability that a head will occur, q is the probability of a tail, and N is the number of flips of the coin. For an honest coin, $P = q = \frac{1}{2}$. For large numbers of events the Bernoulli distribution with P = 1/2 (Eq. 1.2.17 and 1.2.18) can be very accurately approximated by a normal distribution with the same mean and standard deviation. Hence, the area under a normal distribution with a mean of 22 and a standard deviation of 3.32 taken from $-\infty$ to the highest test score gives a confidence measurement of the outcome. For example, the 3 bit ADPCM system got a score of 31 2/3 (2.91 standard deviations above the mean) when compared with the 5 bit log PCM system. The area under the normal distribution is

$$\int_{-\infty}^{2.91} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = .9982 \quad . \qquad (2.1.19)$$

This indicates a high degree of confidence in the superiority of the ADPCM system. On the other hand, the 6 bit log PCM system scored 22 2/3 (only 0.2 standard deviations above the mean for coin flipping) when compared with the 3 bit ADPCM system. Hence, it is concluded that these systems regenerate speech of nearly equal quality because the confidence is only 0.58. The overall scores and confidence ratings are given in Table 8.

It is argued by many people in the behavioral sciences that stimuli can be ranked according to preference only if the subjects agree to some extent on what they perceive and on the criteria for making judgments.

A method for representing data in three dimensions has been programmed and reported on by J. D. Carroll.[3] The paired comparisons given by each subject were processed by Carroll's multidimensional scaling program. In Fig. 21 planes normal to each subject vector are drawn through the stimulus points. The points where these planes intersect the vectors approximately represent the subjects' preferences. The vectors point through the origin in the direction of increasing preference. The least preferred stimuli project on to the negative tails of the vectors.

HIGH SCORE	LOW SCORE	STANDARD DEVIATIONS FROM THE MEAN FOR 44 COIN FLIPS, σ_0	CONFIDENCE, $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sigma_0} e^{-x^2/2}\, dx$
3 bit ADPCM-38	4 bit log PCM-6	4.82	1.0000
3 bit ADPCM-31 $\frac{2}{3}$	5 bit log PCM-12 $\frac{1}{3}$	2.91	0.9982
6 bit log PCM-22 $\frac{2}{3}$	3 bit ADPCM-21 $\frac{1}{3}$	0.20	0.5793
7 bit log PCM-34 $\frac{1}{3}$	3 bit ADPCM-9 $\frac{2}{3}$	3.22	0.9994
4 bit ADPCM-38 $\frac{1}{3}$	4 bit log PCM-5 $\frac{2}{3}$	4.92	1.0000
4 bit ADPCM-38 $\frac{1}{3}$	5 bit log PCM-5 $\frac{2}{3}$	4.92	1.0000
4 bit ADPCM-32 $\frac{1}{3}$	6 bit log PCM-11 $\frac{2}{3}$	3.11	0.9991
7 bit log PCM-23 $\frac{2}{3}$	4 bit ADPCM-20 $\frac{1}{3}$	0.50	0.6915

TABLE 8

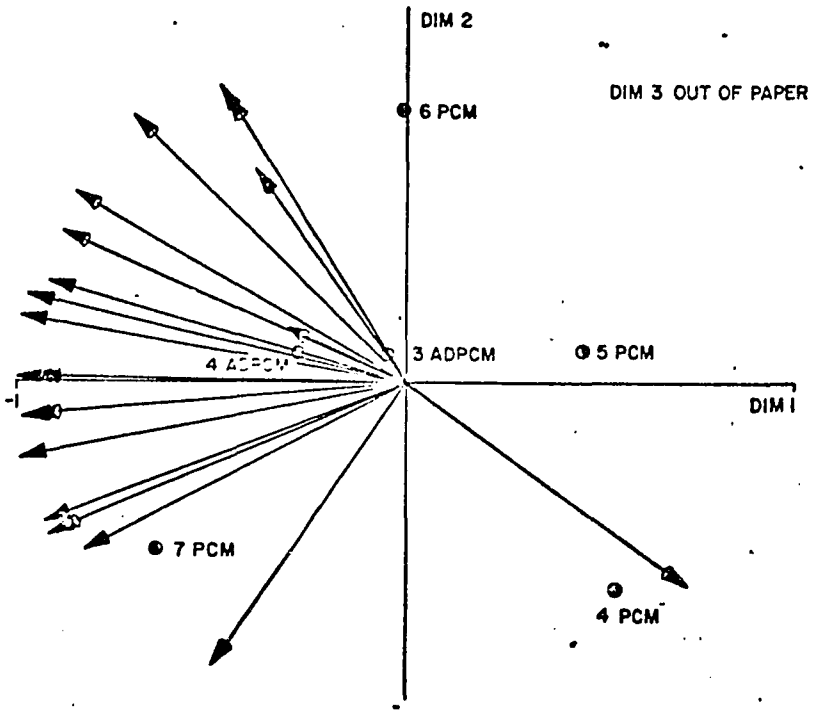


FIG. 21

Seventy-seven percent of the subject variance projects onto the first dimension. This indicates that one criterion is of primary importance. Twelve percent of the variance projects on the second dimension and seven percent on a third. Overall preference ratings for the systems may be obtained by projecting the stimulus points onto the first dimension, where the most negative values are the most preferred. These ratings compare rather well with the overall scores given in Table 8. A small discrepancy exists because the 3 bit ADPCM system is thus ranked above 6 bit log PCM. This is not serious because the difference in the scores in Table 8 and the preference difference plotted on dimension one are both statistically insignificant.

The important results of the subjective tests are that the 3 bit ADPCM system regenerates speech with a quality equal to 6 bit log PCM, and that the quality of speech regenerated by the 4 bit ADPCM system is somewhere between that of 6 bit and 7 bit log PCM.

There is a discrepancy of 4 db to 10 db between the measured signal to noise ratios shown in Fig. 20 and the results of the subjective tests. If more ADPCM quantizing noise can be tolerated than PCM noise, then the two kinds of noise must be qualitatively different. An effort was made to determine this difference.

H. Levitt,[15] N. S. Jayant and A. E. Rosenberg[12] have all found that slope overload in a delta modulator is far more tolerable than granular noise like that generated by a PCM coder. When Jayant's delta-modulator is operating at the subjective optimum (P=1/Q=1.2), most of the noise is slope overload noise (i.e., most of the time the input sample, X_i, does not lie between, the predicted value, Y_i, and the output sample, $\hat{X}_i$). R. A. McDonald[16] found that the spectrum of the quantizing noise in a DPCM coder with a fixed quantizer is relatively white during normal operation; however, during heavy loading when $|\delta_i|$ is often greater than $\hat{\delta}_N$, he found that the noise spectrum resembled that of the input signal. Hence, some masking effects occur and the noise becomes more acceptable during overload.

In the ADPCM system with an eight level quantizer, overload occurs about six percent of the time and accounts for slightly less than half of the noise power (see Fig. 18). Hence, an overall change in the noise spectrum of no more than 3 db can be attributed to overload. In Fig. 22 the long-time power spectra are shown for the input speech signal, the ADPCM quantizing noise and for the noise generated by a $\mu = 100$, 5 bit log PCM coder.

$$\Phi_{xx}(\omega) = \sum_{k=0}^{128} (1 - \frac{k}{128})\rho_{xx}(kT) \cdot \cos(k\omega T) \qquad (2.1.19)$$

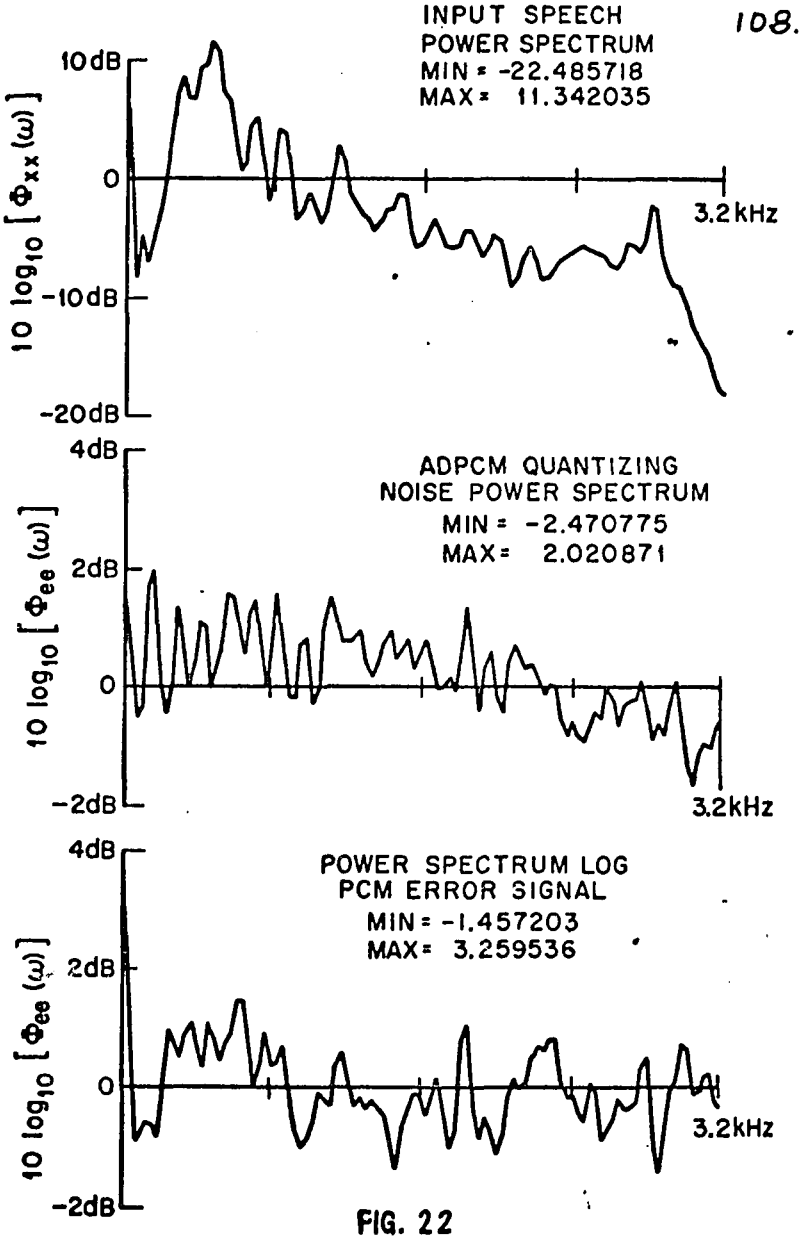

INPUT SPEECH
POWER SPECTRUM
MIN = -22.485718
MAX = 11.342035
10dB
0
-10dB
-20dB
10 log10 [Φxx (ω)]
3.2kHz
ADPCM QUANTIZING
NOISE POWER SPECTRUM
MIN = -2.470775
MAX = 2.020871
4dB
2dB
0
-2dB
10 log10 [Φee (ω)]
3.2kHz
POWER SPECTRUM LOG
PCM ERROR SIGNAL
MIN = -1.457203
MAX = 3.259536
4dB
2dB
0
-2dB
10 log10 [Φee (ω)]
3.2kHz

FIG. 22

where

$$\rho_{xx}(kT) = \frac{E(X_i X_{i-k})}{E(X_{i-k}^2)} \quad , \qquad (2.1.20)$$

T is the sampling interval, and $1 - \frac{k}{128}$ is a triangular window function which smoothly time limits $\rho_{xx}(kT)$. The sampled data power spectrum is made up of impulses (i.e., numerical values) located at 31 Hz intervals.

$$\omega = \frac{\pi \ell}{128T} \approx (2\pi \times 31)\ell \qquad \ell = 1,2,....128 \qquad (2.1.21)$$

The long-term power spectrum of the ADPCM noise rolls off about 2 db between 1.6 kHz and 3.2 kHz; whereas, the log PCM noise is white. This small roll-off is probably due to overload effects.

In Fig. 23 sound spectograms of the input signal, ADPCM quantizing noise and log PCM quantizing noise are shown. The horizontal scale is time, the vertical is frequency and the darkness of the markings is proportional to the amplitude of the short-time power spectrum.[7]

$$\Psi_{xx}(\omega) = \int_{-\infty}^{\infty} \varphi_{xx}(\tau,t) \cdot \cos(\omega\tau)\, d\tau \quad , \qquad (2.1.22)$$

where

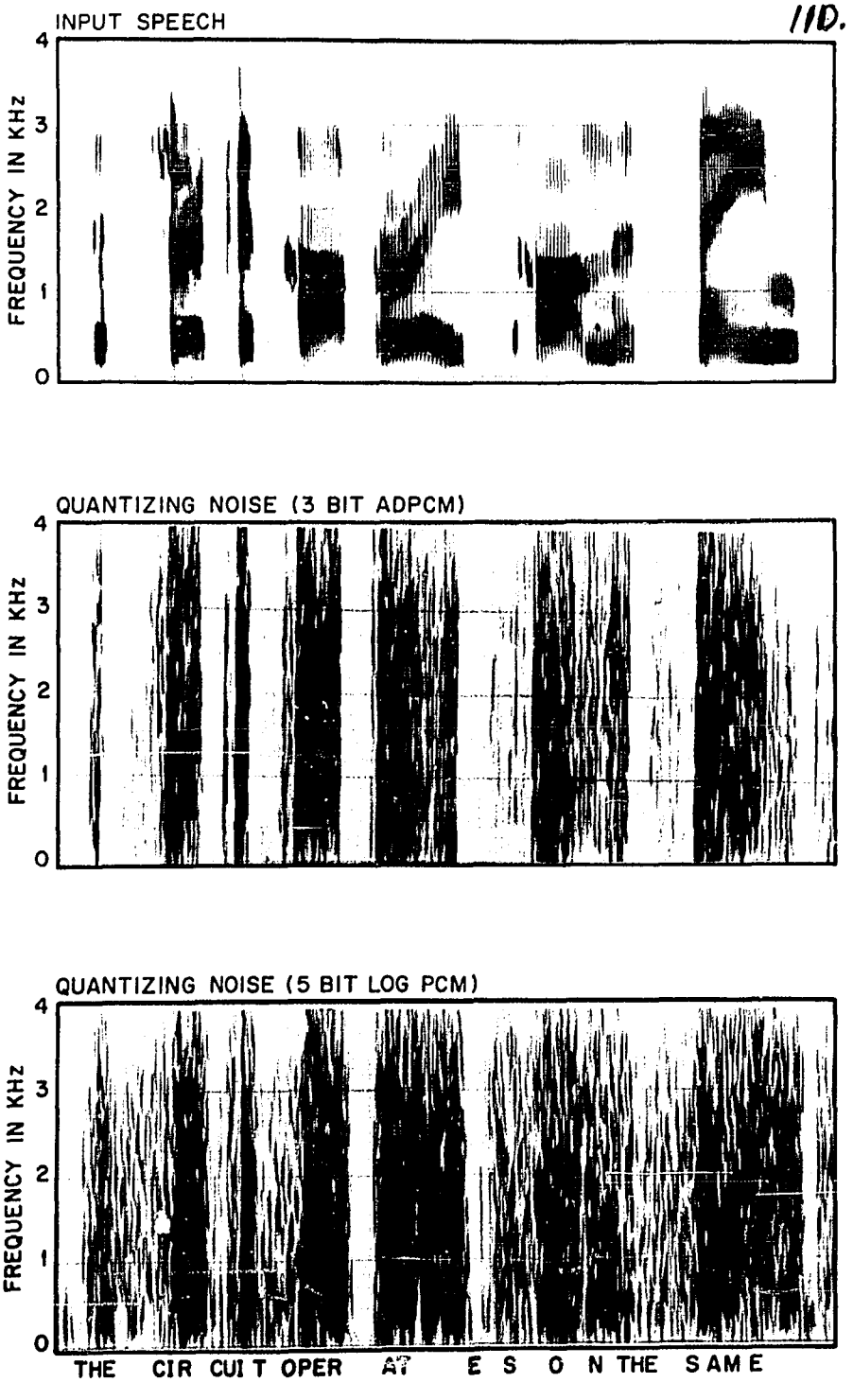

INPUT SPEECH
FREQUENCY IN KHz
4
3
2
1
0
QUANTIZING NOISE (3 BIT ADPCM)
FREQUENCY IN KHz
4
3
2
1
0
QUANTIZING NOISE (5 BIT LOG PCM)
FREQUENCY IN KHz
4
3
2
1
0
THE CIR CUI T OPER AT E S O N THE SAME

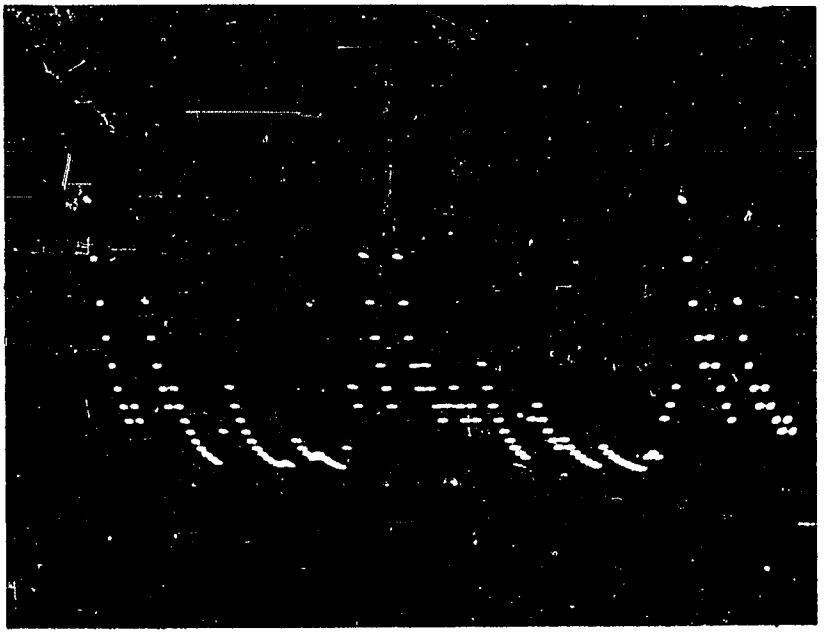
FIG. 23

$$\varphi_{xx}(\tau,t) = \int_{-\infty}^{t} X(\lambda)h(t-\lambda)\, X\, (\lambda+\tau)h(t-\lambda+\tau)\, d\lambda \qquad (2.1.23)$$

and where h(t) is the impulse response of the analyzing filter in the spectograph machine.

The dark bands of energy which appear on the spectogram of the input speech indicate the location of the resonant frequencies of the vocal cavity (i.e., the formant frequencies). The vertical stripes correspond to the peak amplitude regions of each pitch period of the speech waveform. In both spectograms of quantizing noise very little information related to the formant frequencies remains; however, rather clear markings related to vocal-cord excitation and syllabic boundaries remain in the spectogram of the ADPCM noise. This information is not as well preserved in the PCM noise.

In Fig. 24 some photographs of the quantizer companding signal, σ, are shown. Obviously, the amplitude of the quantizing noise samples in the ADPCM system are closely related to σ which in turn is an approximate envelope of the derivative of the input signal. This explains the presence of harmonics of the vocal-cord excitation in the ADPCM noise spectra. The noise in the PCM system is related to the envelope of the



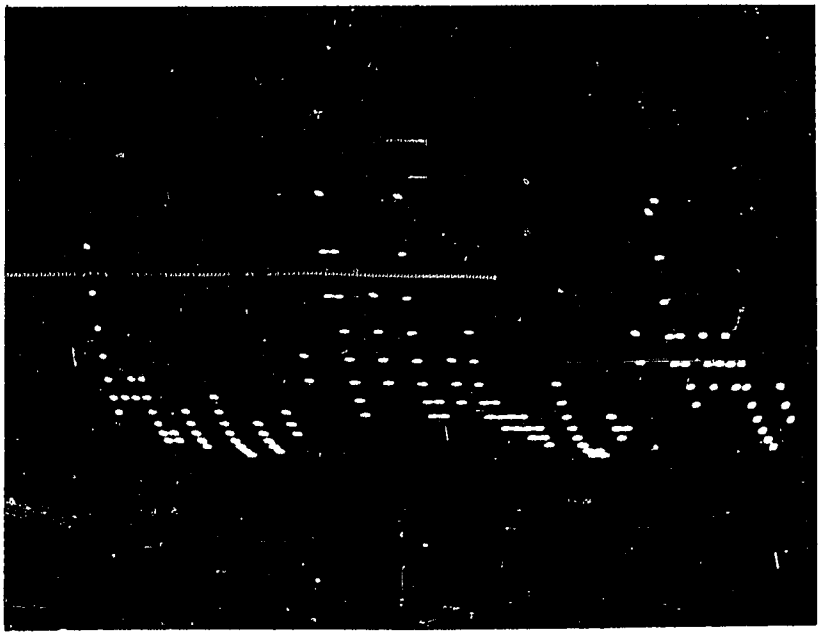


FIG. 24

input signal by the μ law quantization. (Larger amplitudes are more coarsely quantized than smaller amplitudes.) Although both noises contain pitch harmonics, the ADPCM noise sounds buzzier.

The values which σ may take on were limited to a 40 db range. The effective dynamic range of the eight level adaptive quantizer is also about 40 db when processing speech; whereas, that of a 32 level, $\mu = 100$, log PCM quantizer is less than 30 db.[24] This lack of range results in the generation of noticeable noise during quiet intervals as indicated on the spectrograms in Fig. 23.

Hence, the following features may be associated with noise in the ADPCM system:

(1) less energy at high frequencies,

(2) short-time spectra containing harmonics of the vocal-cord excitation function, and

(3) wide dynamic range resulting in lower overall energy while operating at low speech levels.

Further subjective comparisons of log PCM, adaptive delta modulation and adaptive DPCM are about to be conducted by A. E. Rosenberg. These tests will contain a wider selection of listening material. Informal listening to a wide variety of material processed by the hardware

coder described in the next section has convinced the author that an overall subjective advantage of 12 db to 18 db (2 to 3 bits/sample) is held by the ADPCM system over conventional log PCM coding.

Section 2 A Hardware Realization of the DPCM Coder With An Adaptive Quantizer

In order to show that an ADPCM coder is easily realized and to process a variety of speech signals in real time, a hardware coder was developed.

A block diagram of the hardware coder and a timing chart of the more important signals is shown in Fig. 25. The timing signals (CCLK, TCODE and others) are derived from a 768 kHz clock using the digital circuits shown in Appendix No. 2.

The sign of the difference between the input signal and the voltage on the integrator, sgn (x_i-y_i), is detected by the differential comparator and stored in shift register number 1 at the onset of the coding interval, (TCODE). Then either switch S1 or S2 is opened so that the voltage at the output of the integrator will change in the direction of the sign of the difference. After a time interval, τ, the sign of the difference is sampled again and the integrator is switched so that the output voltage will change in the direction of the new difference sgn $(x_i-y_{i+\tau})$. This process is repeated with the time interval reduced to $\tau/2$, $\tau/4$, etc. A DPCM code word is thus generated and stored in the register. The final integration in

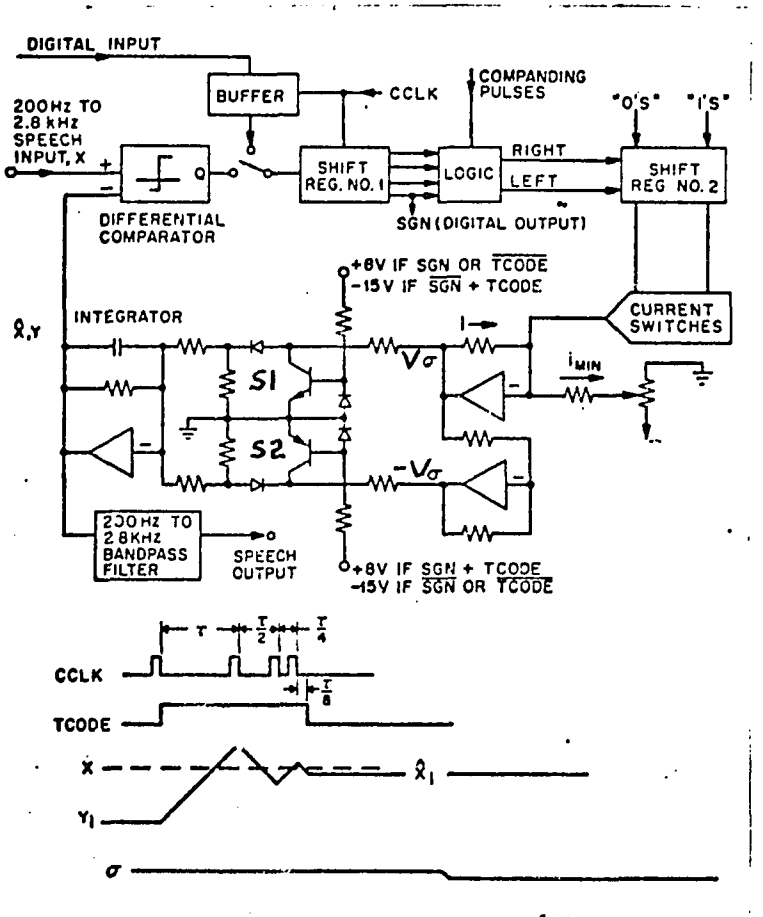

DIGITAL INPUT
BUFFER
CCLK
COMPANDING PULSES
"0"S
"1"S
200Hz TO 2.8 kHz SPEECH INPUT, X
SHIFT REG. NO. 1
LOGIC
RIGHT
LEFT
SHIFT REG NO. 2
DIFFERENTIAL COMPARATOR
SGN (DIGITAL OUTPUT)
+8V IF SGN OR TCODE
-15V IF SGN + TCODE
CURRENT SWITCHES
INTEGRATOR
S1
S2
200 HZ TO 2.8KHZ BANDPASS FILTER
SPEECH OUTPUT
+8V IF SGN + TCODE
-15V IF SGN OR TCODE
CCLK
TCODE

FIG. 25

the coding sequence is usually completed $\tau/4$ or $\tau/8$ microseconds after the entire code word has been clocked into the register.

When an eight level quantizer and a three bit code is used, an optimum quantizer is approximated by a uniform quantizer with enlarged end steps (see Fig. 26). Therefore, when an all ones or all zeros code occurs, the final coding interval is extended to $\tau/2$ microseconds. In order to approximate an optimum quantizer with 16 levels, the last two pulses in the CCLK sequence and the end of the coding interval must be altered in response to the code generated. (See Fig.27.) As of the final writing of the dissentation an optimum 16 level quantizer was not yet implemented; however, an adaptive DPCM coder with a 16 level uniform quantizer was realized using coding intervals of τ, $\tau/2$, $\tau/4$, and $\tau/8$.

When the coding interval ends, both S1 and S2 are closed to ground and a corrected estimate of the input signal, $\hat{X}_i$, is held on the integrator.

$$Y_{i+TCODE} = \hat{X}_i \qquad (2.2.1)$$

If the coding interval is made sufficiently small, the need for sample and hold circuits is eliminated.

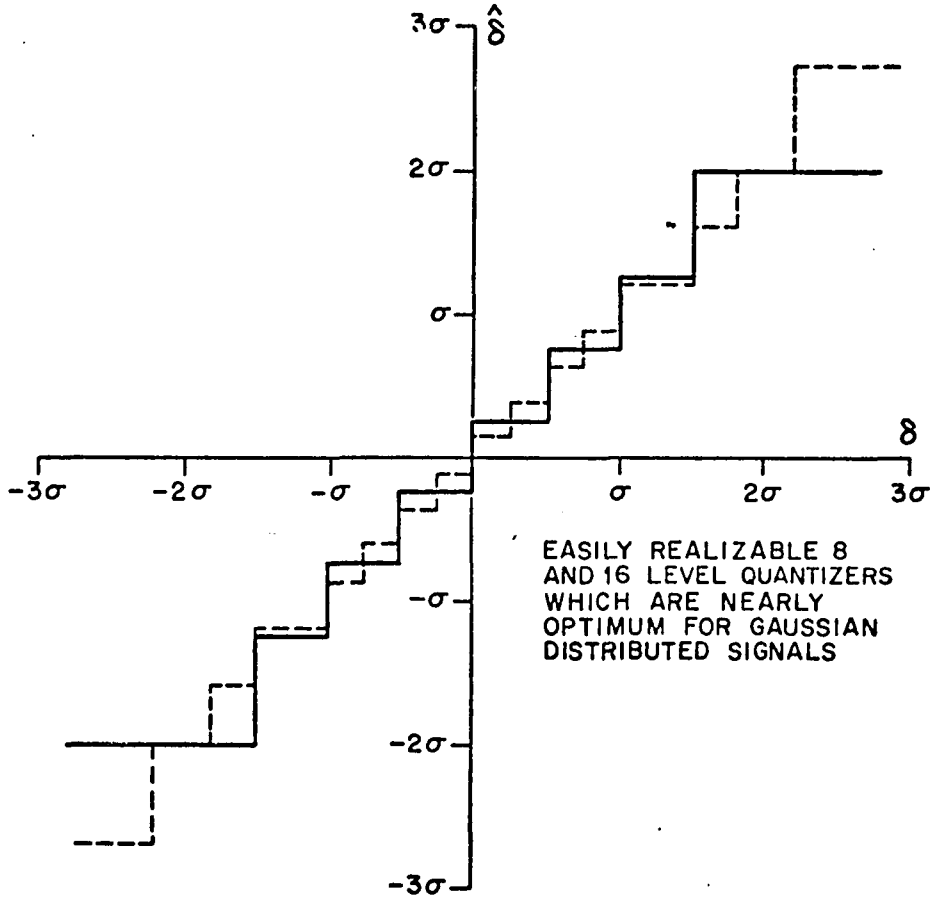
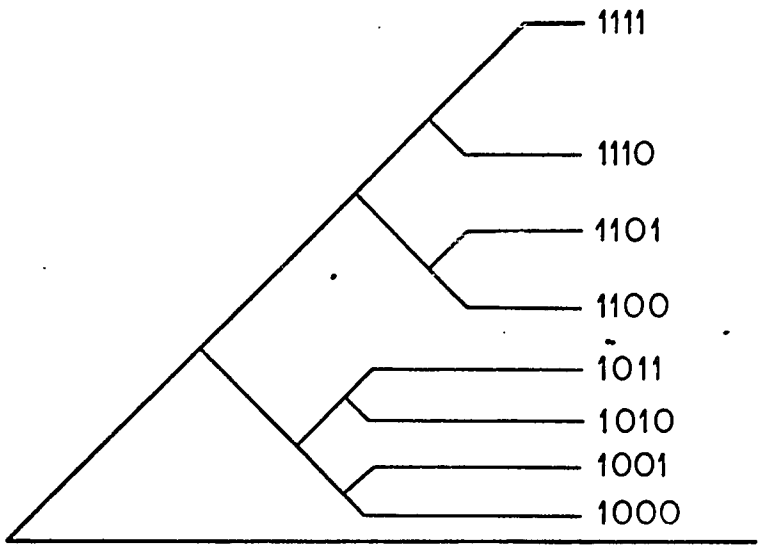

3σ
$\hat{\delta}$
2σ
σ
δ
-3σ
-2σ
-σ
σ
2σ
3σ
-σ
-2σ
-3σ
EASILY REALIZABLE 8 AND 16 LEVEL QUANTIZERS WHICH ARE NEARLY OPTIMUM FOR GAUSSIAN DISTRIBUTED SIGNALS

FIG. 26


1111
1110
1101
1100
1011
1010
1001
1000

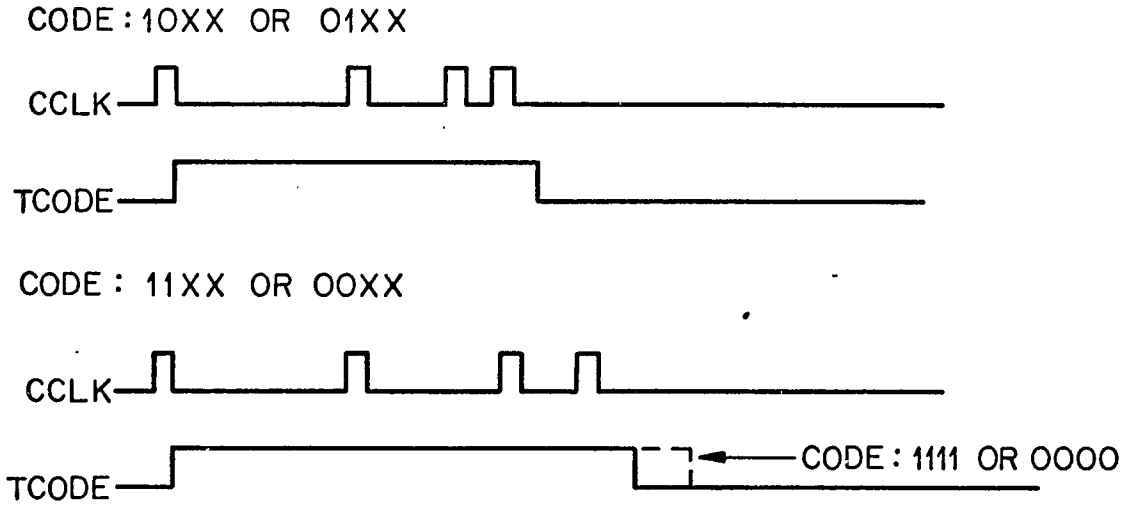

CODE : 10XX OR 01XX
CCLK
TCODE
CODE : 11XX OR 00XX
CCLK
TCODE
CODE : 1111 OR 0000

FIG. 27

The bandlimited input signal need not be sampled and held because it cannot change significantly during the coding interval (TCODE=20μSEC). Furthermore, the output need not be sampled and held because the coding circuit performs the holding function and the desampling filter removes the fast coding transitions.

The size of the quantizer is directly related to the companding voltage, Vσ, which in turn is adapted during the holding interval by the same rules obtained in the simulation for 6 kHz sampling. The companding circuit consists of a shift register, a set of twenty current switches and two operational amplifiers. The current switches are controlled from the shift register. A logic "zero" level turns the switches on, and a "one" turns them off. The current is increased by shifting "zeros" into one end of register number 2 or decreased by shifting "ones" into the other end. The currents are programmed so that each time a switch is turned on in sequence the total current is increased by twenty-five percent. Hence, the quantizer multipliers required (3.0, 2.0, 1.25 and 0.8) are obtained by a simple set of shift operations. The appropriate shift pulses are channeled from the clock circuit through the logic to the shift register under the control of the code word. The companding voltage is proportional

to the current from the switches plus a certain minimum current. A negative companding voltage, $-V\sigma$, is obtained at the output of a second inverting amplifier.

At any time during the holding interval the code word can be strobed into an output register and subsequently shifted onto a digitally multiplexed line for transmission. The timing circuitry can therefore serve several coders. Unlike analog scanning switches this form of multiplexing is virtually free of cross talk.

At the receiving end of the channel the code words are strobed into an input buffer and shifted serially into the decoder by a regenerated set of coding pulses (CCLK). The decoder operates in exactly the same manner as the encoder, except that the code words are obtained from the transmission channel by way of the input register rather than from a comparator.

If the switches, S1 and S2, grounded perfectly there would be no need for the holding diodes shown in Fig. 25. In this case the minimum companding voltage would be given by

$$V\sigma_{min} = (0.8)^{20}\, V\sigma_{max} \simeq .01\, V\sigma_{max} \qquad (2.2.2)$$

The transistor switches do not ground perfectly, nor identically. On a shorted switch there is a small collector to emitter voltage which varies with changes in temperature and collector current. Hence, the blocking diodes are necessary and $V\sigma_{min}$ must be increased so that

$$V\sigma_{min} = (0.8)^{20} V\sigma_{max} + V_D \qquad (2.2.3)$$

where V_D is the turn-on voltage of the blocking diode. There is some deviation from the desired companding rules due to the nonlinear characteristic of the diodes; however, these effects only become apparent when the input signal is reduced to 30 db below full level, and the output level is restored by an amplifier in series with the desampling filter. This is the same level at which the idle channel noise becomes audible.

The wide dynamic range of the coder is demonstrated in Fig. 28 where signal to noise ratios taken with 400 Hz, 800 Hz and 2400 Hz sinewaves are shown. The sinewave signal to noise ratios are defined as

$$SNR = 10 \log_{10} \left(\overline{\hat{X}_*^2} / \overline{N_*^2} \right) \qquad (2.2.4)$$

where $\overline{\hat{X}_*^2}$ is the power in the signal at the output of the desampling filter and $\overline{N_*^2}$ is the power that remains

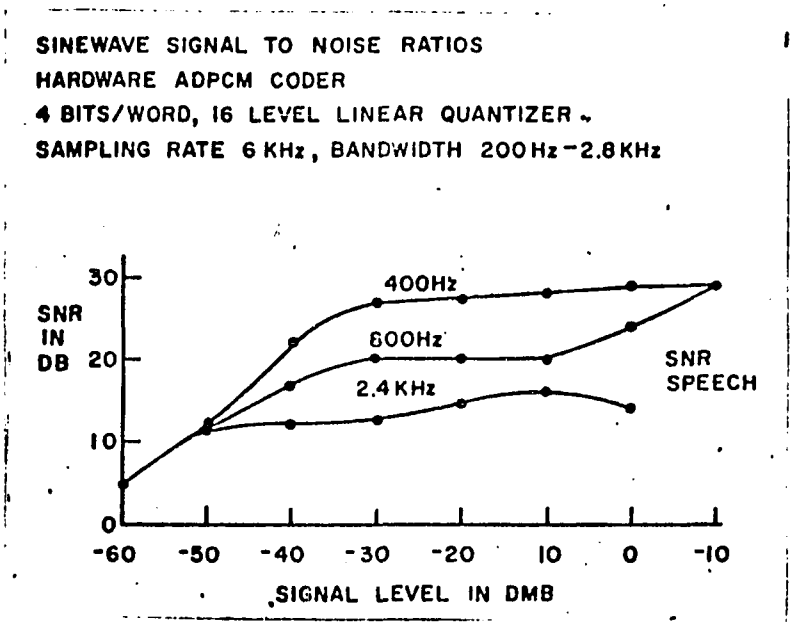
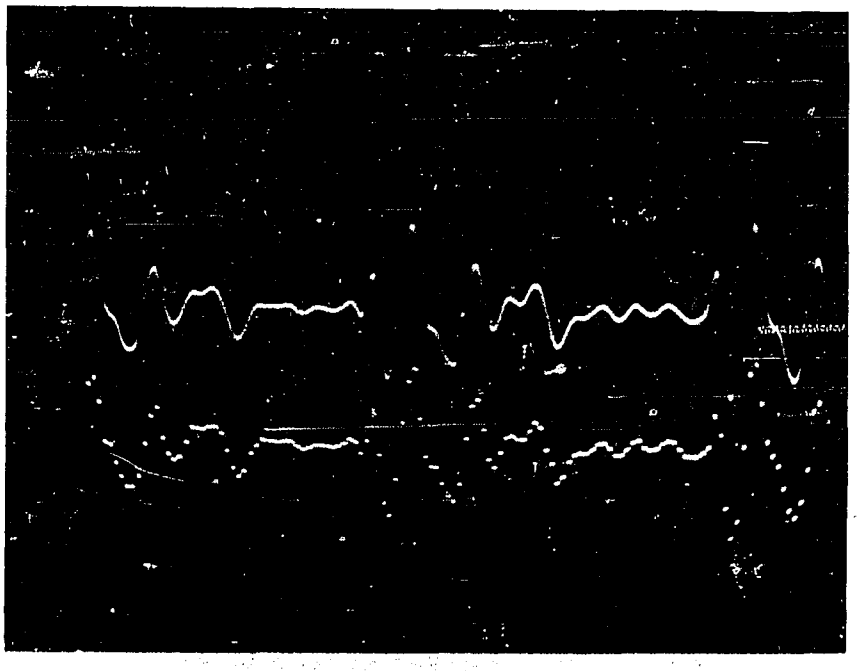

SINEWAVE SIGNAL TO NOISE RATIOS
HARDWARE ADPCM CODER
4 BITS/WORD, 16 LEVEL LINEAR QUANTIZER
SAMPLING RATE 6 KHz, BANDWIDTH 200 Hz - 2.8 KHz
SNR IN DB
30
20
10
0
400Hz
600Hz
2.4 KHz
SNR SPEECH
-60
-50
-40
-30
-20
10
0
-10
SIGNAL LEVEL IN DMB

FIG. 28

after $\hat{X}_*$ has been processed by a notch filter which rejects the input sinewave. It has often been implied that the signal to noise ratio taken in this manner with an 800 Hz sinewave is a good measurement of how the coder will perform when processing speech. No rule of thumb could be further from the truth. It happens that the signal to noise ratio taken with an 800 Hz sinewave is roughly equal to that obtained while processing speech. It should be noted, however, that the sinewave SNR rises well above the SNR for speech at levels of 0 dbm and +10 dbm. At these levels the companding circuit saturates ($V\sigma \rightarrow V\sigma_{max}$); hence, it is obvious that the quantizer's size could be adjusted to give much larger sinewave signal to noise ratios than those obtained with the companding law that is optimum for speech processing. The sinewave signal to noise ratios were used only to compare the performance of the hardware to the simulated coder and to demonstrate the dynamic range of the system.

Pictures of the companding signal, σ, are shown in Fig. 24. In fig. 29 the signal at the integrator is compared with the input, and in Fig. 30 the time scale is expanded so that the coding sequences can be seen.



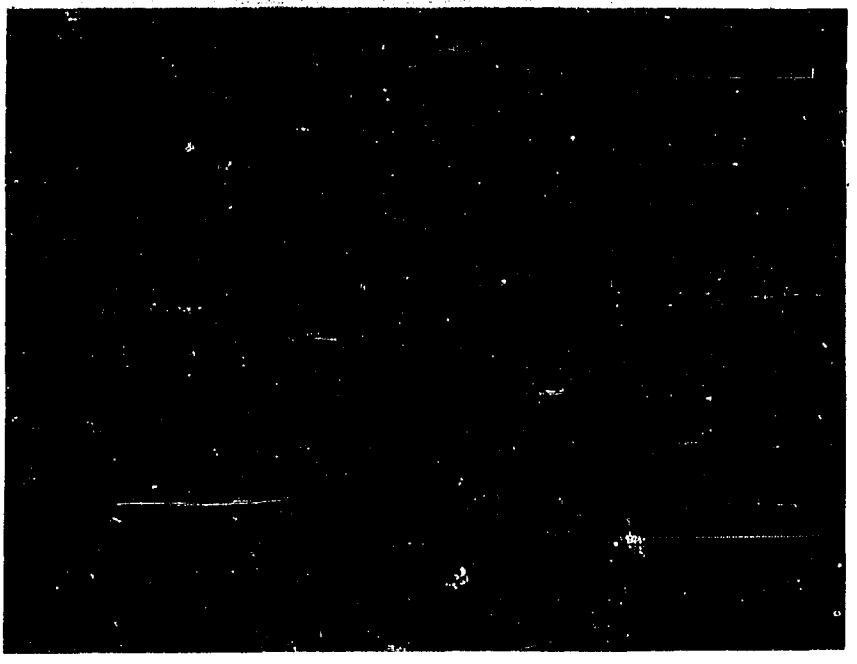
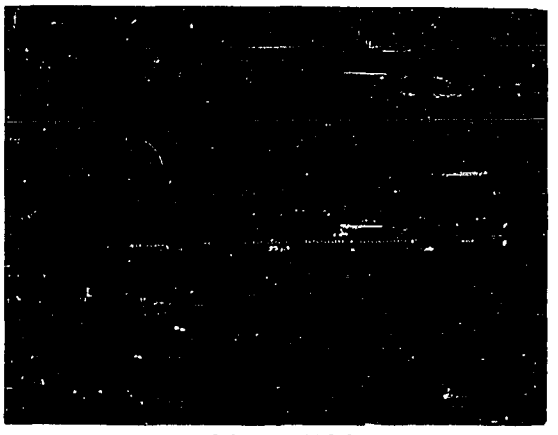
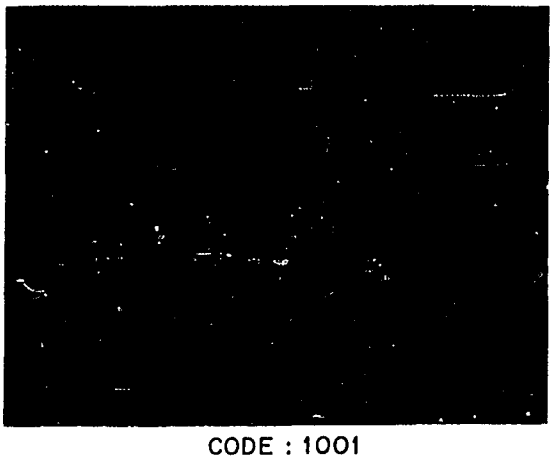


FIG. 27



CODE : 1100

CODE : 1011



CODE : 1001

FIG. 30

A unique application has been found for the ADPCM coder with a 16 level uniform quantizer in an automatic voice response system. Complete diagrams of the overall coder, the companding circuit, the clock circuit and part of an interface to a Data General Model 800 Nova computer are included in Appendix No. 2. The coder is used to encode speech and store the digital output on a computer memory or to regenerate speech from code stored on the memory. The gates incorporated in the coder between the comparator and the shift register (see Fig. A2.1) are used to change the coder from a decoder to an encoder or vice-versa under program control. A set of coders, the computer and a disc memory (hopefully to be replaced by a magnetic bubble memory) comprise an automatic voice-answer back system. Speech segments are recorded on the disc, and later played back in combinations which form useful messages.

The system is to be used to convert wire lists punched on cards to analog speech recordings. The recordings are played to wiremen so that they may wire without looking away from their work. This procedure is known to increase the wireman's speed and to reduce errors. A similar system will be used to channel wiring information to linemen on utility poles. Other automatic voice response systems are used to inform subscribers of

the fact that they have dialed a disconnected number or to tell them the time or give them a weather report, etc.

Section 3 A DPCM Coder with an Adaptive Predictor and an Adaptive Quantizer

In this section a DPCM coder is described in which a higher-order predictor is adapted to the changing characteristics of the input signal so as to minimize the mean square prediction error, $E(\delta^2)$. The quantizer in this coder is also adapted so that the remaining error might be optimally coded.

The performance of the coder described in the previous section could not be improved by applying fixed higher-order prediction. A 3.6 db improvement was realized by R. A. McDonald[16] using a second order fixed predictor in a conventional DPCM coder. This measured improvement was due primarily to matching of the predictor to the speech spectrum below 200 Hz (i.e., below the telephone band). The long term normalized autocovariances at 1 and 2 samples delay are given in Table 9 for the low pass filtered speech used by McDonald and for the 200 Hz to 3.2 kHz bandpass filtered speech used by the author. By applying Eqs. 1.2.6 and 1.2.51, the optimum predictor coefficients and the resulting signal to noise improvements are obtained.

	R. A. McDonald	P. Cummiskey
	.864	.793
	.557	.504
Optimum		
2nd order	$a_1 = 1.507$	$a_1 = 1.060$
Predictor	$a_2 = -.744$	$a_2 = -.337$
SNR improvement		
1st order predictor	6.0 db	4.3 db
2nd order predictor	9.6 db	4.7 db

TABLE 9

It is obvious that there is very little to be gained by incorporating a fixed higher order network in the ADPCM coder. This result was confirmed experimentally by the author and by P. Noll.[17] It has also been found that the subjective performance of McDonald's DPCM coder with a fixed higher order predictor is inferior to that of a conventional DPCM coder.[10]

In view of the success achieved by B. S. Atal and M. R. Schroeder[2] with adaptive prediction, it was decided that similar techniques should be explored with emphasis placed on simplifying the coder where possible. Hence, only the predictor coefficients described in Section II of part I (i.e., a_1----a_N) are used.

Parameters related to the fundamental frequency or pitch of the input speech are not explored in this paper.

The steepest descent gradient search techniques described in Section 2 of Part I were used to solve for the optimum predictor coefficients. Both of the search techniques described earlier were tried and it was found that the simpler of the two performed best. Hence,

$$\Delta a_{o_j} \Big|_{t=iT} = \frac{K \cdot Xo_{i-j} \ \mathrm{sgn}(\delta_{oi})}{\sum_{j=1}^{N} | \ Xo_{i-j} \ |} \qquad (2.3.1)$$

where $Xo_{i-j} = Xo((i-j)T)$ and a_{o_j} was adjusted once every sampling interval, T = 125 milliseconds. (See the left-hand side of Fig. 31). The greatest signal to noise improvement, $E(Xo^2)/E(\delta o^2)$, was achieved with an 8th order predictor and K = 0.094. The level of the error signal, δo, was thus reduced to 8.5 db below that of the input signal. This is a 4.7 db improvement over that achieved by the fixed first order predictor (i.e., the integrator) used in the previous ADPCM coders. If the autocorrelation function of the input signal is taken over a 5 or 10 millisecond interval and Eq. 1.2.6 is solved directly for each interval, an optimum adaptive

predictor is obtained. With such an optimum adaptive predictor the error signal would be reduced to about 12 db below the level of the input signal.[17] In this case, however, the predictor coefficients must be transmitted.

It is interesting to note that the amount of computation required to adjust the coefficients by the gradient search technique is approximately half of that required to obtain the autocorrelation function and solve Eq. 1.2.6 by the direct techniques described in Part I of section II. If the sampling rate is doubled but the taps on the predictor remain at 125 millisecond intervals, the gradient search may be refined (i.e., ao_j may be incremented twice as often). Then if K = .05, the error signal is reduced to 12 db below the level of the input signal. Hence, it is observed that as the number of computations per second used in the gradient search approaches the rate required by a direct solution, the SNR improvements obtained approach the same figure. Therefore, the search techniques described earlier offer a useful alternative method for adapting a predictor.

The optimum value for K is the value which equalizes the error due to coefficient oscillations and the error associated with an inability to track changes in the characteristics of the input speech. Fortunately, the

order of prediction required for speech signals[1] is not so large as to cause the steepest descent search to become too sluggish for effective tracking. The value of K can be diminished in inverse proportion to the rate of coefficient adjustment without increasing tracking noise. Meanwhile, the noise due to coefficient oscillations is diminished. This explains the further reduction of the prediction error achieved by doubling the sampling rate.

During development of the hardware coder described in the previous section, some experimentation with slow or continuous companding was carried out. The quantizer companding scheme was similar to that used by J. A. Greefkes and F. DeJager to vary the step size in their delta modulator. This brief trial convinced the author that performance similar to that achieved with instantaneous companding could be realized while adjusting the quantizer slowly in response to the magnitude of the prediction error averaged over a 5 to 10 millisecond interval.

---

[1] It is useful to increase the order of prediction until there are two coefficients for each resonant mode occurring in the speech waveform. In the 200 Hz to 3.2 kHz band there are four such resonances (3 formants and a vocal chord characteristic).

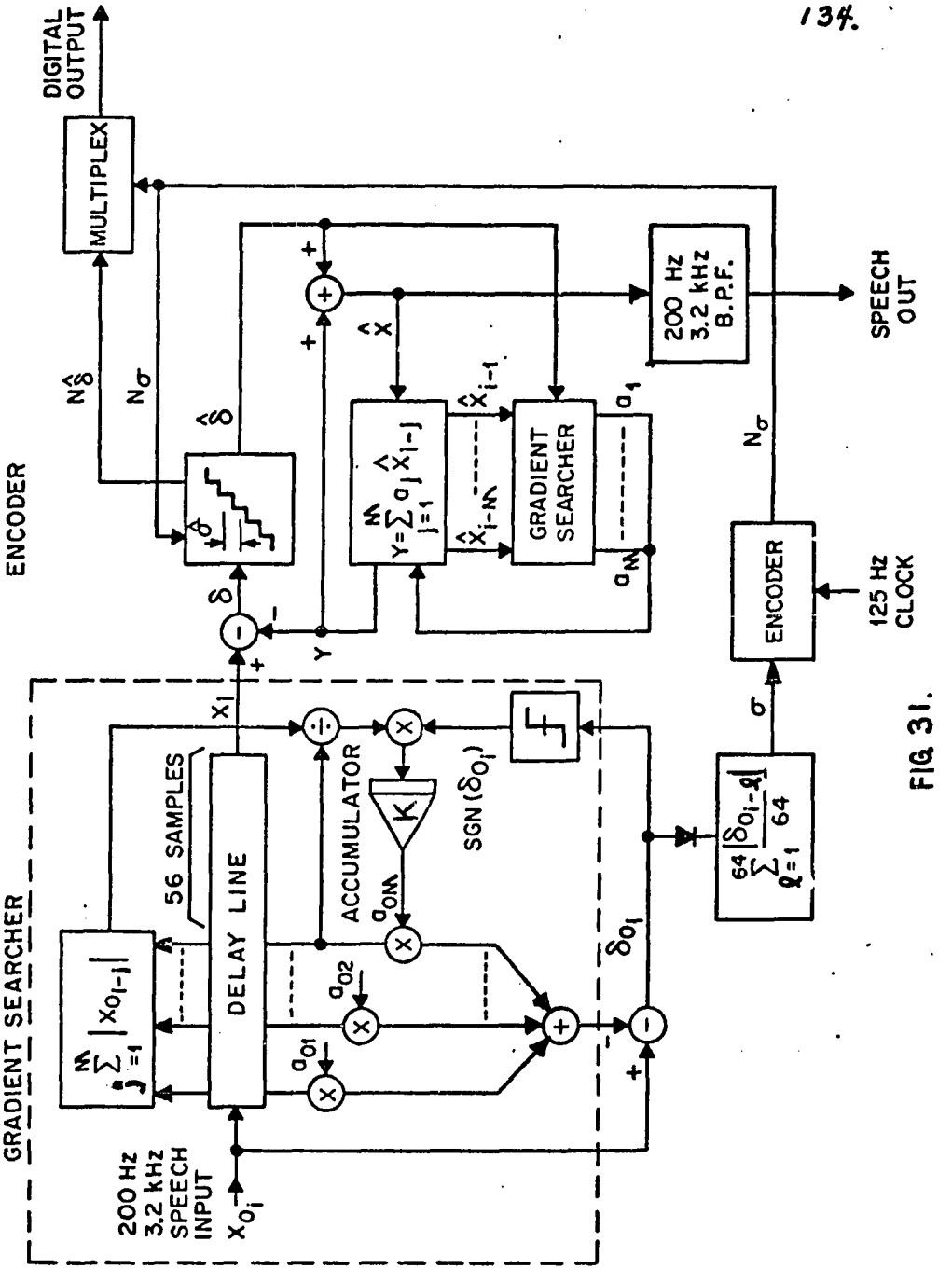

DIGITAL OUTPUT
MULTIPLEX
ENCODER
GRADIENT SEARCHER
200 Hz 3.2 kHz SPEECH INPUT
DELAY LINE
56 SAMPLES
ACCUMULATOR
GRADIENT SEARCHER
200 Hz 3.2 kHz B.P.F.
SPEECH OUT
ENCODER
125 Hz CLOCK

FIG 31.

The program for a Gradient Search was expanded and a new coder was simulated, where the predictor coefficients are adapted at the sampling rate, 8 kHz. The error signal generated by the predictor shown on the left-hand side of Fig. 31, δ_0, is rectified and averaged over 64 samples (i.e., an 8 millisecond interval). This average is used to adjust the quantizer in the encoder at the right. When the companding is slow, it is reasonable to expect that the probability distribution of δ will be similar to that of speech and its derivatives; therefore, an optimum Laplacian quantizer was used in this coder.

With the quantizer preset to embrace the expected prediction error, the predictor in the encoder is adjusted by a steepest descent gradient search routine where the searcher is fed the sign of the quantized prediction error and the previous corrected estimates of the input signal, $\hat{X}_{i-j}$.

$$\Delta a_j \Big|_{t=iT} = \frac{K\hat{X}_{i-j}\,\mathrm{sgn}(\hat{\delta}_i)}{\sum_{j=1}^{m} |\hat{X}_{i-j}|} \qquad (2\cdot3\cdot2)$$

It is obvious from Fig. 32 that this technique eliminates the need to transmit the predictor coefficients. Little or no data is required to transmit the companding

DECODER

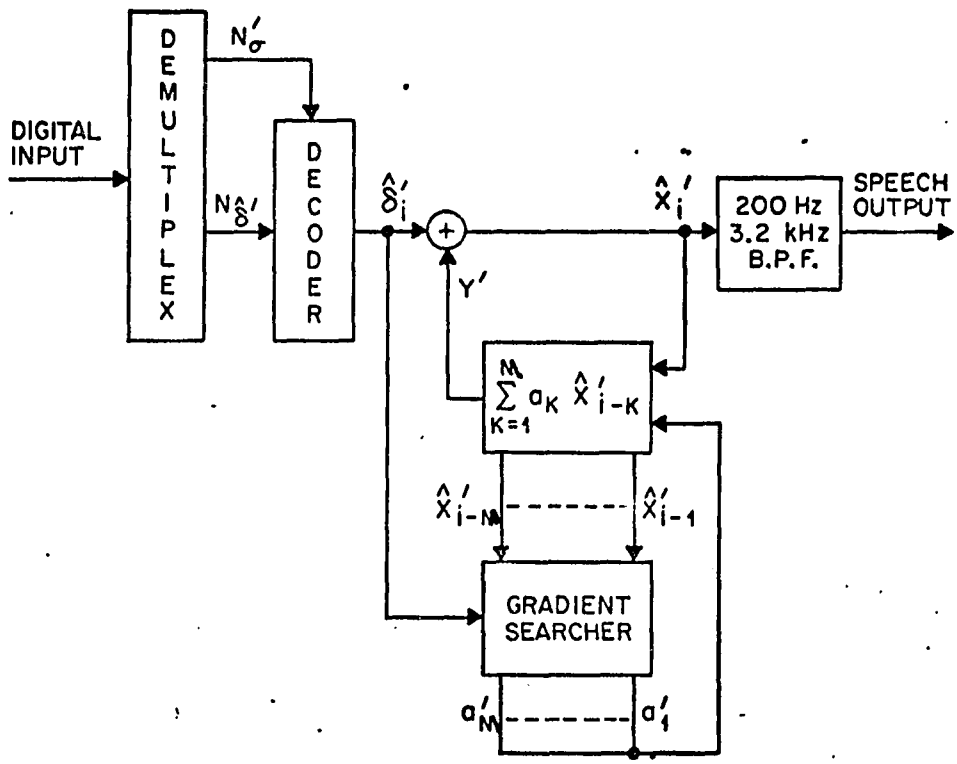


FIG. 32

information. Although attempts to adapt the quantizer using previous outcomes ($\hat{\delta}_{i-1} \cdots \hat{\delta}_{i-j}$) failed miserably in the presents of the adaptive predictor, it is likely that the technique used by Greefkes and DeJager will work well here, and no additional information will be required.

In Fig. 33 the signal to noise curves for the adaptive-predictive DPCM system are compared with the other systems discussed thus far. The ADPCM coder with adaptive-prediction is clearly superior to ADPCM, PCM and delta modulation encoders. Informal listening tests also indicate a subjective advantage over the other systems. The system shown in Figs. 31 and 32 regenerates telephone quality speech at only 3 bits/sample with an 8 level quantizer. At only 2 bits/word (4 levels) the output speech is still highly intelligible. The signal to noise curve for systems where more optimum prediction coefficients are transmitted would be about the same as the dotted curve in Fig. 33 because the additional information required roughly compensates for the gain in SNR.

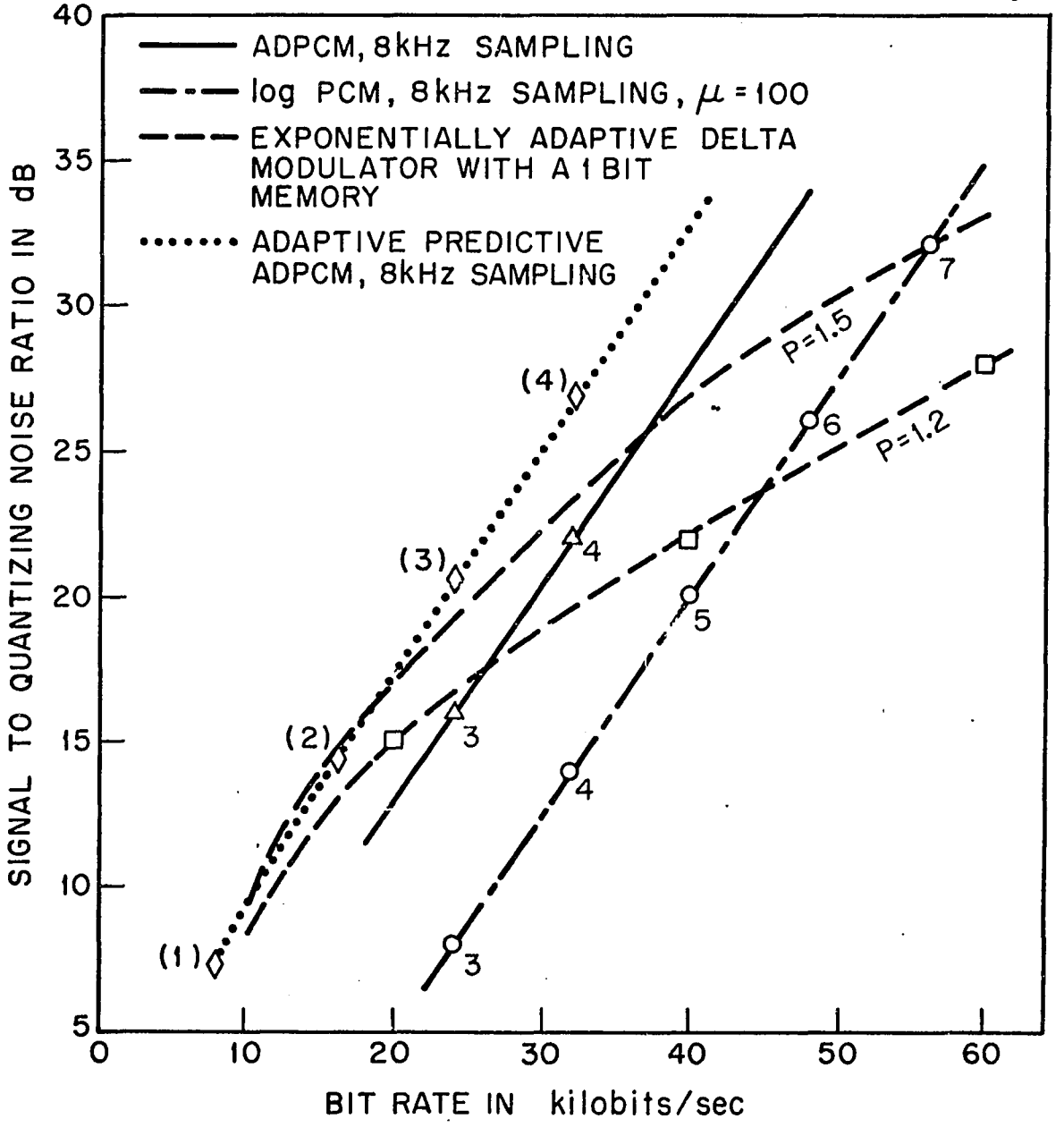

ADPCM, 8kHz SAMPLING
log PCM, 8kHz SAMPLING, μ = 100
EXPONENTIALLY ADAPTIVE DELTA MODULATOR WITH A 1 BIT MEMORY
ADAPTIVE PREDICTIVE ADPCM, 8kHz SAMPLING
SIGNAL TO QUANTIZING NOISE RATIO IN dB
BIT RATE IN kilobits/sec
P=1.5
P=1.2
(1)
(2)
(3)
(4)
3
4
5
6
7
0
10
20
30
40
50
60
5
10
15
20
25
30
35
40

FIG. 33

CONCLUSIONS

The DPCM system with an adaptive quantizer offers an 8 db advantage over conventional log PCM coding. When the adaptive DPCM coder is compared to log PCM on a subjective basis, an even larger advantage is observed. With the sampling rate reduced to 6 kHz, the signal to noise performance of the ADPCM coder is approximately equal to that of an adaptive delta modulator operating below 24 kilobits/sec (see Fig. 19). When a quantizer of 16 or more levels is used, the performance of the DPCM coder surpasses that of the delta modulator. An additional 4.5 db signal-to-noise advantage is obtained by incorporating a higher-order, adaptive predictor in the ADPCM coder. In this case the ADPCM coder is superior to the adaptive delta modulator regardless of the bit rate.

The quantizer in a DPCM system may be adapted instantaneously in response to the previous code word or slowly in response to the input level averaged over 5 or 10 milliseconds. It should be noted that slow companding must include some anticipation as in the coder with adaptive prediction. The initial simulations described in Section 1 of Part II show that the quantizer cannot be optimally adjusted in response to an average of previous quantizer outputs. The best performance was achieved with instantaneously-companded, optimum-Gaussian

quantizers or slowly companded, optimum-Laplacian quantizers.

The DPCM coder with an instantaneously adaptive quantizer was realized in integrated circuit hardware. A simple serial coding strategy was described with variations which allow for the realization of nearly optimum, nonuniform quantizers. The coding interval is made short so that no sample and hold circuits are required. The circuit is thus reduced to the same order of complexity required in PCM and adaptive delta modulation coders.

It was also shown that higher-order adaptive prediction is required if the ADPCM coder with a fixed integrator is to be improved. Nearly optimum predictor coefficients were obtained by adapting the coefficient vector in a direction opposite to the gradient with respect to the error magnitude. The gradient search technique has been shown to be a simple and computationally efficient method for calculating predictor coefficients. Finally, it was demonstrated that the gradient search technique can be applied to the quantized signals generated in the decoder, and that the need to transmit predictor coefficients may thus be avoided.

RECOMMENDATIONS

It is recommended that further study be given to adaptive prediction, to instantaneous and syllabic companding of the quantizer, and to coder performance in the presence of transmission errors.

IMPROVED PREDICTION

It has been shown that better predictor performance $\left(\text{i.e., higher SNR improvement, } \frac{E(\delta^2)}{E(x^2)}\right)$ is possible if the predictor is adjusted at a 16 kHz rate. Therefore the author recommends that the coefficients (a_j's in Fig. 31) be adjusted at 16 kHz by using interpolated values of the quantized signals, $\hat{X}_{i-j-\frac{1}{2}}$ and $\hat{\delta}_{i-\frac{1}{2}}$, where

$$\hat{X}_{i-j-\frac{1}{2}} = \frac{\hat{X}_{i-j} + \hat{X}_{i-j-1}}{2} , \tag{R.1}$$

$$\hat{\delta}_{i-\frac{1}{2}} = \frac{\hat{\delta}_i - \hat{\delta}_{i-1}}{2} , \tag{R.2}$$

$$\Delta a_j \Big|_{t=(i-\frac{1}{2})T} = \frac{K\hat{X}_{i-j-\frac{1}{2}} \operatorname{sgn}(\hat{\delta}_{i-\frac{1}{2}})}{\sum_{j=1}^{M} \left| \hat{X}_{i-j-\frac{1}{2}} \right|} \tag{R.3}$$

$$\text{and } \Delta a_j \Big|_{t=iT} = \frac{K\hat{X}_{i-j} \operatorname{sgn}(\hat{\delta}_i)}{\sum_{j=1}^{M} \left| X_{i-j} \right|} . \tag{R.4}$$

The author also recommends that adaptive prediction be used in a delta modulator where the signal is over-sampled at 8 kHz to 16 kHz. If the delay line taps in an adaptive predictor are spaced one sampling interval apart, the prediction error diminishes rapidly as the sampling frequency is increased (1). When such a predictor was incorporated in a delta modulator, the delta modulator became unstable due to the fact that some of the poles of $\frac{\hat{X}}{\hat{\delta}}(Z)$ may lie outside the unit circle.

$$\frac{\hat{X}}{\hat{\delta}}(Z) = \frac{1}{1 - \sum_{k=1}^{N} a_k Z^{-k}} \qquad \text{(R.5)}$$

The predictor may be realized using a series of second order sections. Then the coefficients can be limited so that the poles of $\frac{\hat{X}}{\hat{\delta}}(Z)$ will remain inside the unit circle. In this case,

$$\frac{\hat{X}}{\hat{\delta}}(Z) = \frac{1}{\prod_{k=1}^{N/2} (1-a_{2k-1}Z^{-1}-a_{2k}Z^{-2})} \qquad \text{(R.6)}$$

A block diagram of such a delta modulator is shown in Fig. 34.

(1) At 16 kHz sampling the prediction error was reduced to 24 db below the level of the input signal.

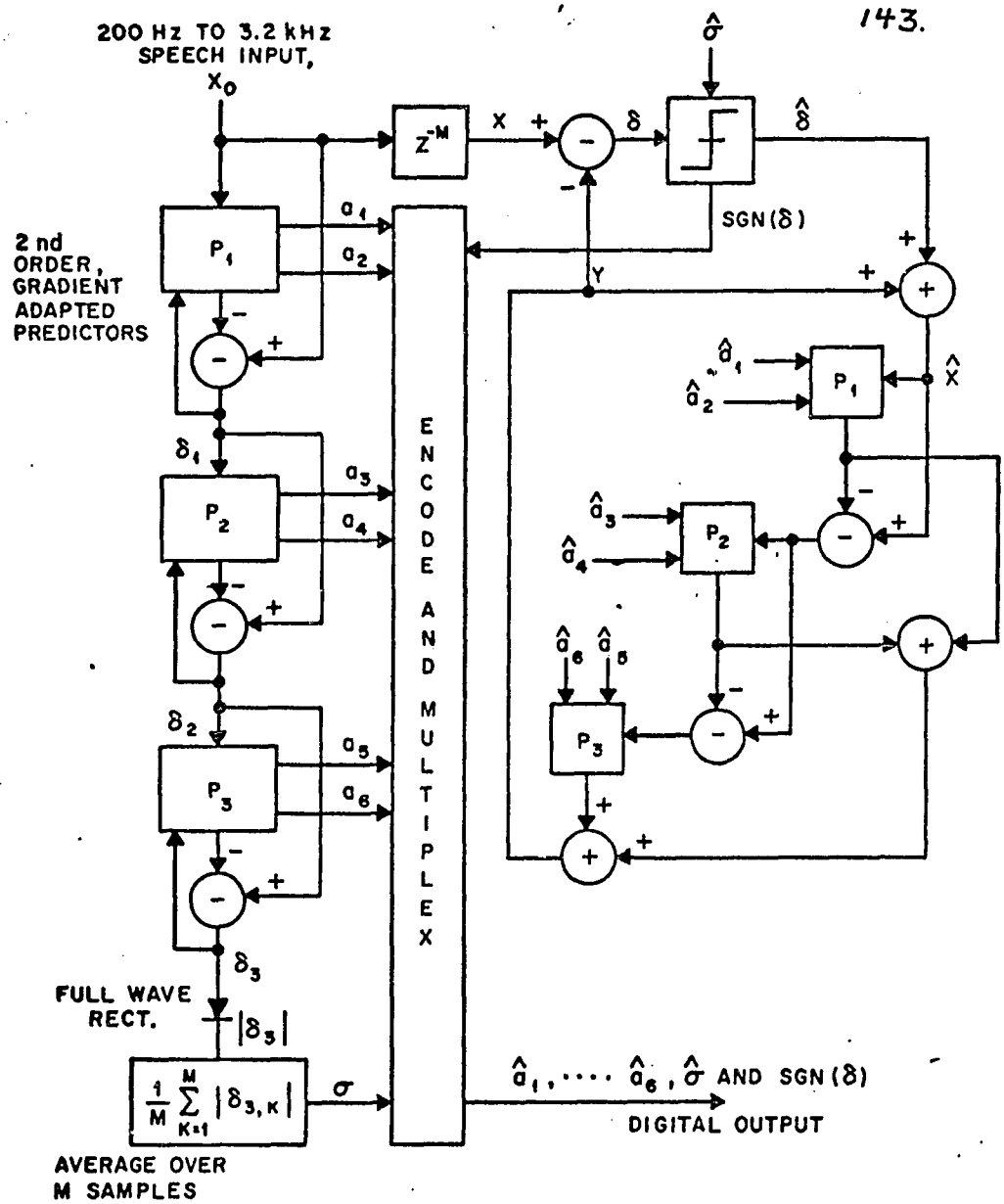

200 Hz TO 3.2 kHz
SPEECH INPUT,
X_0
Z^{-M}
X
δ
$\hat{\sigma}$
$\hat{\delta}$
SGN(δ)
2 nd
ORDER,
GRADIENT
ADAPTED
PREDICTORS
P_1
P_2
P_3
a_1
a_2
a_3
a_4
a_5
a_6
δ_1
δ_2
δ_3
Y
$\hat{a}_1$
$\hat{a}_2$
$\hat{a}_3$
$\hat{a}_4$
$\hat{a}_5$
$\hat{a}_6$
$\hat{X}$
ENCODE AND MULTIPLEX
FULL WAVE
RECT.
$|\delta_3|$
$\frac{1}{M}\sum_{K=1}^{M}|\delta_{3,K}|$
σ
AVERAGE OVER
M SAMPLES
$\hat{a}_1, \ldots \hat{a}_6, \hat{\sigma}$ AND SGN(δ)
DIGITAL OUTPUT

FIG. 34

Further effort could be invested toward realizing an efficient means for detecting the fundamental frequency of the speech signal In Fig. 29, it can be seen that during voicing the speech waveform is quasi-periodic and that a good estimate of the next sample is a sample taken one pitch interval in the past. Unfortunately, reliable pitch period detection is a task which is an order of magnitude more complex than computing the other predictor coefficients [-2].

IMPROVED COMPANDING

Both instantaneous and syllabic (slow) techniques for companding the quantizer require further study.

A more precise theory is needed for deriving the optimum, instantaneous, multipliers used to compand the quantizer in Fig. 16. The present theory only gives rough estimates for the multipliers; hence, an extensive search is required before the best values can be found.

In the coder with the adaptive predictor (see Fig. 31) a companding scheme is required so that the quantizer size need not be transmitted. If previous code words or quantizer outputs can be used to accomplish this task, the predictor and the gradient searcher at the left may be eliminated. Hence the encoder could be greatly simplified.

Slow or syllabic companding of the quantizer should also be persued in coders with fixed predictors.

PERFORMANCE IN THE PRESENCE OF CHANNEL ERRORS

It is also recommended that coder performance in the presence of transmission errors be studied. Subjective as well as objective measurements of performance must be taken on a variety of coders at various error rates. Further studies are also needed to evaluate digital coder performance on mobile radio links where fading is a major consideration.

Appendix No. 1

THE ON LINE COMPUTING FACILITY

The on-line computing facility shown in Fig. A1.1 was used to simulate adaptive DPCM coders. Input speech samples were recorded directly from the sound booth through the Spectrum Type LH42D filter, the Xerox Data Systems A/D converter, and the DDP 516 computer onto a disc memory. For permanent storage, the digitally coded speech was transferred from the disc memory to magnetic tape by way of the Hewlett Packard 3030 digital tape unit. A Xerox Data Systems D/A converter and another Spectrum filter are available for converting digitized speech back into continuous waveforms for listening and for recording onto analog tape. A display terminal is also connected to the system so that waveforms, graphic information, programs and other alpha-numeric information might be displayed. Hard copies of programs, data, etc. are available from an Inktronic printer and an ASR/33 teletypewriter. The Inktronic is used for printing out large blocks of data and entire programs; where as, the teletypewriter initiates, records and terminates transactions between the user and the computer.

A library of subroutines was available for transferring data between the computer and the peripheral

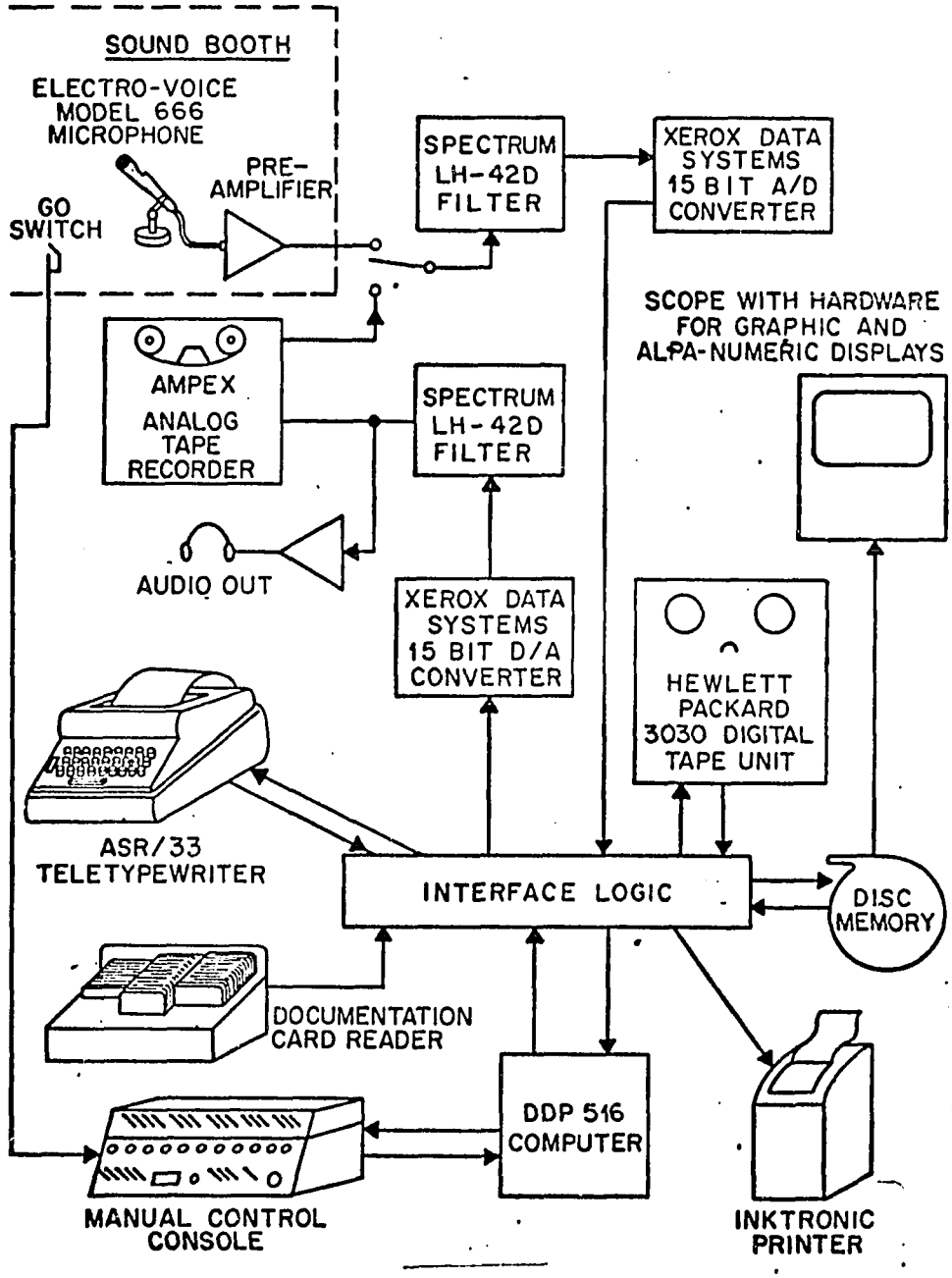

SOUND BOOTH
ELECTRO-VOICE MODEL 666 MICROPHONE
PRE-AMPLIFIER
GO SWITCH
SPECTRUM LH-42D FILTER
XEROX DATA SYSTEMS 15 BIT A/D CONVERTER
SCOPE WITH HARDWARE FOR GRAPHIC AND ALPA-NUMERIC DISPLAYS
AMPEX ANALOG TAPE RECORDER
SPECTRUM LH-42D FILTER
AUDIO OUT
XEROX DATA SYSTEMS 15 BIT D/A CONVERTER
HEWLETT PACKARD 3030 DIGITAL TAPE UNIT
ASR/33 TELETYPEWRITER
INTERFACE LOGIC
DISC MEMORY
DOCUMENTATION CARD READER
DDP 516 COMPUTER
MANUAL CONTROL CONSOLE
INKTRONIC PRINTER

FIG. A1.1

devices mentioned above. Some of these subroutines can be called manually from the console; whereas, most of them are called from Fortran IV or machine code (DAP) statements. The author wrote his main program for data handling in Fortran IV, and several machine code subroutines for simulating DPCM systems with fixed and adaptive quantizers, and predictors. Other subroutines for calculating autocorrelation functions, probability distributions and signal to noise ratios, and for simulating log PCM were also programmed.

When all of the software is loaded in the computer and an input speech sample is stored on the disc, it is possible to simulate a coding scheme and play back the processed speech within two minutes time. All of the other programs mentioned above have similar run times. It is also possible to alter programs and recompile or reassemble within five or ten minutes. Hence, by taking advantage of repeated, short turn around, man-machine interactions, it was possible to try several companding techniques and to optimize parameters rapidly.

The precise diagrams of the ADPCM circuit are not required for an understanding of the dissertation. Due to difficulties involved in reducing this information onto 8-1/2" x 11" paper, Appendix No. 2 (pages 149-164) has been omitted from this copy. The reader may, however, obtain this data from the bound copy of the dissertation at the Newark College of Engineering Library.

REFERENCES

1. Abate, J., "Linear and Adaptive Delta Modulation", Proceeding of the IEEE, vol. 55, No. 3, March 1967, also "Linear and Adaptive Delta Modulation", Newark, N.J.: D. Eng. Sc. dissertation, Newark College of Engineering, April 1967.

2. Atal, B. S. and M. R. Schroeder, "Adaptive-Predictive Coding of Speech Signals", Bell Sys. Tech. J., vol. 49, No. 65, 1970, pp. 1973-1986.

3. Carroll, J. D., "Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1, Theory", New York and London: Seminar Press, Inc., 1972.

4. Cummiskey, P., "A Hardware Implementation of an Exponentially Adaptive Delta Modulator", Unpublished, Bell Telephone Laboratories Memorandum and a talk presented to the Acoustical Society of America, April 1971.

5. DeJager, F., "Delta Modulation, a Method of PCM Transmission Using a One-Unit Code", Philips Research Report, vol. 7, 1952, pp. 442-466.

6. DeJager, F. and J. A. Greefkes, "Continuous Delta Modulation", GLOBECOM, vol VI, 1964 and Philips Res. Reports, vol 23, 1968, pp. 233-246.

7. Faddeev and Faddeeva, "Computational Methods of Linear Algebra", San Francisco and London: W. H. Freeman and Co.

8. Flanagan, J. L., "Speech Analysis, Synthesis and Perception, New York: Academic Press, 1965.

9. Fletcher, H., "Speech and Hearing in Communication", Princeton, N.J.: Van Nostrand, 1953.

10. Grether, C. B. and R. W. Stroh, "Subjective Evaluation of Differential Pulse Code Modulation using the Speech Goodness Rating Scale", IEEE Conference on Speech Communication and Processing, IEEE Cat. No. 72CH0596-7AE, April 1972.

11. Jayant, N. S., "An Adaptive Delta Modulation with a One-Bit Memory", Bell Sys. Tech. J., vol. 49, No. 3, March,1970, pp. 321-342.

12. Jayant, N. S. and A. E. Rosenberg, "The Preference of Slope Overload to Granularity in Delta Modulation of Speech" Bell Sys. Tech. J., December, 1971, pp. 3117-3125.

13. Max, Joel, "Quantizing for Minimum Distortion", IRE Transactions on Information Theory, vol. IT-6, March, 1960, pp. 7-12.

14. Lee, Y. W., "Statistical Theory of Communication", New York: John Wiley and Sons, Inc., 1960.

15. Levitt, H., C. A. McGonegal and L. L. Cherry, "Perception of Slope-Overload Distortion in Delta Modulated Speech Signals", IEEE Trans. on Audio and Electro-Acoustics, vol. 18, No. 3. Sept., 1970, pp. 240-247.

16. McDonald, R. A., "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems - Particular Applications to Voice Signals", Bell Sys. Tech. J., vol. 45, Sept., 1966.

17. Noll, P. W., "Non-Adaptiveand Adaptive DPCM Coding of Speech Signals", Overdruk Vit Polytechnish Tijdschrift, Editie Electro-technick-Electronikca, Nr. 19, 1972.

18. Paez, M. D. and T. H. Glisson, "Minimum Mean Squared-Error Quantization in Speech, PCM and DPCM Systems", IEEE Trans. on Comm., vol. 20, No. 2. April, 1972, pp. 225-230.

19. Panter, P. F., "Modulation, Noise and Spectral Analysis", New York: McGraw Hill, 1965.

20. Rabiner, L. R., "Techniques for Designing Finite Duration Impulse Response Digital Filter", IEEE Trans. Com. Tech., vol. 20, No. 2, April, 1971, pp. 188-195.

21. Schouten, J. F., F. DeJager and J. A. Greefkes, "Delta Modulation, a New Modulation System for Telecommunication", Philips Tech. Review, vol. 13, March, 1952, pp. 237-245.

22. Smith, B. "Instantaneous Companding of Quantized Signals", Bell Sys. Tech. J., vol. 36, 1957, pp. 653-709.

23. Tomozawa, A. and H. Kaneko, "Companded Delta Modulation for Telephone Transmission", Proc. N.E.C., 1966.

24. "Transmission Systems for Communications", Members of the technical staff Bell Telephone Laboratories, Feb., 1970, pp. 566-607.

25. Winkler, M. R., "High Information Delta Modulation", IEEE International Conv. Record, Part 8, 1963, pp. 260-265.

VITA

The author was born on , at . He graduated from Wood-Ridge High School in 1955 and then served as an electrician's mate in the United States Navy (1955-1959). After being honorably discharged from the service, Mr. Cummiskey attended Fairleigh Dickinson University where he received the Bachelor of Science degree in Electrical Engineering. In June 1962, he was employed by Bell Telephone Laboratories as a technical aide. In June 1966, he received a Master of Science in Electrical Engineering from Newark College of Engineering and advanced to Associate Member of Staff at Bell Laboratories. During the past ten years at Bell Laboratories, Mr. Cummiskey developed experimental circuits used in speech research. The topic of his master's thesis was the design of a terminal-analog speech synthesizer. The author was also called upon to interface equipment to small digital computers and to program them. This led to the work with adaptive delta modulation and ADPCM.