THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Volume 58 July–August 1979 Number 6, Part 2

Evaluation of Adaptive Speech Coders Under Noisy Channel Conditions

By C. SCAGLIOLA

(Manuscript received June 20, 1978)

An experiment has been performed in which the digital transmission of speech coded by adaptive differential PCM was simulated under noisy channel conditions. The experiment was done with two aims: (i) to get information on the subjective effect of channel errors and the influence of various design parameters on the speech quality under various conditions and (ii) to find objective measures for predicting the overall quality of the processed speech over a wide range of circuit conditions. The subjective results show that, for a speech transmission through a channel with bit error probability up to 1/256, best results can be obtained with a slow error recovery, associated with fast quantizer adaptation. The use of slow error recovery and slow quantizer adaptation is preferable for channels with very high bit error rates, like 1/32. Overall subjective quality is well predicted by the sum of two terms: (i) an objective performance measure of the noise present on the output signal, disregarding any effect of level mismatching due to the sensitivity of the adaptation algorithms to channel errors and (ii) a measure of the level mismatching which takes into account both the average gain on the output signal and its fluctuation in time. The best prediction scores are achieved by three newly defined objective performance measures, two-level compensated segmental SNRs, and a spectral signal-to-distortion ratio.

I. INTRODUCTION

The design of digital waveform coders for speech communications must face the inevitable presence of channel errors. Adaptive coders, like ADPCM (adaptive differential PCM), in which the adaptation of the

quantizer step-size is derived from the transmitted binary stream and no error-protected side information is sent to the receiver, may be particularly sensitive to this problem. In fact, a single channel error may cause a multiplicative offset between the signal level at the receiver and that at the transmitter. This offset may persist indefinitely if no error dissipation mechanism is provided.

Recently, some algorithms of quantizer adaptation have been developed that make the effect of a single transmission error die out over time, so the transmitter and receiver can resynchronize their step-size estimates.[1,2] The possibility of obtaining such results is physically due to the fact that these algorithms have an imperfect adaptation: the step size increases more quickly and decreases more slowly for low input levels than for high ones. In this way, the step size is overestimated for low input levels and underestimated for high ones, thereby reducing the dynamic range of the coder. Dynamic range and error dissipation rate vary inversely, and the designer has to balance between them.

In the case of speech transmission, the choice of the appropriate design parameters must be based on a precise evaluation of the subjective quality of the coded speech. Use of the conventional long-term signal-to-noise ratio as an estimator of the subjective quality would be, in this instance, completely misleading, at least because an offset in the signal amplitude between input and output, due to an offset in step sizes caused by an error, will be reflected in a noticeable squared difference between the two waveforms, while it may not be subjectively disturbing.

To study the subjective performance of ADPCM coders operating under both error-free and noisy channel conditions, an experiment has been conducted, as summarized in Section II. The following three sections provide a brief description of the coding method, the definition of several objective performance measures, and the description of the experimental design and testing procedure. Sections VI and VII provide analyses of subjective and objective measurement data. In Section VIII, the results of the previous two sections are discussed and a physical interpretation is given of the principal findings on quality prediction.

II. OVERVIEW OF THE EXPERIMENT

The experiment included 12 different ADPCM coding schemes, which comprised all combinations of two bit rates, two adaptation time constants, and three error dissipation rates. These systems processed a total of 288 speech samples from four talkers (two male and two female), at two different power levels (24 dB apart) and with three different probabilities of independent errors on the channel.

Twenty listeners rated the quality of the processed speech samples

on a scale from 1 to 9. The odd values were associated with the adjectives: unsatisfactory, poor, fair, good, excellent. In addition to the subjective data, a fairly large number of objective performance measures were also taken on the processed speech samples.

The aims of the experiment included the study of:

(*i*) The influence on speech quality of the above design parameters.

(*ii*) The optimum combination of parameters for a given error probability.

(*iii*) The objective measures or combinations of objective measures which are good predictors of speech quality even under noisy channel conditions.

The principal conclusions drawn from the analyses of the subjective and objective data are:

(*i*) Contrary to a common feeling, a very slow error dissipation is sufficient to ensure good robustness of the ADPCM coder to error rates even much higher than those encountered in a normal telephone connection. When no recovery mechanism is provided, a fairly slow adaptation makes the system not very sensitive to channel errors in the range of error rates typical of a telephone connection.

(*ii*) For speech transmission through a channel with bit error rate up to 1/256, best results can be obtained with a very slow error dissipation, associated with fast quantizer adaptation; when the slow error dissipation is associated instead with a slow adaptation, the system becomes fairly robust to very high error rates, like 1/32, at the expense of a slight quality deterioration at low error rates.

(*iii*) Good predictors of subjective quality were found to be two-segmental SNR measures in which a compensation of the level mismatching between input and output was performed on a frame-by-frame basis. The combination of any of these measures with two separate measures of level mismatching further improved the prediction accuracy.

III. ROBUST ADPCM SYSTEM: A BRIEF DESCRIPTION

Figure 1 is a block diagram of the ADPCM coder-decoder used in the experiment. The predictor is a second-order transversal filter, with tap coefficients 1 and −0.5. The step size $\Delta(k)$ is adapted according to the robust algorithm described in Ref. 1, which permits synchronizing the step-size estimates at transmitter and receiver, after a transmission error occurs, during a period of error-free transmission

$$\Delta(k+1) = \Delta^{\beta}(k) \cdot M(I(k)), \tag{1}$$

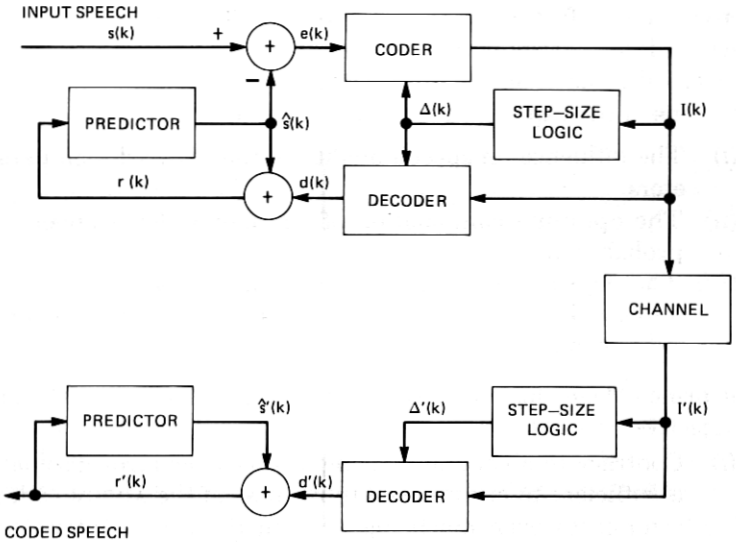

Fig. 1—Block diagram of ADPCM coder-decoder.

where the exponent β, $0 < \beta < 1$, is the decay constant and $M(I(k))$ is the step-size multiplier at time k. The multiplier $M(I(k))$ depends only on the actual code word $I(k)$ and assumes $N = 2^{B-1}$ distinct values $\{M_1, M_2, \cdots M_N\}$, where B is the number of bits used to encode the prediction error $e(k)$.

The decay speed is shown to be independent of the actual code word transmitted and also of the values of the multipliers, being only a function of the decay constant β.[1] When $\beta = 1$, the decay speed is zero and the decay time infinite. With $\beta < 1$, the decay speed increases, but at the same time the dynamic range decreases.

It has been shown that the loading factor (ratio between range of the quantizer and rms quantizer input) is a constant as a function of the input level if $\beta = 1$, but it is a decreasing function of the input level if $\beta < 1$.[1] The form of their relationship is almost linear, with slope approximately inversely proportional to $\log(M_N/M_1)/(1 - \beta)$, as indicated by Fig. 6 of Ref. 1 and as recently proved theoretically by D. Mitra.[3] Therefore, with a small β and a small ratio between the maximum and minimum multipliers, the coder will produce more granular noise for low input level and more overload distortion for high levels.

The values of the multipliers play another important role in the overall performance of the coders under noisy channel conditions. In fact, they determine the magnitude of the initial offset after a single

transmission error. In the worst case, the offset is given by

$$\left| \frac{\Delta'(k)}{\Delta(k)} \right|_{\max} = \frac{M_N}{M_1}. \tag{2}$$

In the present experiment, the multipliers M_i were related by the linear relationship:

$$M_i = [\alpha + C(1 - \alpha)(i - 0.5)]\hat{\Delta}^{(1-\beta)}, \tag{3}$$

where $\hat{\Delta}$ is the step size that gives optimum performance at the desired nominal input level. This was fixed at −21 dBm, i.e., 27 dB under the saturation threshold for the signal that in the internal 16-bit computer representation is 32767. This relationship was chosen because, for $\beta = 1$, the adaptation algorithm coincides with the magnitude estimation algorithm described by Castellino et al.,[4] and for this algorithm more information about the subjective effects of its parameters are available.[5] In eq. (3), C essentially determines the mixture of granular noise and clipping distortion in the decoded prediction error at the nominal level. The parameter α controls mainly the speed of adaptation and hence the ratio between maximum and minimum multipliers.

IV. OBJECTIVE PERFORMANCE MEASURES

Several objective performance indices were measured for each utterance in the experiment. The speech samples used as input to the coders were low-pass filtered at 3.4 kHz before being sampled and converted into digital form by a 16-bit A/D converter operating at 8-kHz sampling rate. Figure 2 is a block diagram of the simulated circuit arrangement for performing coding and measurements. A filter after the ADPCM decoding limits the bandwidth of the output speech as in a real situation. A secondary path provides the reference signal $s_0(k)$ with which the filtered output $r_0(k)$ is compared to compute the objective performance measures. With two identical filters in the main

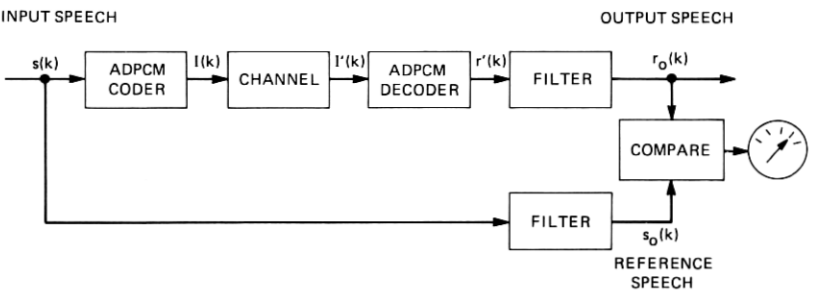

Fig. 2—Block diagram of simulated circuit arrangement for coding and measurements.

and reference paths, only the distortion introduced by the coder is measured. The filters are 5th-order elliptic low-pass with 3.4-kHz cutoff frequency, 0.25-dB in-band ripple, and at least 40 dB stopband attenuation.

The measures are classified in the two categories of time domain measures and frequency domain measures. Frequency-weighted signal-to-noise ratios are included in the first category because they rely strongly on the exact time synchronization of the two waveforms and on the absence of phase distortion.

4.1 Time domain measures

4.1.1 Long-term signal-to-noise ratio (SNR)

$$\text{SNR} = 10 \log \frac{\sum_k s_o^2(k)}{\sum_k [s_o(k) - r_o(k)]^2}, \tag{4}$$

where k ranges over all the samples of the utterance. SNR is the ratio between the long-term signal energy and the long-term noise energy, the noise being defined as the difference between reference signal $s_0(k)$ and output signal $r_0(k)$.

4.1.2 Segmental signal-to-noise ratio (SNR_{seg})

Here the utterance is divided into adjacent segments of J samples each, and the signal-to-noise ratio in each segment is measured in decibels. The noise is still defined as the difference between corresponding samples of reference and output speech. The segmental SNR is the average of these measures over the M segments of the utterance.

$$\text{SNR}_{\text{seg}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\sum_{j=1}^{J} s_o^2(j + mJ)}{\sum_{j=1}^{J} [s_o(j + mJ) - r_o(j + mJ)]^2}. \tag{5}$$

In this experiment, $J = 128$, corresponding to 16 ms segments. This measure, proposed by Noll,[6] was recently found to correlate very nicely with subjective ratings of ADPCM-coded speech.[5] A very important feature added to this basic formula consists in discarding from the computation those segments in which the signal power is below −54 dBm. This threshold, whose value was found to be appropriate for high quality speech,[7] was introduced to avoid a slight idle channel noise having an unduly great negative weight in the overall performance measure. This is done also in all the following time domain measures.

4.1.3 Compensated signal-to-noise ratio (SNR_{com})

This measure was specifically formulated to compensate for the level variations that may occur under channel error conditions when the coder has a slow error dissipation. The difference between reference and output signals due to level offset should not be measured in fact as noise.

To compensate for these level variations, let us formulate the coding process in the mth segment as composed of an amplification of the input signal, the addition of an uncorrelated random noise, and a possible dc component. Therefore, the output process in the mth segment can be written as

$$r_o(k) = g(m)s_o(k) + q(k) + Q. \tag{6}$$

This coincides with the simple linear regression model of $r_0(k)$ on $s_0(k)$. Therefore, the gain factor $g(m)$ is the slope of the regression line of the ouput on the reference signal in the mth segment, and the noise term $q(k)$ is the minimum error made in predicting $r_0(k)$ from $s_0(k)$.[8]

Let us define the signal-to-noise ratio in the mth segment as the ratio between the variances of $g(m) \cdot s_0(k)$ and $q(k)$. This can be shown to be only a function of the correlation coefficient $\rho(m)$ between the reference and output signals in the mth segment:[8]

$$\rho(m) = \frac{JS_{sr}(m) - S_s(m)S_r(m)}{\sqrt{[JS_{ss}(m) - S_s^2(m)][JS_{rr}(m) - S_r^2(m)]}}, \tag{7}$$

where $S_x(m)$ indicates the summation of $x(j)$ and $S_{xy}(m)$ the summation of $x(j) \cdot y(j)$ over the J samples of the mth segment.

Averaging in decibels across the M segments in the utterance, the compensated SNR turns out to be:

$$\text{SNR}_{\text{com}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\rho^2(m)}{1 - \rho^2(m)}. \tag{8}$$

4.1.4 Average gain (G)

This is the average in decibels of the gain factor $g(m)$ defined before, across the M segments. From simple linear regression analysis, the gain $g(m)$ is:[8]

$$g(m) = \frac{JS_{sr}(m) - S_s(m)S_r(m)}{JS_{ss}(m) - S_s^2(m)}. \tag{9}$$

The average gain, which is an indication of how much the output level was increased or decreased on the average, with respect to the input level, is:

$$G = \frac{1}{M} \sum_{m=0}^{M-1} 20 \log g(m). \tag{10}$$

4.1.5 Gain fluctuation (σ_g)

This is simply the standard deviation of the gain $g(m)$, measured in decibels, across the M segments. It is a measure of how much the output level fluctuates owing to transmission errors.

$$\sigma_g = \left[\frac{1}{M}\sum_{m=0}^{M-1}(20 \log g(m))^2 - G^2\right]^{1/2}. \tag{11}$$

4.1.6 Maximum signal-to-noise ratio (SNR_{max})

An alternate way of compensating for the level mismatching between reference and output speech signals was found by defining the noise in the mth segments as

$$\epsilon(k) = s_o(k) - \hat{s}_o(k), \tag{12}$$

where

$$\hat{s}_o(k) = a_1(m) \cdot r_o(k) + a_o \tag{13}$$

is the least-square estimate of $s_0(k)$ based on $r_0(k)$.

The ratio between the variances of the reference signal and the minimum estimation error $\epsilon(k)$ (hence the name "maximum SNR") is again only a function of the correlation coefficient $\rho(m)$, defined by eq. (7).

Averaging again in decibels across the M segments in the utterance, the maximum SNR is

$$\text{SNR}_{\max} = \frac{1}{M}\sum_{m=0}^{M-1} 10 \log \frac{1}{1-\rho^2(m)}. \tag{14}$$

It is readily seen that $\text{SNR}_{\max}$ is always greater than SNR_{com} and that the two measures give essentially different results only for low-quality coding conditions.

4.1.7 Frequency-weighted, segmental, signal-to-noise ratios

This term indicates a fairly large class of measures. In these measures, the frequency axis is partitioned into many bands, usually nonuniform, the reference and output spectra are compared, some performance measure is then computed over each band, and these measures are averaged across the bands. In the measures described below, the spectra are computed over 256 points (32 ms). The segmental measures are obtained by averaging the measures taken every 128 samples (16 ms).

Three measures are reported here. They are described more in Refs. 7 and 9. The partitioning of the frequency axis is effected in those three cases according to the 16 classical articulation bands.[10]

$$\text{SNRF}_1 = \frac{1}{M}\sum_{m=0}^{M-1}\left[\frac{1}{16}\sum_{j=1}^{16} 10\log\frac{S_j(m)}{N_j(m)}\right] \tag{15}$$

$$\text{SNRF}_2 = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{16} L_j(m)\ 10\log\dfrac{S_j(m)}{N_j(m)}}{\sum_{j=1}^{16} L_j(m)} \tag{16}$$

$$\text{SNRF}_5 = \frac{1}{M}\sum_{m=0}^{M-1} 10\log\frac{1}{1+\sum_{j=1}^{16}\dfrac{N_j(m)}{S_j(m)}}, \tag{17}$$

where $S_j(m)$ is the energy of the reference signal $s_0(k)$ in the jth frequency band in the frame m and $N_j(m)$ is the corresponding noise energy. The noise is again defined as the difference between reference and output signals, the latter being preventively divided by the gain $g(m)$ previously defined to compensate for level fluctuations. However $g(m)$ is computed on the 256-point analysis window. In eq. (16), a "loudness weighting" has been introduced. The term $L_j(m)$ corresponds approximately to the subjective loudness in band j, and is computed as

$$L_j(m) = \int_{\text{band } j} |S_r(f)|^{1/2}\, df, \tag{18}$$

where $S_r(f)$ is the spectrum of the gain-compensated output speech $r_0(k)/g(m)$.

4.2 Frequency domain measures

All the spectral measures here presented and used in the experiment are based on the concept of linear prediction or inverse filtering.[11] The speech signal is represented by the pth order autoregressive model:

$$s(k) = \sum_{i=1}^{p} a_i s(k-i) + u(k), \tag{19}$$

where $u(k)$ is the white spectrum excitation function and the a_i's are the coefficients of the inverse filter

$$A_s(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}. \tag{20}$$

The coefficients a_i's are computed to minimize the residual power of the signal at the inverse filter output.

In this paper, the dissimilarity between the spectra of reference and output speech in a given frame is computed essentially by comparing the residual powers of the signals $s_0(k)$ and $r_0(k)$ filtered by the inverse filters $A_s(z)$ and $A_r(z)$, derived from the same two signals. Four residual powers can be computed in the mth signal frame:

(*i*) $P_e(m)$ obtained passing $s_0(k)$ through $A_s(z)$.
(*ii*) $P_d(m)$ obtained passing $r_0(k)$ through $A_s(z)$.
(*iii*) $P_{e'}(m)$ obtained passing $r_0(k)$ through $A_r(z)$.
(*iv*) $P_{d'}(m)$ obtained passing $s_0(k)$ through $A_r(z)$.

Four objective measures based on these concepts are presented in the following paragraphs.

4.2.1 LPC distance measure (D_1)

This measure, proposed by Itakura,[12] is also called log likelihood ratio. The distance between output and reference speech in the mth frame is defined as

$$D_1 = \ln \frac{P_d(m)}{P_{e'}(m)}. \tag{21}$$

It can be shown that D_1 can be expressed in terms of spectral differences between the LPC models of the two frames of speech.[13] Moreover, it results that the spectral difference is most heavily weighted in the peaks of the input speech smoothed spectrum, i.e., in the speech formants.

Interchanging the roles of reference and output speech, a different log likelihood ratio is obtained:

$$D_2 = \ln \frac{P_{d'}(m)}{P_e(m)}, \tag{22}$$

which has the same basic properties as D_1.

The measure used here is actually the arithmetic mean of D_1 and D_2, averaged across the M segments of the utterance:

$$D_I = \frac{1}{2M} \sum_{m=0}^{M-1} \left[\ln \frac{P_d(m)}{P_{e'}(m)} + \ln \frac{P_{d'}(m)}{P_e(m)} \right]. \tag{23}$$

4.2.2 Bharucha index (D_B)

This index is again a distance measure, similar to the log likelihood ratio. It has been formulated by Bharucha[14] and its definition is also

reported in Ref. 7. The basic idea is that of measuring the noise introduced by the coder by "notching out" the speech spectrum by means of a time-varying linear filter, whose transfer function is matched to the inverse of the short-term spectral envelope. The quality index proposed by Bharucha is essentially the average increase in the residual power at the output of the "notch filter," due to coding:

$$D_B = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \left[\frac{P_s(m)}{P_r(m)} \cdot \frac{P_d(m)}{P_e(m)} \right], \tag{24}$$

where $P_s(m)$ and $P_r(m)$ are the powers of the reference and output speech in the mth segment and provide the appropriate scaling in the measure. It can be shown that an uncorrelated noise component in the output speech is inversely weighted, frame by frame, by the smoothed LPC spectrum of the input speech signal.[7,14] Therefore, the noise has more weight in those frequency bands where the signal energy is low; this is probably in conformity with subjective noise evaluation.

4.2.3 Spectral signal-to-distortion ratios

Following the same basic idea of the Bharucha index, two other measures were derived in the form of signal-to-distortion ratios. In fact, they are measured in decibels, and they increase with increasing quality, like the time domain SNRs. With the first SDR, a distortion power is defined as the difference between the prediction error powers $P_d(m)$ and $P_e(m)$ defined above.

Before taking the difference, however, the term $P_d(m)$ is multiplied by $P_s(m)/P_r(m)$ to compensate for level differences between input and output.

$$\text{SDR}_1 = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{P_s(m)}{P_d(m) \cdot [P_s(m)/P_r(m)] - P_e(m)}. \tag{25}$$

In the second SDR measure, the difference between the signal-to-prediction error ratios in decibels that is averaged to give D_B is instead computed relative to the input signal-to-prediction error ratio, and this new ratio is again averaged in decibels across the segments of the utterance.

$$\text{SDR}_2 = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{10 \log[P_s(m)/P_e(m)]}{10 \log[P_s(m)/P_e(m)] \cdot [P_d(m)/P_r(m)]}. \tag{26}$$

In this experiment, D_B, SDR_1, and SDR_2 are computed with an analysis window of 160 samples (20 ms) that was shifted by 128 samples (16 ms) every frame. The inverse filters for computing $P_e(m)$ and $P_d(m)$ were always of the 20th order.

V. EXPERIMENTAL DESIGN AND PROCEDURE

5.1 Circuit conditions

The choice of the experimental design variables was dictated by the criterion of a broad quality range and of each value the variables assumed having caused an effect that could be perceived for at least some combination of the other variables.

Two bit/sample values were chosen, $B = 3$ and $B = 4$, and three values of the decay constant, $\beta = 1$, $\beta = 255/256$, and $\beta = 63/64$.

The condition $\beta = 1$ was included because it should give the broadest dynamic range and because it makes the ADPCM system identical to the earlier schemes.[4,15]

For the step-size multipliers, given by formula (3), C was kept constant for each bit rate and equal to the values 0.65 and 0.41 that were found to be subjectively optimum for $B = 3$ and $B = 4$, respectively.[5] The parameter α was given two values, 0.75 and 0.96875. The first value produces a rather fast adaptation, with a corresponding time constant of 0.5 ms. With $\alpha = 0.75$, the ratio M_N/M_1 is about 4 dB and 5.5 dB for $B = 3$ and $B = 4$, respectively. The second value of α produces a rather slow adaptation, the corresponding time constant being 4 ms. The ratio M_N/M_1 is much smaller, about 0.5 dB and 0.7 dB for 3- and 4-bit/sample, respectively.

Two different input levels, 24 dB apart, were used for every combination of the other conditions. They were $L = -33$ dBm and $L = -9$ dBm, i.e., symmetrical around the nominal input level -21 dBm for which the coder was designed to have optimum performance. The level -21 dBm was not included to keep the dimension of the experiment within reasonable limits of feasibility.

Finally the channel was characterized by three different probabilities of independent errors, $P(e) = 0$, $P(e) = 1/256$, $P(e) = 1/32$.

Summarizing, the experiment included 72 conditions which comprised all the combinations of $B = 3, 4$; $\alpha = 0.75, 0.96875$; $\beta = 1,255/256, 63/64$; $L = -33, -9$; $P(e) = 0, 1/256, 1/32$.

5.2 Preparation of stimuli

Each experimental condition was simulated four times, using as input signals four sentences spoken by two male and two female talkers. Each talker spoke into a high-quality dynamic microphone, while seated in a sound-proof booth. The digital recordings had been generated by low-pass filtering the amplified microphone signal at 3.4 kHz, and then sampling and converting it into digital form by a 16-bit A/D converter operating at 8-kHz sampling rate.

For simulating the different input levels, the sentences, all previously adjusted to the same mean power level of -21 dBm, were multiplied by a constant factor at the coder input, and then divided by the same

factor at the output. In this way, each processed sentence was listened at the same level, unless channel errors and/or slope overload of the coder had caused output level variations.

From the 288 simulations, two analog test tapes were generated, each containing in a different random order two simulations of each experimental condition, one with a male and one with a female talker. For each talker, 18 different sentences read from a different phonetically balanced list were used, so that in each tape the same sentence appeared only twice.

5.3 Testing procedure

Twenty paid subjects (10 for each tape), all students from junior and senior classes of local high schools, judged the 288 stimuli. They listened to the processed speech binaurally over Pioneer SE 700 earphones at a nominal level of 80 dB_{SPL}, while seated in a double-walled sound booth. As pointed out before, the level of individual sentences varied according to the particular experimental conditions. The total listening time for each group of subjects was about 30 minutes, with a short break after the 80th sentence. After each stimulus, the subjects had 4 seconds to record their judgments. They were asked to rate the quality of the stimuli according to the adjectives: excellent, good, fair, poor, unsatisfactory. Their answer sheet contained 144 rows of short lines divided into nine columns, with the odd ones labeled with the adjectives. In this way, the subjects were allowed to check intermediate ratings, if they chose to do so.

The categorical judgments expressed by the listeners were subsequently converted into numerical scores, assigning value 1 to the category "unsatisfactory," value 9 to the category "excellent," and intermediate integer values to intermediate categories.

Before the actual test sessions took place, the subjects listened to 12 practice sentences different from those used in the experiment, spoken by the same four talkers, and representative of the range of quality they expected in the test.

VI. ANALYSIS OF SUBJECTIVE RESULTS

6.1 Control variables

The purpose of the subjective test was to assess the different behavior of coders in the presence of different operating conditions, namely, channel error rate and input level. Other sources of variability in listener responses are expected to cancel out in the average data for each experimental condition. Before averaging the data for each experimental condition, it was necessary to assess the importance of such extraneous sources of variability, as differences in the way the listeners judged the stimuli and differences due to talker voices.

6.1.1 Listeners

To assess the variability due to listener differences, their responses were analyzed according to MDPREF.[16,17] This is a factor analytic procedure that derives a geometrical multidimensional space representation, in which the stimuli are represented by points and the subjects by vectors. The projections of the points on a vector are the best fit with the scores given to the stimuli by that subject. Basically, MDPREF reveals whether the subjects attended to different psychological attributes in the stimuli or if they attached different weights to each of the various attributes. In the solution for the 72 experimental conditions and the 20 subjects, the first principal component accounted for only 55 percent of the variance, while the remaining 45 percent was distributed over all the other components: 4.2 percent for the 2nd, 3.8 percent for the 3rd, 3.4 percent for the 4th, 3 percent for the 5th, etc.

The fact that 45 percent of the total variance was accounted for by so many axes in an almost uniform fashion indicates that these axes do not represent different perceptual attributes of the stimuli, but that they account only for the "noise" in the subjective data. In other words, in spite of the low variance accounted for by the first axis, it is evident that the listeners attended essentially to the same attributes with the same weights and then only a unidimensional solution exists. Therefore, the mean of the listeners' ratings for each condition were used for the subsequent analyses.

6.1.2 Talkers

An analysis of variance was computed to study the variability of the scores obtained by the talkers of different sex and to assess the validity of averaging the ratings across talkers to perform an analysis of the effect of the design variables. The analysis showed that the difference due to the sex of the talker was highly significant. The average score was 4.40 for female talkers and 5.07 for male talkers. On the other side, however, all the interactions between the sex of the talker and the design variables were not significant. This indicates that the sex of the talker influenced the average value of the ratings, but not the relative ranking of the various experimental conditions. Therefore, the mean ratings across listeners and talkers were used for further analyses, reducing the variability of the data to that due to the physical variables of the coders and the circuits.

6.2 Design variables

In Figs. 3 and 4, the mean ratings across listeners and talkers are shown for each bit rate B and decay constant β as a function of the probability of error $P(e)$, with α and the level L as parameters.

When no error dissipation mechanism is provided (Figs. 3a and 4a), a slower adaptation, i.e., $\alpha = 0.96875$, makes the system less sensitive to the errors. With a slow dissipation, i.e., $\beta = 255/256$ (Figs. 3b and 4b), the slow adaptation is advantageous only at the higher bit error rate, while with no errors the performance appears to be worse than with fast adaptation. With faster error dissipation, i.e., $\beta = 63/64$ (Figs. 3c and 4c) and slow adaptation, the unbalancing of the load factor between low and high input level is very high and the performance at the low level is very degraded even with no errors. With fast adaptation, i.e., $\alpha = 0.75$, the dynamic range is instead very high; besides, even if the performance under error-free conditions is lower than with slower error dissipation, the system is very insensitive to channel errors.

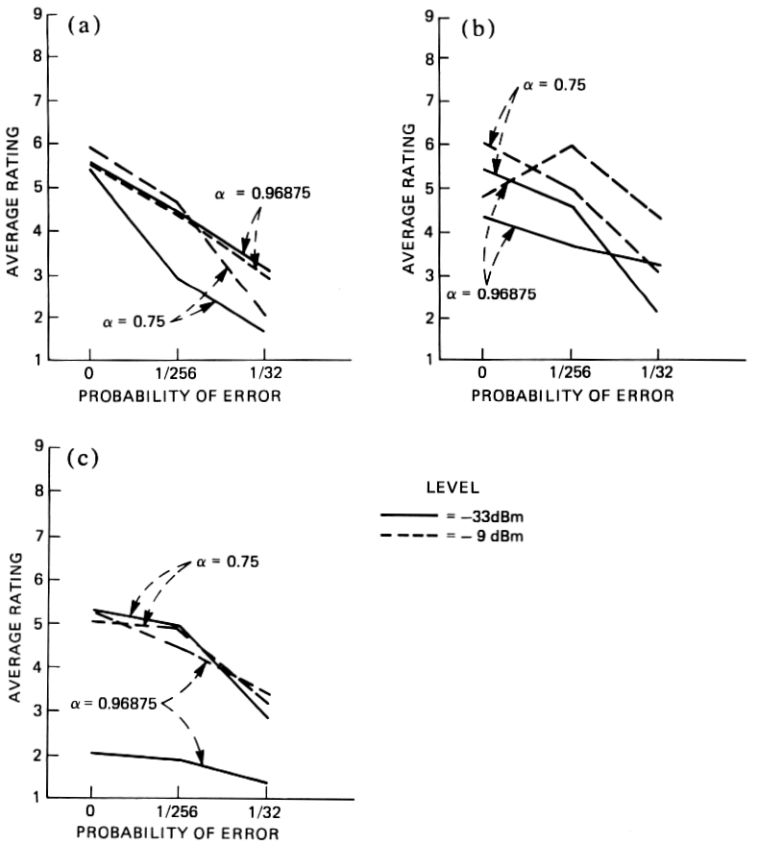

Fig. 3—Mean ratings as a function of probability of error, for $B = 3$: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

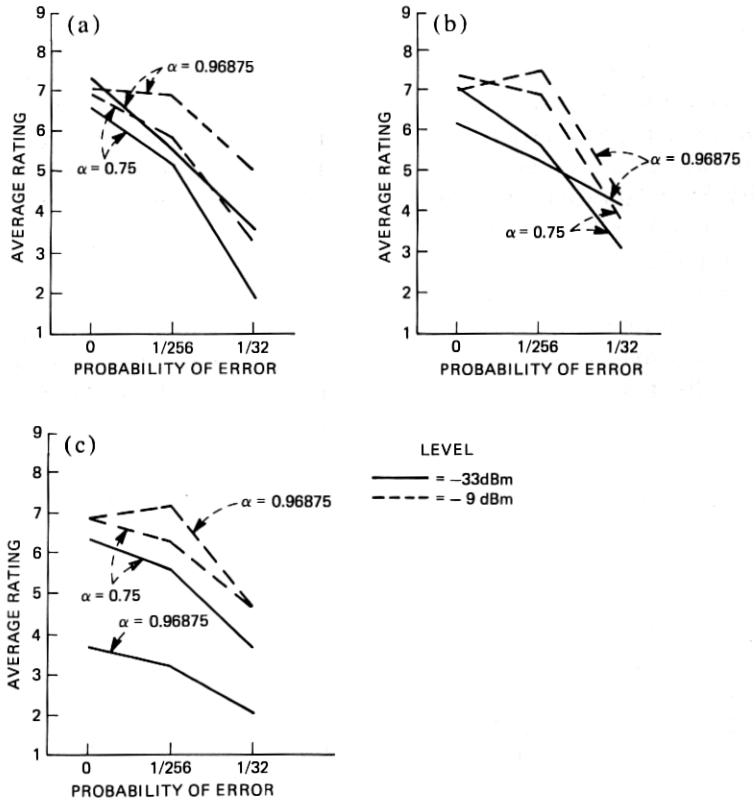

Fig. 4—Mean ratings as a function of probability of error, for $B = 4$: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

6.2.1 *Analysis of variance*

A five-way analysis of variance was performed to evaluate the effect of the experimental variables. The results of the analysis are reported in Table I. The last column reports the P value, that is, the probability that the test statistics assume a value greater than or equal to the computed F ratio, under the null hypothesis, compared with the two significance levels 0.05 and 0.01.

The analysis showed that all the main effects except that due to α are highly significant. The fact that α has no significant effect means that it has a positive effect for certain combinations of parameters and a negative one for others, as shown by the significant interactions.

The interactions between B and α and between B and β were found to be not significant, indicating that α and β have the same effects on the quality of coded speech whether it is a three-bit or a four-bit one. All the other two-way interactions are highly significant. Only the

Table I—Analysis of variance of the mean scores across listeners and talkers

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Ratio	Significance
B	1	31.8928	31.8928	204.7	$P<0.01$
α	1	0.1733	0.1733	1.11	>0.05
β	2	4.9167	2.4583	15.77	<0.01
L	1	19.1570	19.1570	122.96	<0.01
$P(e)$	2	85.5395	43.7697	274.52	<0.01
$B \times \alpha$	1	0.2854	0.2854	1.84	>0.05
$B \times \beta$	2	0.0034	0.0017	0.01	NS
$B \times L$	1	0.7642	0.7642	4.93	<0.01
$B \times P(e)$	2	1.8088	0.9044	5.84	<0.01
$\alpha \times \beta$	2	11.0713	5.5356	35.78	<0.01
$\alpha \times L$	1	3.6933	3.6933	23.87	<0.01
$\alpha \times P(e)$	2	5.0383	2.5191	16.28	<0.01
$\beta \times L$	2	3.8310	1.9155	12.38	<0.01
$\beta \times P(e)$	4	4.7084	1.1772	7.60	<0.01
$L \times P(e)$	2	1.5342	0.7671	4.95	<0.05
$\alpha \times \beta \times L$	2	6.2810	3.1405	20.16	<0.01
Residual	43	6.9980	0.1558		

interaction between L and $P(e)$ is significant at $P < 0.05$ but not at $P < 0.01$. This indicates that the level has an effect almost independent of the probability of error.

The significant interaction between B and L is due to the fact that the difference between the ratings at the two levels is greater on the average for $B = 4$ than for $B = 3$. The significant interaction between B and $P(e)$ is instead due to the fact that the coder with higher bit rate has a greater loss in quality in passing from $P(e) = 0$ to $P(e) = 1/32$.

Of the three-way interactions, only that among α, β, and L is significant. All the other three- and four-way interactions were not significant, and they were pooled in the residual.

VII. QUALITY PREDICTION BY OBJECTIVE MEASURES

To find an objective predictor of the speech quality, linear regression procedures were used. A linear model was chosen not only for its simplicity, but also because in many cases it proved to be adequate to represent the relationship between objective measures and subjective quality. A linear relationship exists, for instance, between the simple SNR and the quality of speech degraded only by the addition of stationary random noise or of speech dependent noise.[18-20] A linear relationship exists also between signal-to-granular noise ratio and probability of overload, and the quality of speech processed by ADPCM coders when no transmission errors are present.[5]

To perform regression analyses, the subjective ratings were averaged across listeners and talkers, and the objective measures were also

averaged, taking the arithmetic mean of the values obtained for each processed sentence. The gain fluctuation σ_g was instead averaged quadratically, taking the square root of the arithmetic mean of the squared values σ_g^2.

Different sets of regression formulas were computed, in which the objective performance measures, like signal-to-noise ratios or spectral distance measures, were used either singly or in combination with the two measures of level mismatching. Figures 5 and 6 show the gain fluctuation and the average gain, both averaged across bit rate, as a function of the probability of error. A few remarks should be made on these figures. Although the two sets of measures have a fairly low correlation of 0.55, the patterns are much alike for low input level and fast adaptation. For the 18 conditions with $L = -33$ dBm and $\alpha = 0.75$, the correlation between G and σ_g is, in fact, 0.96. Therefore, even if in

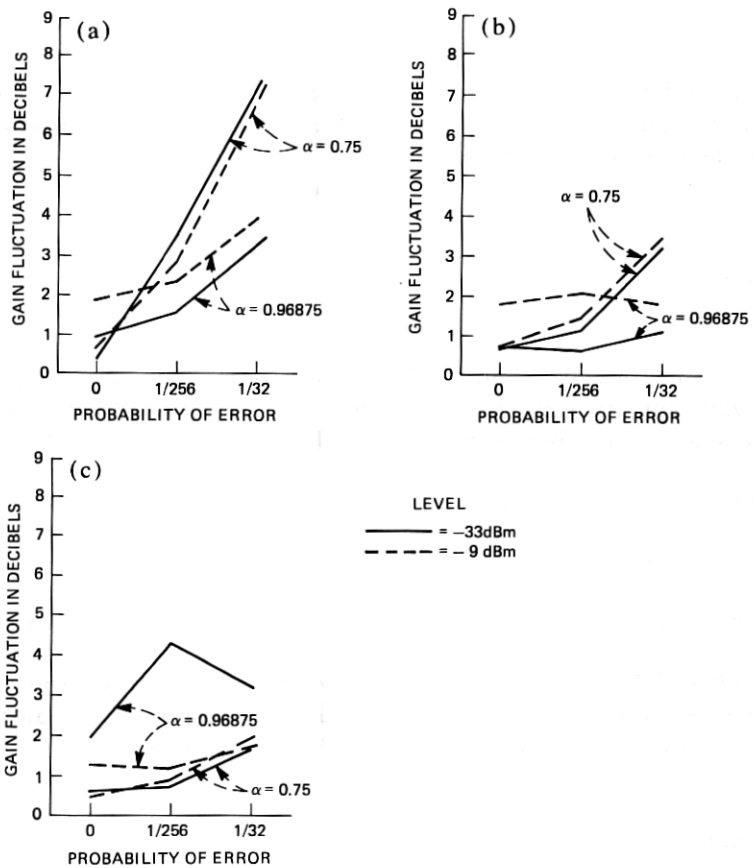

Fig. 5—Gain fluctuation, averaged across bit rate, as a function of probability of error: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

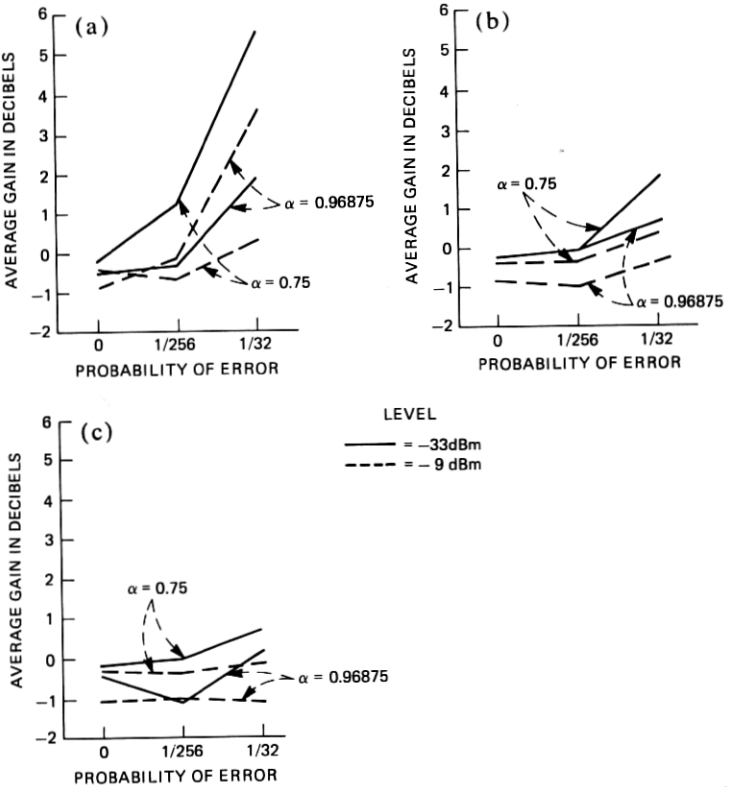

Fig. 6—Average gain, averaged across bit rate, as a function of probability of error: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

a general case the two measures are independent each of the other, in this particular experiment a high value of σ_g is normally associated with a high average gain. This is particularly true if the input level is low, because, owing to the channel errors, the output level tends to be increased. If the input level is high and the quantizer step size is close to its maximum value, the output level is likely to increase only to a smaller extent.

The form of the relationship between the two level mismatching measures and the loss in quality due to the sensitivity of the adaptation algorithms to channel errors was not clear *a priori*, and therefore various nonlinear transformations were tried on those measurement data. No transformation on gain fluctuation proved useful in regression equations, while a compression of the average gain, given by

$$\tilde{G} = G/\sqrt{|G|}, \tag{27}$$

gave better predictions than G, when associated with the other performance measures.

Table II reports the results of the regression analysis. The prediction accuracy is indicated by both the correlation coefficient between the true and predicted subjective scores and the rms prediction error. After each regression analysis, however, a goodness-of-fit test was performed to test normality of prediction errors. The Kolmogorov-Smirnov test[8] was used and in each case the hypothesis of normal distribution was accepted at the 0.20 significance level.

Table II—Formulas for predicting ratings using objective measures

	Formula for Predicting Rating	Correlation	rms error
1	$\hat{R} = 0.156\ \mathrm{SNR} + 2.702$	0.667	1.202
2	$\hat{R} = 0.247\ \mathrm{SNR}_{\mathrm{seg}} + 1.369$	0.873	0.787
3	$\hat{R} = 0.271\ \mathrm{SNR}_{\mathrm{seg}} + 0.260\bar{G} + 1.071$	0.881	0.765
4	$\hat{R} = 0.274\ \mathrm{SNR}_{\mathrm{seg}} + 0.110\sigma_g + 0.767$	0.875	0.780
5	$\hat{R} = 0.302\ \mathrm{SNR}_{\mathrm{seg}} + 0.270\bar{G} + 0.122\sigma_g + 0.393$	0.883	0.756
6	$\hat{R} = 0.336\ \mathrm{SNR}_{\mathrm{com}} - 0.486$	0.911	0.665
7	$\hat{R} = 0.316\ \mathrm{SNR}_{\mathrm{com}} - 0.394\bar{G} - 0.228$	0.935	0.571
8	$\hat{R} = 0.292\ \mathrm{SNR}_{\mathrm{com}} - 0.179\sigma_g + 0.569$	0.923	0.621
9	$\hat{R} = 0.297\ \mathrm{SNR}_{\mathrm{com}} - 0.337\bar{G} - 0.088\sigma_g + 0.254$	0.938	0.561
10	$\hat{R} = 0.389\ \mathrm{SNR}_{\mathrm{max}} - 1.529$	0.913	0.656
11	$\hat{R} = 0.367\ \mathrm{SNR}_{\mathrm{max}} - 0.335\bar{G} - 1.212$	0.931	0.590
12	$\hat{R} = 0.343\ \mathrm{SNR}_{\mathrm{max}} - 0.160\sigma_g - 0.449$	0.923	0.622
13	$\hat{R} = 0.345\ \mathrm{SNR}_{\mathrm{max}} - 0.283\bar{G} - 0.088\sigma_g - 0.668$	0.933	0.581
14	$\hat{R} = 0.267\ \mathrm{SNRF}_1 + 3.906$	0.878	0.773
15	$\hat{R} = 0.251\ \mathrm{SNRF}_1 - 0.295\bar{G} + 3.920$	0.891	0.731
16	$\hat{R} = 0.219\ \mathrm{SNRF}_1 - 0.268\sigma_g + 4.606$	0.909	0.673
17	$\hat{R} = 0.216\ \mathrm{SNRF}_1 - 0.150\bar{G} - 0.235\sigma_g + 4.527$	0.912	0.662
18	$\hat{R} = 0.238\ \mathrm{SNRF}_2 + 3.290$	0.887	0.744
19	$\hat{R} = 0.224\ \mathrm{SNRF}_2 - 0.272\bar{G} + 3.339$	0.899	0.708
20	$\hat{R} = 0.197\ \mathrm{SNRF}_2 - 0.249\sigma_g + 4.050$	0.914	0.656
21	$\hat{R} = 0.195\ \mathrm{SNRF}_2 - 0.140\bar{G} - 0.219\sigma_g + 3.983$	0.916	0.647
22	$\hat{R} = 0.307\ \mathrm{SNRF}_5 + 9.836$	0.855	0.838
23	$\hat{R} = 0.286\ \mathrm{SNRF}_5 - 0.386\bar{G} + 9.499$	0.879	0.770
24	$\hat{R} = 0.244\ \mathrm{SNRF}_5 - 0.323\sigma_g + 9.461$	0.905	0.688
25	$\hat{R} = 0.243\ \mathrm{SNRF}_5 - 0.194\bar{G} - 0.278\sigma_g + 9.318$	0.909	0.671
26	$\hat{R} = -6.514\ D_I + 7.613$	0.797	0.975
27	$\hat{R} = -6.035\ D_I - 0.474\bar{G} + 7.345$	0.837	0.883
28	$\hat{R} = -4.994\ D_I - 0.388\sigma_g + 7.737$	0.878	0.772
29	$\hat{R} = -4.977\ D_I - 0.225\bar{G} - 0.334\sigma_g + 7.592$	0.885	0.751
30	$\hat{R} = -0.519\ D_B + 9.154$	0.826	0.910
31	$\hat{R} = -0.482\ D_B - 0.423\bar{G} + 8.788$	0.856	0.834
32	$\hat{R} = -0.404\ D_B - 0.306\sigma_g + 8.797$	0.868	0.800
33	$\hat{R} = -0.405\ D_B - 0.262\bar{G} - 0.241\sigma_g + 8.647$	0.878	0.773
34	$\hat{R} = 0.436\ \mathrm{SDR}_1 - 1.268$	0.850	0.849
35	$\hat{R} = 0.407\ \mathrm{SDR}_1 - 0.442\bar{G} - 0.923$	0.883	0.757
36	$\hat{R} = 0.348\ \mathrm{SDR}_1 - 0.281\sigma_g + 0.518$	0.885	0.751
37	$\hat{R} = 0.353\ \mathrm{SDR}_1 - 0.304\bar{G} - 0.203\sigma_g + 0.254$	0.897	0.712
38	$\hat{R} = 0.947\ \mathrm{SDR}_2 - 0.249$	0.874	0.784
39	$\hat{R} = 0.893\ \mathrm{SDR}_2 - 0.513\bar{G} - 0.028$	0.917	0.642
40	$\hat{R} = 0.766\ \mathrm{SDR}_2 - 0.306\sigma_g + 1.331$	0.918	0.641
41	$\hat{R} = 0.785\ \mathrm{SDR}_2 - 0.353\bar{G} - 0.212\sigma_g + 0.998$	0.934	0.578

Among the objective performance measures taken singly, the best one turns out to be $\text{SNR}_{\max}$, with a correlation coefficient of 0.913 and an rms error of 0.656 [formula 10 in Table II]. The compensated signal-to-noise ratio SNR_{com} gives almost the same results, while all the other measures achieve a correlation lower than 0.9. In particular, the conventional, long-term, signal-to-noise ratio has a correlation of only 0.667 and an rms prediction error almost double that of $\text{SNR}_{\max}$. The log likelihood ratio D_I is the second-worst predictor when used singly, with a correlation of only 0.797.

When the two measures of level mismatching, i.e., the average gain and the gain fluctuation, are included in the quality prediction formulas, the prediction accuracy is significantly improved, the rms prediction error having a 16-percent decrease on the average. The smallest improvement is displayed by SNR_{seg}. Among all the other measures, SNR_{com} gives the best prediction when combined with $\tilde{G}$ and σ_g (formula 9), with a correlation of 0.938 and an rms error of 0.561, about one-quarter of a category. Formulas 13 and 41, which use $\text{SNR}_{\max}$ and SDR_2, are almost as good as formula 9. The frequency-weighted SNRs also give a fairly good prediction, with correlations over 0.9 and the remaining frequency domain measures, D_I, D_B, and SDR_1 give a slightly poorer prediction.

VIII. DISCUSSION

8.1 *Effects of coder design parameters*

The subjective data have displayed complicated interactions among all the experimental design variables, the strongest interaction being the one between the adaptation constant α and the decay constant β. In fact, each of these two parameters affects different phenomena:

(*i*) The *dynamic range* is reduced when β decreases from unity, but this reduction does not seem to be perceptible for any β if $\alpha = 0.75$. If $\alpha = 0.96875$ and $\beta = 255/256$, a certain reduction in the dynamic range begins to be perceived, producing a loss in quality at the low level of about 1.5 points with respect to the high level. When $\alpha = 0.96875$ and $\beta = 63/64$, the dynamic range is reduced still further, and the loss in quality of the low level with respect to the high one is very large, the average score dropping down from 5.3 to 2.4.

(*ii*) The *effect of the errors* is smaller when α increases or β decreases. For instance, with $\alpha = 0.75$, the loss in quality passing from $P(e) = 0$ to $P(e) = 1/256$ averages 0.49 when $\beta = 63/64$, while it averages 1.59 for $\beta = 1$ and 0.93 for $\beta = 255/256$.

(*iii*) The difference in the effect of the errors between the two input levels is higher for faster adaptation. For instance, passing again from $P(e) = 0$ to $P(e) = 1/256$, the difference between the losses at the two levels averages 1.0 when $\alpha = 0.75$, while it averages 0.45 when $\alpha = 0.96875$.

8.2 Optimum coders

Given a fixed number of bits per sample, a combination of decay constant β and adaptation constant α provides the best output quality for a given probability of error.

In the case of error-free transmission, optimum quality should be attained with no error dissipation mechanism, i.e., $\beta = 1$ which produces theoretically infinite dynamic range. The parameter α is not very critical in that case.[5]

When the coder operates under noisy conditions and the probability of error is in the range of the values encountered in a normal telephone connection or even higher than that (as is the case of $P(e) = 1/256$), a very slow error dissipation associated with fast adaptation provides good robustness to channel errors, without impairing the dynamic range. Actually, in this experiment, the combination $\beta = 255/256$ and $\alpha = 0.75$ provided optimum performance even under error-free conditions.

If the probability of error is as great as 1/32, more typical of mobile radio communications, the best compromise between dynamic range and error sensitivity is obtained by a slow adaptation constant, combined again with a slow error dissipation rate. However the use of faster error dissipation and faster adaptation could be almost as good for this very high error rate.

8.3 Objective measures

One aim of the experiment was to examine a certain number of objective measures of coder performance and to compare them in the light of the actual subjective quality ratings, obtained under very different conditons. With the results of correlation and regression analyses reported in Table II, it is possible to observe strengths and weaknesses of the different measures and to derive general indications on which are the desirable properties of an objective quality measure.

A first indication that emerges from the experimental data is that the conventional long-term SNR is a very poor indicator of the quality of ADPCM coders under noisy channel conditions; this confirms the results obtained with PCM[18] and ADPCM coders[5] in the case of error-free transmission. Therefore, the use of SNR can be completely misleading, when comparing different coders operating under noisy conditions. A noticeable improvement in prediction accuracy is obtained simply by

measuring signal-to-noise ratio segmentally. Being time-segmental is a necessary property of any successful objective quality measure of coded speech.

Table II demonstrates also that, when a coder incorporating an adaptive quantizer is operating under noisy channel conditions, an objective performance measure must not be sensitive to changes or fluctuations of the output speech level. These fluctuations can be measured separately and the value obtained can be combined with the performance measure to improve the accuracy of the subjective quality prediction. It should be noticed, for instance, that $\bar{G}$ and σ_g have a positive coefficient when combined with SNR_{seg} (formula 5 in Table II). This indicates that the level mismatching is weighted too much in SNR_{seg}, which did not incorporate any level compensation.

An important consideration on the subject of objective quality measures is that results from recent experiments[7,9] indicate that frequency-weighted signal-to-noise ratios improve the prediction accuracy, especially when largely different noise spectra are produced by the coders.[7,9] In this experiment, actually the frequency-weighted SNRs did not predict the subjective ratings as accurately as the simpler level-compensated SNRs, namely, SNR_{com} and SNR_{max}. This may depend on the fact that the level compensation is effected by minimizing the rms error on the whole bandwidth; this fact may worsen the measure in same band. If this is the case, it would be a weakness of the frequency-weighted SNR's when measuring coder performance in the presence of channel errors.

A final remark on frequency domain measures: These measures are more general than time domain ones because they are insensitive to short delays or to phase distortion.[9,14] Therefore, they are more easily applicable to the test of coders whose input and output signals are in analog form or which include digital filters. On the other hand, the performance measures used here incorporate a spectral noise weighting that cannot be directly controlled. However, it is encouraging to see that the newly defined spectral signal-to-distortion ratio, SDR_2, provides a very good prediction of subjective ratings when combined with the level mismatching measures.

More work is needed in the field of objective prediction of coder quality. In particular, frequency weighted SNRs should be evaluated more carefully, to derive the appropriate frequency weighting mechanism. In addition, the difference in behavior of the various spectral distortion measures need to be analyzed in more depth.

8.3.1 Estimation of the subjective effect of level mismatching

8.3.1.1 A Physical Interpretation of Prediction Formulas. The results of Table II lend themselves to a nice interpretation. The quality of an

adaptive coder operating in a noisy channel environment (high probability of error) may be considered as composed of two terms: (*i*) "intrinsic" goodness of the speech reproduction, which takes into account the noise due to the coding and to the errors, but not the level mismatching, (*ii*) loss in quality due to the level variations caused by the sensitivity of the adaptation algorithm to channel errors. In formulas, we can write

$$\hat{R} = \hat{R}_I - \hat{R}_L. \tag{28}$$

$\hat{R}_I$ can basically be estimated by any of the performance measures which incorporate gain compensation or, in any case, which are not sensitive to alterations in the output signal level. $\hat{R}_L$ is instead estimated by a linear combination of the gain fluctuation and the average gain, modified according to eq. (27).

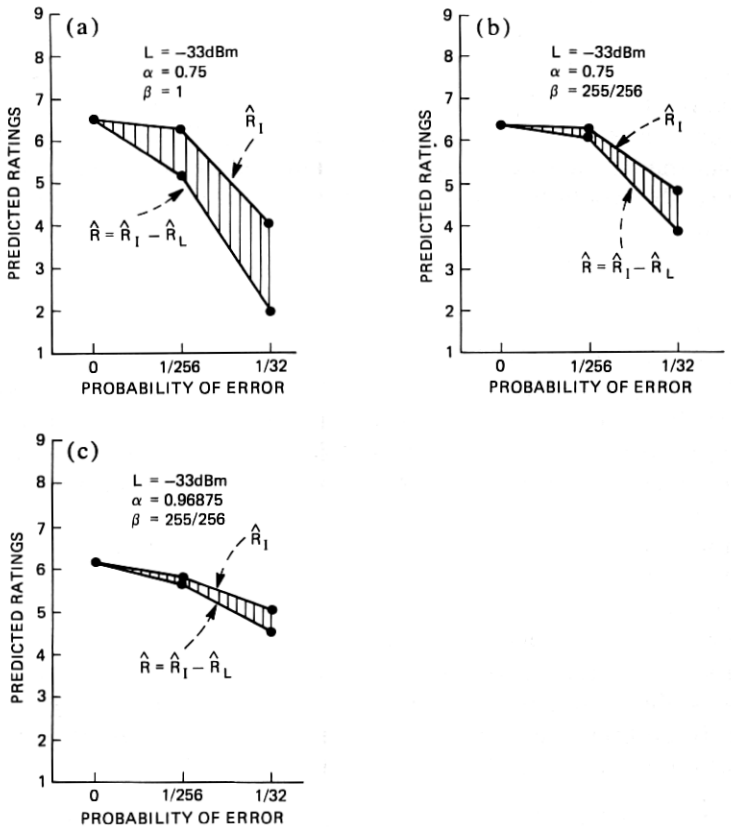

Fig. 7—Predicted overall rating $\hat{R}$ and "intrinsic" goodness $\hat{R}_I$, as a function of probability of error, for 4-bit ADPCM and different combinations of design parameters: (a) $\alpha = 0.75$, $\beta = 1$. (b) $\alpha = 0.75$, $\beta = 255/256$. (c) $\alpha = 0.96875$, $\beta = 255/256$.

8.3.1.2 An Example. In the light of the interpretation given in the previous section, it is possible to give at least a qualitative answer to the question on which is the subjective effect of level mismatching. Figure 7 shows the predicted overall rating $\hat{R}$ and the "intrinsic" goodness $\hat{R}_I$ as a function of the probability of error, for the 4-bit ADPCM coder with low level input and three different combinations of design parameters. $\hat{R}$ was computed according to formula 41 in Table II, while $\hat{R}_I$ was computed discarding the terms involving $\hat{G}$ and σ_g from the same formula:

$$\hat{R}_I = 0.785\ SDR_2 + 0.998.$$

In the case of fast adaptation and absence of error recovery (Fig. 7a), the loss in quality due to level mismatching can be estimated as half category (1 point) for the intermediate error rate and 1 category (2 points) for the high error rate. In the case of fast adaptation and slow error recovery (Fig. 7b), the loss $\hat{R}_L$ is instead reduced to about only half category (1 point) for the high error rate.

Finally, when slow adaptation and slow error recovery are used, simultaneously (Fig. 7c), the loss due to level mismatching can be estimated as only about 0.4 point, i.e., less than a quarter of a category.

IX. ACKNOWLEDGMENTS

The author wishes to thank B. J. McDermott, B. H. Bharucha, R. E. Crochiere, and J. M. Tribolet for several useful discussions and for having provided subroutines for computing some of the objective measures. The author thanks also J. Coker for recruiting the subjects and running the listening tests.

REFERENCES

1. D. J. Goodman and R. M. Wilkinson, "A Robust Adaptive Quantizer," IEEE Trans. Commun., *COM-23,* No. 11 (November 1975), pp. 1362–1365.
2. C. Scagliola, "An Adaptive Speech Coder with Channel Error Recovery," International Conference on Communications, Chicago, Ill., June 1977.
3. D. Mitra, "An Almost Linear Relationship Between the Step-Size Behavior and the Input Signal Intensity in Robust Adaptive Quantization," IEEE Trans. Commun., *COM-27,* No. 3 (March 1979).
4. P. Castellino, G. Modena, L. Nebbia, and C. Scagliola, "Bit Rate Reduction by Automatic Adaptation of Quantizer Step-size in DPCM Systems," International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.
5. B. J. McDermott, C. Scagliola, and D. J. Goodman, "Perceptual and Objective Evaluation of Speech Processed by Adaptive Differential PCM," B.S.T.J., *57,* No. 5 (May–June 1978), pp. 1597–1618.
6. P. Noll, "Adaptive Quantizing in Speech Coding Systems," International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.
7. R. E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A Study of Objective Measures of Speech Waveform Coders," International Zurich Seminar, Zurich, Switzerland, March 1978.
8. A. A. Afifi and S. P. Azen, *Statistical Analysis a Computer Oriented Approach,* New York: Academic Press, 1972.

9. J. M. Tribolet, P. Noll, B. J. McDermott, and R. E. Crochiere, "Complexity vs. Quality for Speech Waveform Coders," IEEE International Conference on Acoustics Speech and Signal Processing, Tulsa, Oklahoma, April 1978.
10. N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Amer. *19* (January 1947), pp. 90–119.
11. J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, New York: Springer Verlag, 1976.
12. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust. Speech and Sig. Proc., *ASSP-23*, (February 1975), pp. 67–72.
13. A. H. Gray and J. D. Markel, "Distance Measures for Speech Processing," IEEE Trans. Acoust. Speech Sig. Proc., *ASSP-24*, (October 1976), pp. 380–391.
14. B. H. Bharucha, "An Objective Measure of Codec Speech Quality," unpublished paper, 1976.
15. N. S. Jayant, "Adaptive Quantization with One-Word Memory," B.S.T.J., *52*, No. 7, (September 1973), pp. 1119–1144.
16. P. Slater, "Analysis of Personal Preferences," Brit. Journal of Statistical Psychology, *13* (November 1960), pp. 119–135.
17. J. D. Carroll, "Individual Differences and Multidimensional Scaling," in *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. I, Shepard, Romney, Nerlove (Eds.), New York: Seminar Press, 1972, pp. 105–155.
18. D. J. Goodman, B. J. McDermott, and L. H. Nakatani, "Subjective Evaluation of PCM Coded Speech," B.S.T.J., *55*, No. 8, (October 1976), pp. 1087–1109.
19. D. L. Richards, *Telecommunications by Speech*, London: Butterworths, 1973, Ch. 4.
20. L. Nebbia and P. Usai, "Influence of Some Types of Noise on Telephone Digital Transmissions," Symposium, "Speech Intelligibility," Liege, November 1973.