THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 8, October 1978.
Printed in U. S. A.

An Analysis of 16 kb/s Sub-Band Coder Performance: Dynamic Range, Tandem Connections, and Channel Errors

By R. E. CROCHIERE

(Manuscript received December 26, 1977)

In this paper, we examine the performance of sub-band encoding under a number of constraints which can exist in practical digital communications systems. In particular, we investigate the effects of varying input signal levels, tandem connections, and channel errors on the performance of sub-band coders. A coder bit rate of 16 kb/s is used in all the simulations. The dynamic range performance is evaluated for a 50-dB range of input signal levels. Tandem connections of up to four sub-band coders in tandem are examined. Finally, the effects of random channel errors on the performance of sub-band coders is examined for bit error probabilities of up to 10 percent. A robust coder design with partial bit error protection is also proposed for use in very high channel error environments.

Three different performance measures were used in these simulations, the conventional signal-to-noise ratio, a segmental signal-to-noise ratio, and an LPC distance measure. By comparing the results of these various performance measures and from informal assessments of subjective quality, we gain some new insights into the advantages and disadvantages of these measures in terms of their usefulness in predicting coder quality.

I. INTRODUCTION

Sub-band coding has recently been proposed as a technique for obtaining relatively good quality digital speech at a bit rate of 16 kb/s.[1–3,16] This quality is subjectively comparable to that of 24 kb/s ADPCM (adaptive differential PCM),[1,2] and it is generally acceptable for some types of digital communications applications where relatively low transmission rates are required.

In a practical communications system, the quality of digital speech can be affected and degraded by a number of factors. The input speech

levels to the coders may vary over a relatively broad range, and the coders may not necessarily be driven at their optimum input levels. In a communications network, digital coders may be linked with other types of digital or analog systems, and it is possible that several tandem connections of the same type of coder may occur in a given transmission path. Finally, channel errors may occur in a digital system, and it is important to understand how the performance of digital coders are affected by these errors.

In this paper, we present the results of a series of experiments designed to assess the performance and robustness of 16-kb/s sub-band coding in practical communications environments. The dynamic range of the coder is evaluated over a 50-dB range of input signal levels. Tandem connections of sub-band coders of up to four links are examined. The effect of channel errors is examined for error probabilities as high as 0.1. Finally, several methods for improving the robustness of sub-band coding in the presence of channel errors are examined.

II. THE SUB-BAND CODER

Sub-band coding is a waveform coding technique in which the speech band is partitioned into typically four or five sub-bands by bandpass filters. Each sub-band is then lowpass-translated to dc, sampled at its Nyquist rate, and then digitally encoded using adaptive PCM (APCM) encoding. By this process of dividing the speech band into sub-bands, each sub-band can be preferentially encoded according to perceptual criteria for that band. On reconstruction, sub-band signals are decoded and bandpass-translated back to their original bands. They are then summed to give a replica of the original speech signal.

A particularly attractive implementation of the sub-band coder, in terms of hardware, is based on an integer band sampling approach.[1,2] This approach uses the samplers both for discretizing the sub-band signals as well as for doing the lowpass and bandpass translations, i.e., the modulation is achieved with an impulse train instead of with sine and cosine signals. This implementation is illustrated in Fig. 1. Bandpass filters BP_1 to BP_N in the transmitter and receiver serve to partition the input speech into sub-bands. The coders and decoders encode the sub-band signals and the multiplexer combines these digital signals into a single bit stream for transmission over the digital channel. In addition, the multiplexer inserts synchronizing bits into the bit stream for the purpose of synchronizing the operation of the transmitter and receiver.

Table I shows the choice of bands and bit allocations used in the 16-kb/s coder. The coder is a 5-band design which was proposed in Ref. 2. Column 2 shows the frequency range covered by each sub-band. The bit allocation refers to the number of bits/sample used by the coders in each

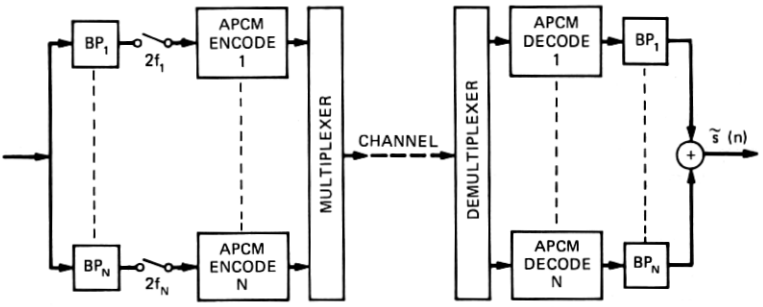


Fig. 1—An integer-band sampling implementation of the sub-band coder.

Table I — 16 kb/s 5-band sub-band coder

Band	Band Edges (Hz)	Sampling Freq (Hz)	Min Step-Size (dB)	Bit Allocation	Kb/s
1	178–356	356	(Ref)	4	1.42
2	296–593	593	0	4	2.37
3	533–1067	1067	0	3	3.20
4	1067–2133	2133	−3	2	4.27
5	2133–3200	2133	−8	2	4.27
SYNC	—			—	0.47
					16.00

sub-band. As seen from the table, more accuracy is allowed for encoding the lower bands for reasons explained in Ref. 2.

The frequency range of the coder extends from 200 to 3200 Hz. A plot of this frequency response is shown in Fig. 2. As seen in this figure, two small notches appear in the frequency response at 1067 and 2133 Hz.

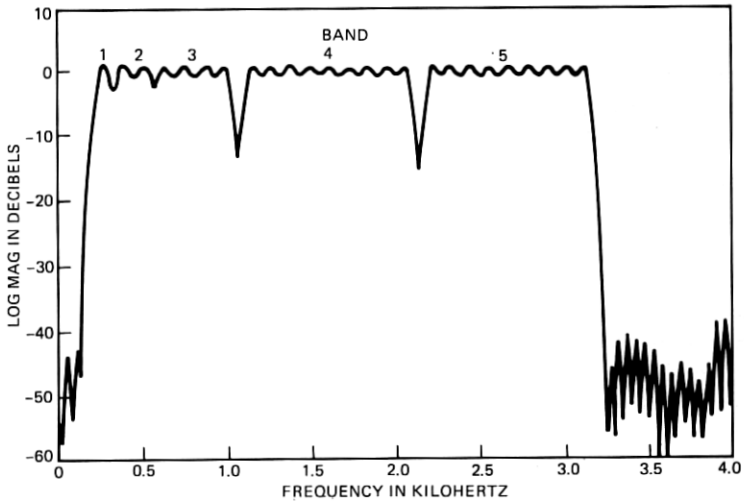


Fig. 2—Frequency response of the 5-band coder in Table I.

These notches are due to the transition bands of the filters in bands 4 and 5. Subjectively, these notches are not very perceptible. Bands 1 to 3 are overlapped to avoid such notches at lower frequencies. The filters are sharp cutoff, 200 tap, FIR filters.

Column 4 in Table I refers to the ratio of minimum allowed step sizes of the APCM coders (expressed in decibels), with the minimum step size of band 1 being the reference. This choice of minimum step sizes is different than that suggested in Ref. 2 and was found to give a better matching of the dynamic ranges of the sub-bands.

In Section VI, we propose an alternate design for a 4-band coder which can be used in a high channel error environment.

III. PERFORMANCE MEASURES

3.1 Conventional s/n

Several objective measures were used to evaluate the performance of the sub-band coder. In this section we briefly define each of these measures.

The most commonly used measure of performance of digital coders has been the conventional signal-to-noise ratio (s/n) evaluated over an utterance of speech. The speech power is defined as

$$\hat{s} = \sum_m x^2(m) \tag{1}$$

and the noise power is defined as

$$\hat{n} = \sum_m (x(m) - y(m))^2, \tag{2}$$

where $x(m)$ and $y(m)$ are the input and output signals of the coder, respectively, and the summations in (1) and (2) are taken over the entire speech utterance. The conventional s/n is then defined as

$$s/n = 10 \log(\hat{s}/\hat{n}). \tag{3}$$

In measuring the input and output signals to the sub-band coder, it is generally desirable to compensate for the effects of filtering in the coder, particularly effects of group delay. This is done by the circuit arrangement shown in Fig. 3. The input speech signal $s(m)$ is sub-band-coded to form the output speech signal $y(m)$. It is also filtered with the same filters used in the sub-band coder to generate a compensated reference signal $x(m)$ which is used as the input signal in (1) and (2). Thus, the s/n ratio defined here is strictly a measure of coder distortions and is not affected by bandlimiting or group delay in the coder.

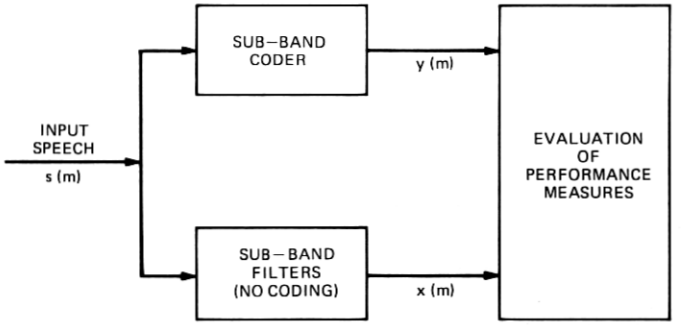


Fig. 3—Circuit for evaluating signal-to-noise ratios of the sub-band coder.

3.2 Segmental s/n

While the s/n measure is perhaps the most widely used criterion in measuring coder distortion, it has also long been known that it does not correlate well with subjective performance.[2,4,5] Another definition of signal-to-noise ratio, however, recently proposed by Noll,[6,7] does appear to correlate better with subjective performance. This measure is based on s/n measurements made over short segments of speech which are typically about 20 ms in duration. An average over all of the segments in the speech utterance is then taken to obtain a composite measure of performance for the entire utterance. If $(s/n)_i$ corresponds to the signal-to-noise ratio in decibels for a segment, i (computed in the same manner as in (3)), the segmental s/n, (SEG), is then defined as

$$\text{SEG} = \frac{1}{N} \sum_{i=1}^{N} (s/n)_i, \tag{4}$$

where it is assumed that there are N 20 ms segments in the speech utterance.

Several problems occur in this definition of segmental s/n when regions of silence exist in the speech utterance. In segments where the input signal $x(n)$ is essentially zero, any slight noise will give rise to large negative $(s/n)_i$ ratios, and these segments may unduly dominate the average in (4). To prevent this anomaly, we first identify those segments which correspond to silence and exclude them from the average in (4). This is achieved by means of a simple threshold. Let $\hat{s}_i$ represent the speech energy in a segment, i, so that

$$\hat{s}_i = \frac{1}{K} \sum_{m=1}^{K} x^2(m), \tag{5}$$

where K corresponds to the number of speech samples in the segment. Then the segment will be included in the computation of SEG in (4) if its energy exceeds a threshold σ_t^2; that is, if $\hat{s}_i > \sigma_t^2$. If it does not exceed this threshold, it is not included in the average in (4). Furthermore, to

prevent any one segment from dominating the average, we also limit the value of $(s/n)_i$ to a range of -10 to $+80$ dB. That is, $-10 \leq (s/n)_i \leq 80$ dB.

It remains to determine the threshold σ_t^2 for the speech/silence decision. To establish this threshold, we coded a speech utterance composed of three concatonated sentences in which there was about 30 percent silence in the entire utterance. Figure 4 shows a plot of SEG as a function of σ_t. The threshold σ_t was varied from 0 to 32767 corresponding to the range of signal values representable in the 16-bit integer word length of the computer. The dashed line in Fig. 4 shows the number of segments included in the SEG measure. As seen in the figure when the threshold σ_t was below 3, virtually all the silence intervals were included in the SEG measure. The low $(s/n)_i$ in these regions essentially dominated the sum, resulting in values of SEG of about 1 dB. When the threshold, σ_t, was raised to a value of 10, nearly all the silence regions were eliminated from the measure, and the value of SEG rose to about 9 dB. At a threshold of $\sigma_t = 30$, the value of SEG reached a plateau of about 10.3 dB. Therefore, the threshold was chosen to be $\sigma_t = 30$ for all SEG measurements. The conventional s/n for this same utterance was 10.8 dB.

3.3 LPC distance measure

A third performance measure that was used is the LPC distance measure proposed by Itakura.[8,9] This measure is based on an all-pole model of speech of the form

$$s(n) = \sum_{m=1}^{P} a(m)s(n-m) + Gu(n), \tag{6}$$

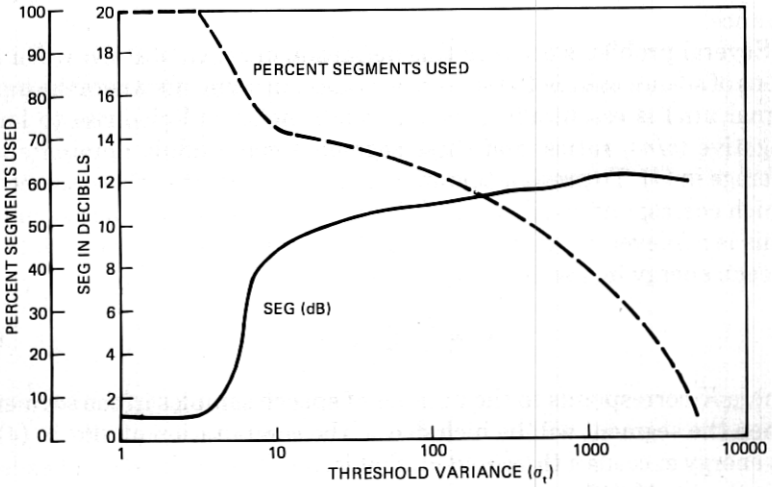


Fig. 4—Segmental s/n as a function of the speech/silence threshold σ_t.

where $s(n)$ is the sampled speech signal, $a(m)(m = 1, \ldots p)$ are the coefficients of an all-pole filter which models the resonances of the speech production mechanism, p is the number of modeled poles, G is the gain of the filter, and $u(n)$ is the excitation source for the all-pole filter.

The LPC distance measure for a segment, K, of speech (typically 20 ms in duration) is then defined as

$$d_{1k} = \log\left[\frac{\mathbf{a}_k V_b \mathbf{a}_k^t}{\mathbf{b}_k V_b \mathbf{b}_k^t}\right], \tag{7}$$

where

$\mathbf{a}_k$ = LPC coefficient vector $(1, a_1, \ldots a_n)$ measured for the kth frame of the original (reference) speech signal $s(n)$,

$\mathbf{b}_k$ = LPC coefficient vector measured for the kth frame of the coded (or processed) speech is $s'(n)$,

and V_b is the speech correlation matrix of $s'(n)$ whose elements V_{ij} are defined as

$$V_{bij} = v(|i - j|) = \sum_{n=1}^{N-|i-j|} s'(n)s'(n + |i - j|), \tag{8}$$

where $s'(n)$ is the processed speech signal. The overall distance measure for the speech utterance is then determined as the average over the N segments in the utterance,

$$\bar{d}_1 = \frac{1}{N}\sum_{k=1}^{N} d_{1k} \tag{9}$$

By interchanging the roles of the reference and processed speech, a second distance measure can similarly be defined in the form[10]

$$d_{2k} = \log\left[\frac{\mathbf{b}_k V_a \mathbf{b}_k^t}{\mathbf{a}_k V_a \mathbf{a}_k^t}\right] \tag{10}$$

and

$$\bar{d}_2 = \frac{1}{N}\sum_{k=1}^{N} d_{2k}. \tag{11}$$

An average distance measure can now be defined as

$$\bar{d} = \frac{1}{2}(\bar{d}_1 + \bar{d}_2) \tag{12}$$

which is the measure used in this paper. The LPC distance measure $\bar{d}$ is basically a measure of dissimilarity between the spectra of the processed and unprocessed speech. It is therefore useful in measuring the spectral distortion introduced by the coder. If the processed and unprocessed utterances are identical, then the distance $\bar{d}$ is zero. For small differences

between the processed and unprocessed signals, $\bar{d}$ will typically have a small positive value less than 1. Large values of $\bar{d}$, greater than 1, generally indicate significant spectral differences between the processed and unprocessed signals.

IV. DYNAMIC RANGE

The dynamic range of the sub-band coder is determined by the ratio of maximum to minimum allowed step-sizes in the APCM quantizers. In this work, we used a ratio of $\Delta_{max}/\Delta_{min} = 128$, which leads to an effective dynamic range of about 30 to 35 dB over which the quality remains relatively constant. Typically, if the $\Delta_{max}/\Delta_{min}$ ratio is increased, the dynamic range of the coder increases (within limits) by about 6 dB per doubling of the ratio.

To improve the performance of the sub-band coder at the low end of its dynamic range we also used a mid-rise/mid-tread switch in the APCM coders.[11] This extended the useful range of the coders by about 6 dB and eliminated the low-level tones and idle channel noise generated by the APCM coders.

Figure 5 shows the results of the s/n and the SEG measures for the coder for input signal levels over a range of about 50 dB. The measurements were made for a speech segment composed of two sentences, "High altitude jets whiz past screaming" and "A lathe is a big tool," spoken by two different male speakers. As seen in the figure, the conventional s/n is high in the granular noise region of the coder (input levels less than −10 dB) and drops rapidly in the over-load region (input levels greater than 0 dB). It is controlled primarily by the high-energy region in the speech utterance. At low input levels, the s/n measure is typically too large. It fails to account for the low-level granular noise of the coders,

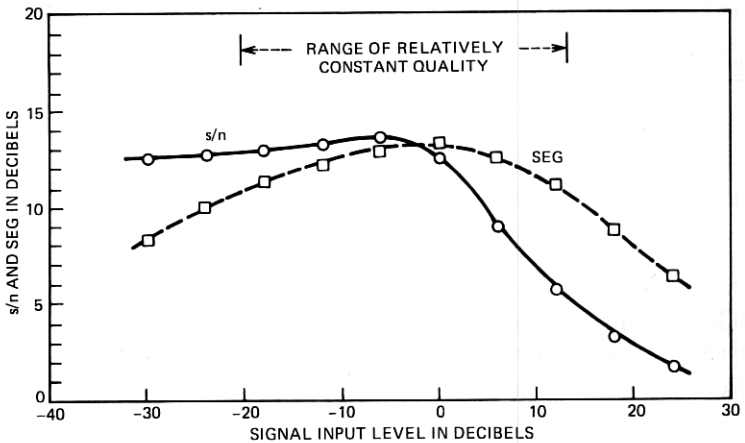


Fig. 5—Dynamic range of the sub-band coder: s/n and SEG measurements.

which can be subjectively disturbing. In the overload region, the s/n measure overemphasizes the clipping in the high-energy parts of the coded speech.

The SEG measure agrees much better with our informal observations of quality. It is more sensitive to the granular noise at low levels and less sensitive to the overload in very loud parts of the speech. It essentially treats all time segments on an equal basis and does not favor high or low parts of the utterance.

In examining the performance of the sub-band coder, it is also instructive to observe the performance of the individual APCM coders used in the sub-bands. Results for s/n and SEG measurements for the 4-, 3-, and 2-bit coders are presented in Fig. 6, where the results of the 4-bit coder are obtained from measurements of sub-bands 1 and 2, the results of the 3-bit coder are obtained from sub-band 3, and results for the 2-bit coder are obtained from sub-bands 4 and 5. The solid lines refer to s/n measurements, and the dashed lines refer to SEG measurements. An important consideration in the design of the sub-band coder is that the dynamic range in each of the sub-bands be aligned so that, at the optimum input level, each sub-band is operating at its peak performance. This alignment is determined by the choice of maximum and minimum step sizes in the coders in each sub-band. The relative values of minimum step sizes (expressed in decibels) that we used are given in column 4 of Table I, which resulted in the alignment of the dynamic ranges shown in Fig. 6.

Figure 7 shows the results of the LPC distance measurements on the sub-band coder. The measure was made between $x(m)$ and $y(m)$ according to the arrangement in Fig. 3 and, therefore, does not take into account the spectral distortions due to the filters or notches between the bands. At the optimum input level, the value of the LPC distance is 0.12.

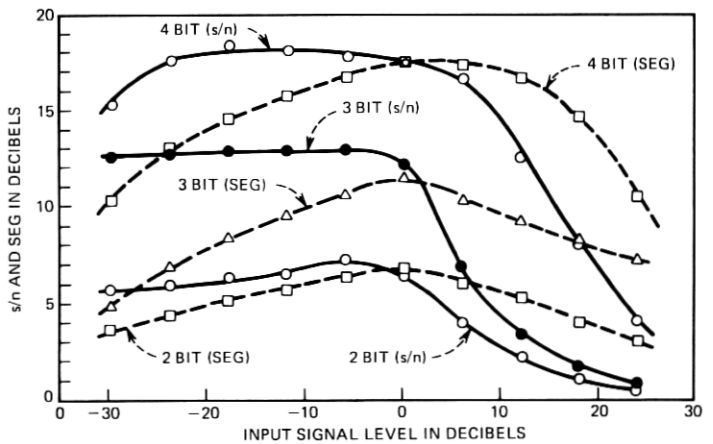


Fig. 6—Dynamic range of the individual APCM coders in the sub-bands.

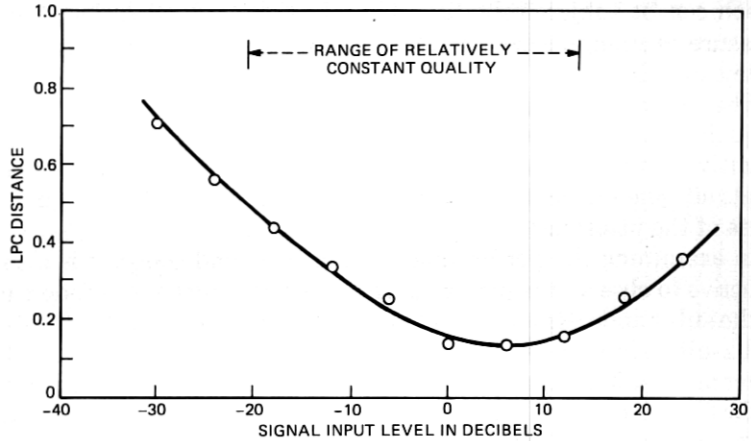


Fig. 7—LPC distance as a function of the input signal level.

For low input levels, i.e., in the granular noise region, it goes up to 0.56 at a −24-dB input level. At high input levels (+24 dB) in the overload or clipping region of the coder, the LPC distance goes up to 0.36. Thus, the spectral distortion is typically greater in the granular noise region than in the overload region of the coder.

To determine the effect of the bandpass filters and the notches in frequency response of the filter, a second LPC distance measurement was made across the sub-band coder according to the arrangement shown in Fig. 8. The input speech was delayed by a flat delay equal to the delay of the filters. This reference signal and the output of the coder were then both filtered with a 200- to 3200-Hz bandpass filter giving the signals $\hat{x}(m)$ and $\hat{y}(m)$. The purpose of the bandpass filters on the outputs is so that the spectral differences outside of the 200- to 3200-Hz band of interest do not affect the LPC distance measure.

When we measured the LPC distance between $\hat{x}(m)$ and $\hat{y}(m)$ by this method, we obtained a distance of 0.58 for the sub-band coder (operating at the optimum input level of 0 dB). We then removed the quantizers

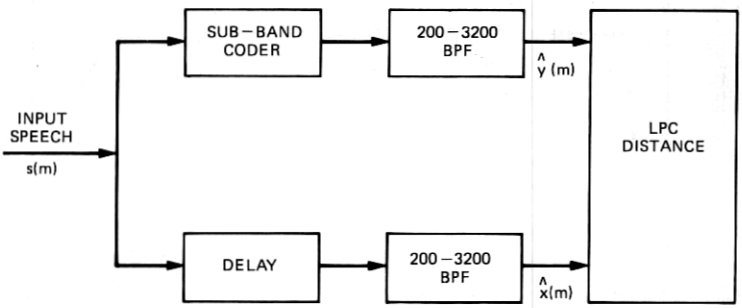


Fig. 8—Circuit arrangement for measuring the total LPC distance of the sub-band coder (including the effect of the filters) in a 200- to 3200-Hz bandwidth.

from the coder and measured only the effects of the sub-band filtering. This resulted in a distance of 0.53 between $\hat{x}(m)$ and $\hat{y}(m)$. This distance is strictly due to the passband ripples, sharp transition bands, and notches in the frequency response of the coder as seen in Fig. 2. Although the contribution to the LPC distance due to the filters was greater than that due to quantization noise, their subjective effects cannot necessarily be weighted in the same way. Subjectively, the effects of the sharp cutoff filters and the notches do not strongly affect the quality or intelligibility of the coder.

V. TANDEM CONNECTIONS

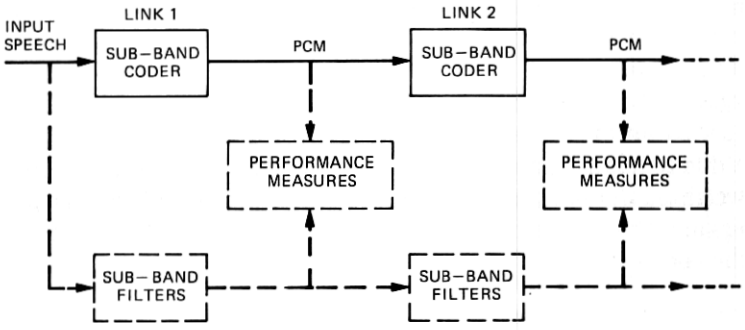
Computer simulations of tandem connections of sub-band coders were made for up to four coders in tandem. Two types of tandem connections were considered in this experiment. The first type of tandem link consists of a sub-band coder followed by 16-bit linear PCM as shown in Fig. 9a. A parallel link of sub-band filters, shown in dotted lines, was also simulated in order to generate reference signals to facilitate s/n and SEG measurements.

In the second type of link shown in Fig. 9b, we simulated the effects of a digital-to-analog conversion and a resampling of the signals between each coder. This simulation was achieved by means of an all-pass filter which was inserted between the tandem links. Again, a reference link of sub-band filters was also simulated to facilitate signal-to-noise ratio measurements. The effect of the all-pass filter is to disperse the phase of the coder output so that the succeeding coders cannot synchronize their levels from link to link. Figure 10 is a plot of the group delay of the all-pass filter that was used.

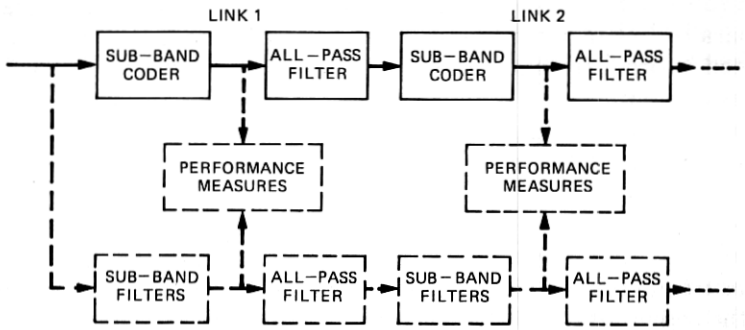
Figure 11 shows the results of s/n and SEG measurements for the tandem connections as a function of the number of tandem links. The solid lines refer to s/n measurements and the dashed lines refer to SEG measurements. The upper two curves refer to measurements made on the sub-band-to-PCM links in Fig. 9a, and the lower curves refer to measurements made on the sub-band-to-analog connections of Fig. 9b.

As seen in the figure, for the sub-band-to-analog connections, the s/n and SEG measures drop by roughly 3 dB per doubling of the number of tandem links, indicating that the quantization noise contributed by each link adds independently of other links.

In the sub-band-to-PCM connection, however, it is seen that the quantizer distortions do not add independently. After the first encoding, the succeeding coders tend to synchronize their quantizer levels to those of the first coder, and in this way they do not add any further distortion to the signal. This result is somewhat surprising in view of the fact that the quantizers in the sub-bands are separated by interpolating and de-



(a)



(b)

Fig. 9—Circuits for measuring performance of tandem connections of sub-band coders. (a) Sub-band/PCM links. (b) Sub-band/analog links.

cimating filters and all the sub-bands are summed at the outputs of the coders between links. In one example, we observed an s/n of 6.8 dB in the fourth sub-band of the first link. In the succeeding links, the s/n of this same coder went up to 18 dB in the fourth sub-band due to this synchronization effect.

Figure 12 shows the corresponding results for the LPC distance measurements on the tandem connections. Here again we see that the sub-band-to-PCM link performs better than the sub-band-to-analog link. A maximum LPC distance of 0.29 was observed for four sub-band-to-analog tandem connections, indicating that successive tandem connections do not excessively distort the spectrum of the coded speech over that of the initial coding.

Based on informal listening, the quality of two tandem connections does not appear to be much different than that of one encoding. For three sub-band/analog encodings, the differences become apparent, and with four links the differences are clearly noticeable.

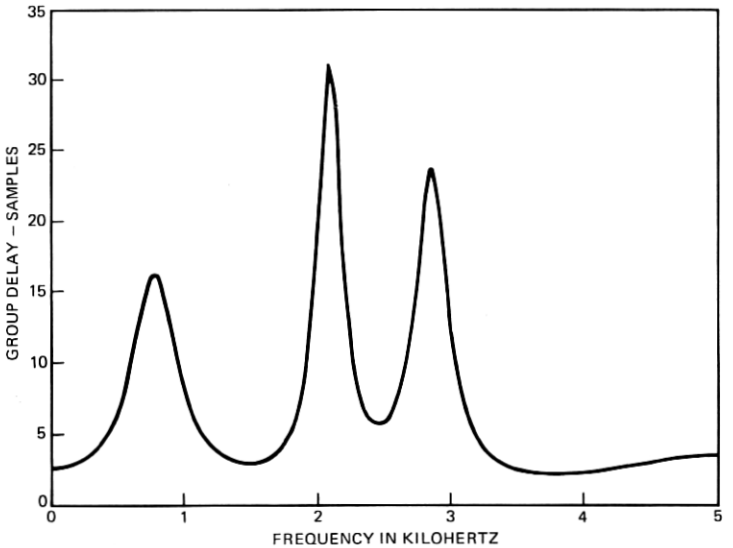


Fig. 10—Group delay as a function of frequency for the all-pass filters used to simulate analog links.

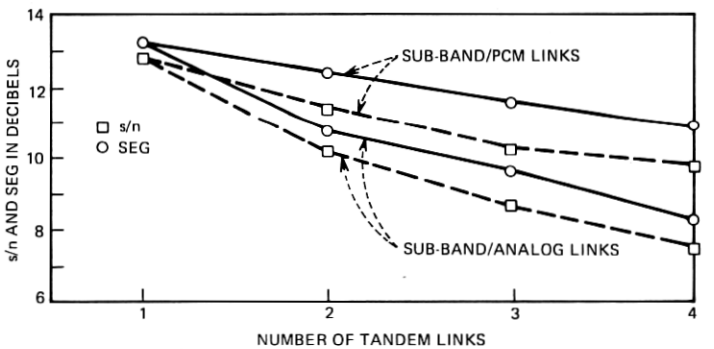


Fig. 11—*s/n* and SEG measurements for the sub-band/PCM and sub-band/analog tandem connections.

VI. CHANNEL ERRORS

The analysis of the sub-band coder performance under channel errors constituted the largest part of our experimental investigations. The coder performance was analyzed for bit error probabilities of up to 10 percent. We first analyzed the individual 4-, 3-, and 2-bit APCM coders in each of the sub-bands in order to assess their performance separately under channel errors. We then examined the use of a robust step-size adaption algorithm[12] in order to enhance the performance of these individual coders. For the 4- and 3-bit coders, we also investigated the use of partial bit error protection of the sign and most significant bits in the coders.

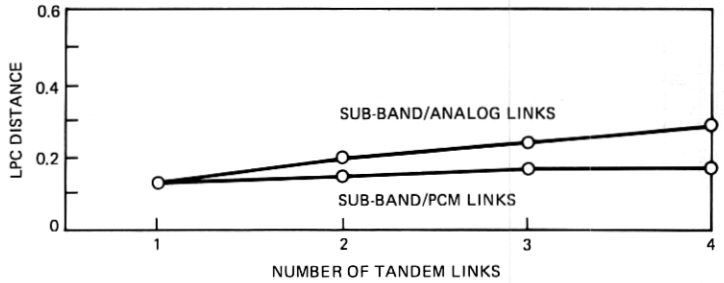


Fig. 12—LPC distance measurements for the tandem connections.

Based on these results, we then considered three overall sub-band coder designs. The first design was the 5-band coder described in Section II. The second design was the same coder with the robust step-size adaption algorithm for its APCM coders. In the third design, we considered a 4-band coder with a reduced bandwidth and a slightly lower bit rate. The remainding bits were applied to a partial bit error protection scheme to enhance its robustness under conditions of very high channel errors. We then analyzed and compared the performance of these three coders under channel errors.

6.1 The robust quantizer

The step-size adaption algorithm used in the sub-band coder is based on the one-word step-size memory scheme proposed by Jayant, Flanagan, and Cummiskey.[4,13] The coder input signal is quantized to one of 2^B levels, where B is the number of bits in the coder. The step-size adaption circuit examines the quantizer output bits for the $(r-1)$th sample and computes the quantizer step-size, Δ_r, for the rth sample according to the relation

$$\Delta_r = \Delta_{r-1} M(L_{r-1}), \tag{13a}$$

where

$$\Delta_{\min} \leqslant \Delta_r \leqslant \Delta_{\max} \tag{13b}$$

and where Δ_{r-1} is the step-size used for the $(r-1)$th sample. $M(L_{r-1})$ is a multiplication factor whose value depends on the quantizer magnitude level L_{r-1} at time $r-1$. It can take on one of 2^{B-1} values $M_1, M_2, \ldots M_{2^{(B-1)}}$. If the lower-magnitude quantizer levels are used at time $r-1$, a value of $M(L_{r-1}) = M_i$ less than one is used to reduce the next step-size. If upper magnitude levels are encountered, a value of M_i greater than 1 is chosen. In this way, the coder continuously adapts its step-size in an attempt to track the short-time variance of the input signal.

A disadvantage of the above adaption scheme is that, once a step-size

error occurs, it remains in error until the maximum or mininum step-size is reached. A modification of this algorithm, proposed by Goodman and Wilkinson[12] allows for the step-size computation to be less sensitive to past errors. This "robust" algorithm is based on the relation

$$\Delta_r = (\Delta_{r-1})^{\beta} \cdot M(L_{r-1}), \tag{14a}$$

where

$$\Delta_{\min} \leq \Delta_r \leq \Delta_{\max}. \tag{14b}$$

The parameter β is chosen to be slightly less than 1, and it determines how rapidly the effects of past errors are dissipated. In the limit when β goes to 1, the algorithm reduces to that of (13a).

• As the value of β is reduced, the M values must be adjusted to compensate for its effect on the step-size adaptation. As shown in Ref. 12, this compensation can be obtained by a simple scaling of the M values. If $\hat{M}_i$ represent the ideal M values for step-size adaption when $\beta = 1$, then the new M values, denoted as M_i, are approximately

$$M_i = G\hat{M}_i \qquad i = 1, 2, \ldots 2^{B-1}, \tag{15}$$

where G is a scaling factor that is dependent on β *and* on the expected value of Δ_r. In computer simulations, we determined G by optimizing the performance of the coders as a function of this scaling factor. Figure 13 is a plot of G as a function β that was used in our simulations. It is based on an expected value of Δ_r in the range of 500 to 5000, typically encountered in our computer simulations. As seen in the figure, when β varies from 1 down to 15/16 optimum scaling factor, G, increases from a value of 1 to about 1.5.

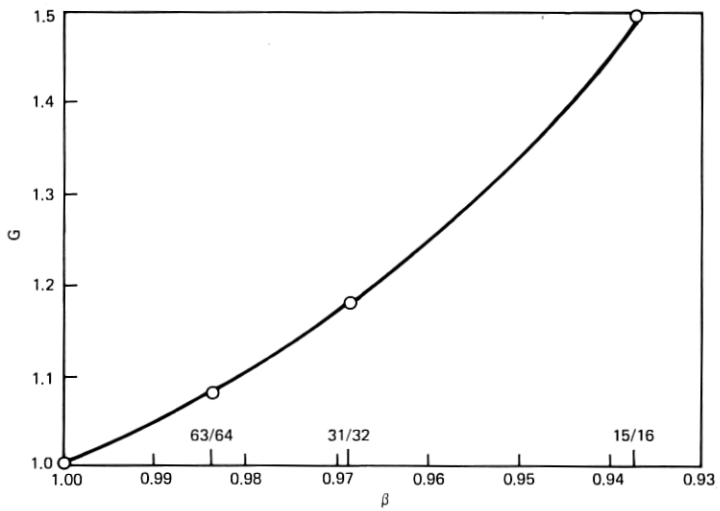


Fig. 13—Multiplier scaling factor, G, as a function of β, used for computer simulations.

6.2 Performance of individual coders under channel errors

The performance of the individual APCM coders was examined, in terms of s/n and SEG measurements, as a function of the bit-error rate and the robust quantizer parameter, β. Figures 14a to 14c show the results for the s/n measurements for the 4-, 3-, and 2-bit coders, respectively, as a function of bit-error rate where the bit-error rate corresponds to random channel errors. Figures 15a to 15c show similar results for the SEG measurements. Four values of β were used: 1, 63/64, 31/32, and

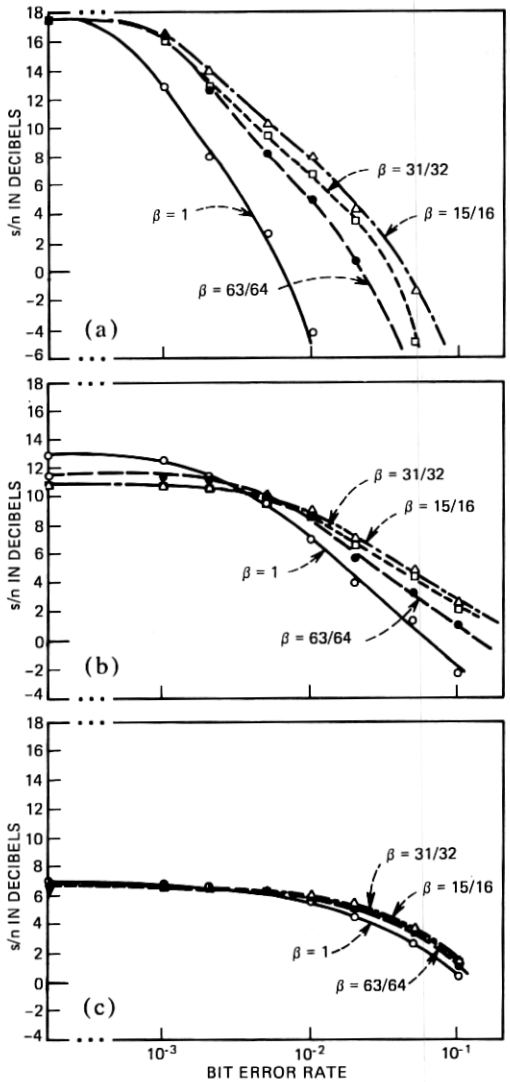


Fig. 14—s/n performance of the APCM coders as a function of the bit error rate. (a) 4-bit coder. (b) 3-bit coder. (c) 2-bit coder.

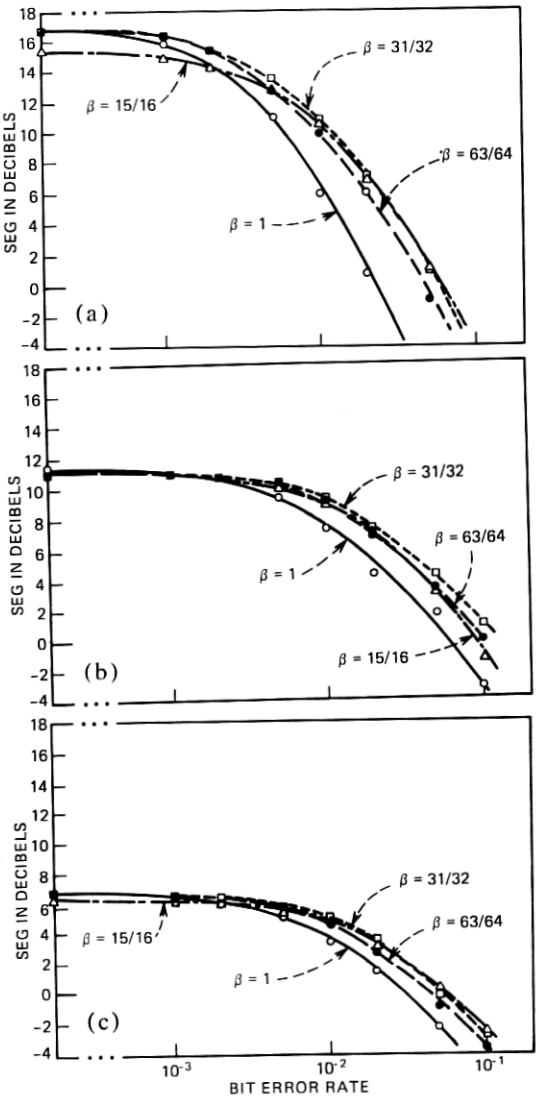


Fig. 15—SEG performance of the APCM coders as a function of the bit error rate. (a) 4-bit coder. (b) 3-bit coder. (c) 2-bit coder.

15/16. Each value of β corresponds to one curve in the plots. The bit error rate is plotted on a log scale and covers a range of 10^{-4} to 10^{-1}.

Several conclusions can be drawn from the results in Figs. 14 and 15. It is seen that the 4-bit coder is the most vulnerable to channel errors and that the 2-bit coder is the least vulnerable. Fortunately, the 4-bit coder receives the most improvement from the use of a robust step-size algorithm. The 2-bit coder, however, receives the smallest improvement from the robust algorithm.

The robust quantizer does not seem to affect the s/n and SEG measurements when no channel errors are present until the value of β is reduced below a value of about 31/32. This is assuming that the M values in the coder are appropriately scaled as discussed in the preceding section. If the M values are not properly scaled, then the performance of the coders will be significantly reduced as β decreases. For example, the performance of the 3-bit coder drops by about 6 dB in s/n and 3 dB in the SEG measure when β is reduced from 1 to 31/32 and the M values are not scaled according to (15).

The optimum choice for the robust quantizer parameter, β, for protection against channel errors appears to be about 31/32.

6.3 Partial bit error correction

Since the 4- and 3-bit coders are the most vulnerable to channel errors, the lower sub-bands which use these coders are affected the most by channel errors. Subjectively, these are also the most important bands since distortions in these bands quickly deteriorate the quality of the sub-band order.

One way to maintain the quality in these lower sub-bands is to provide for some partial bit-error correction in the transmission of these coder bits. In this section, we investigate the effect of sign and/or most significant magnitude bit protection on the performance of the 3- and 4-bit APCM coders.

To provide for error correction of transmitted bits, extra parity bits must be transmitted by the coder.[14,15] The degree of error protection that is achieved is strongly dependent on the design of the error protection block codes, the bit error rate of the channel, and the percentage of additional redundant bits that are transmitted for error protection. Fortunately, since the lower sub-bands typically have low sampling rates and therefore low transmission rates, the additional transmission rate required to provide partial bit-error correction of some of the bits in these lower sub-bands should be relatively small compared to the overall transmission rate of the coder. In this work, we have avoided issues of specific designs of block codes for bit error correction. We have instead assumed that ideal or nearly ideal error protection can be achieved. The results that we present should therefore be interpreted as upper bounds on what can be achieved, given a sufficient amount of extra transmission rate for error protection.

Figure 16a and 16b show results of s/n and SEG measurements on the 4-bit APCM coder, as a function of the bit error rate, for several bit error protection schemes. In all the results, a robust quantizer with $\beta = 31/32$ is used. The solid line shows the performance when no bit-error correction is used. The long dashed curve shows the results when the sign bit is ideally protected, and the short dashed line shows the results when

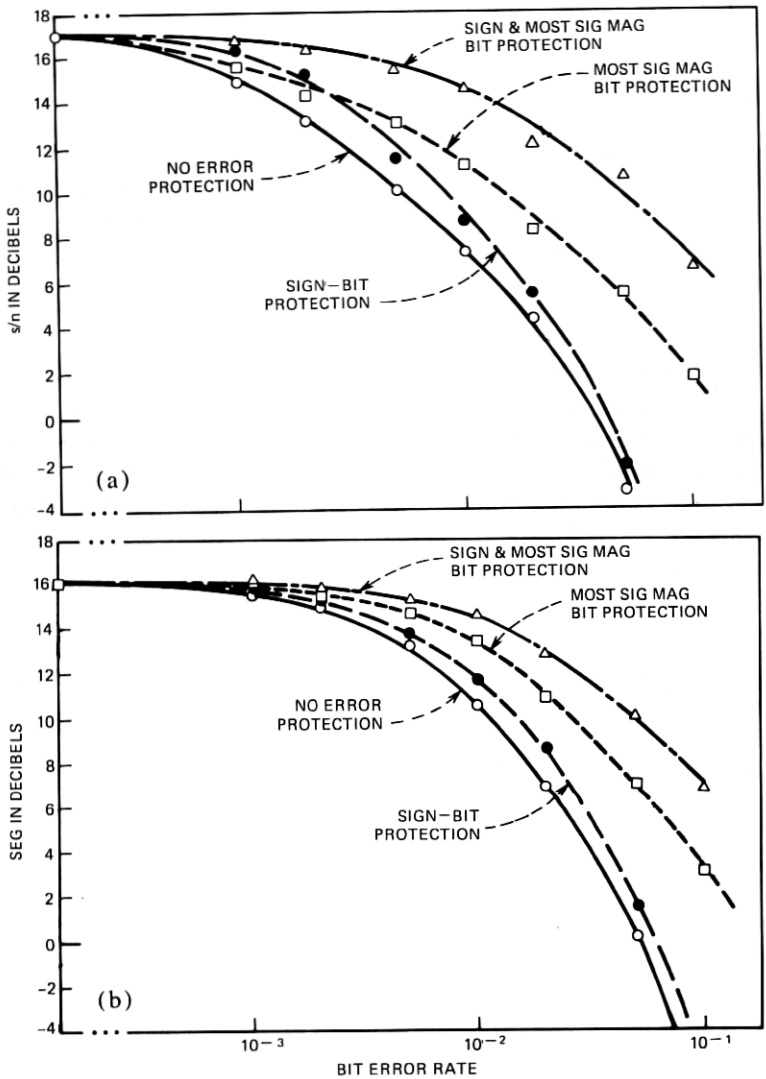


Fig. 16—Performance of the 4-bit APCM coder with partial error protection. (a) s/n measurements. (b) SEG measurements.

the most significant magnitude bit is ideally protected. Finally, the long and short dashed curve shows the performance when both the sign bit and the most significant magnitude bit are protected. As seen in the figure, protection of the most significant magnitude bit alone gives a better performance than when the sign bit is protected alone. This occurs because an error in the sign bit results in a single isolated error, whereas an error in the most significant magnitude bit causes a step-size error which propagates for many samples. At high bit-error rates, significant

improvements in coder performance are possible with bit-error protection.

Figure 17 shows similar results for the 3-bit APCM coder. In this case, the protection of only one bit was considered, either the sign bit (long dashed line) or the most significant magnitude bit (short dashed line). It is seen that protecting the sign bit leads to about the same improvement as the most significant magnitude bit. In comparing Figs. 16 and 17, it can be seen that the improvement of the 3-bit coder performance with error protection in high channel errors is not as large as the improvement obtained for the 4-bit coder.

6.4 A sub-band coder design for high channel errors

As noted in the previous section, when high channel errors are encountered, it is possible to divert a part of the transmission rate to the protection of bits in the lower sub-band(s). In this way, some of the coder quality at low channel error rates can be traded for more robustness of the coder at high channel error rates. In this section, we consider an example of such a design.

Table II shows the choice of bands and bit allocations for a 4-band

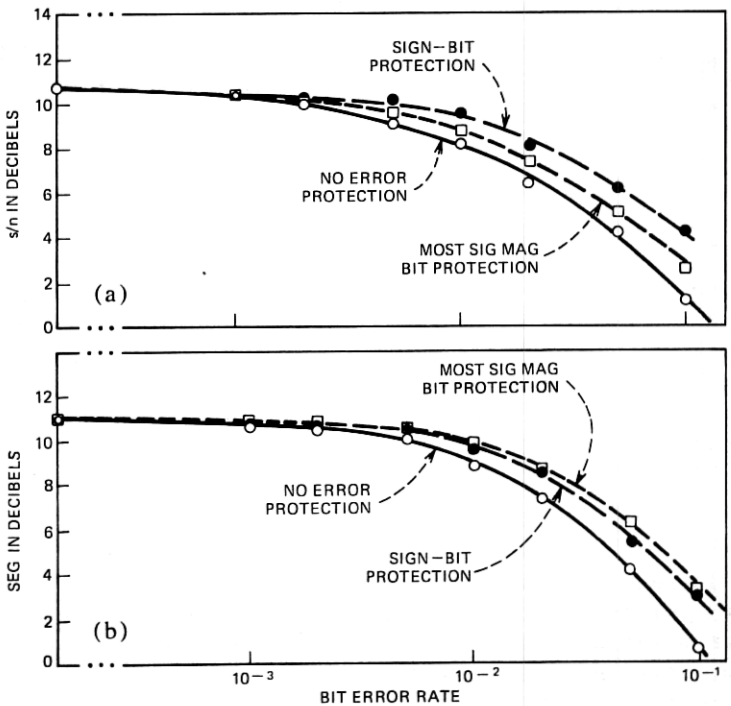


Fig. 17—Performance of the 3-bit APCM coder with partial error protection. (a) s/n measurements. (b) SEG measurements.

Table II — 16 kb/s 4-band coder with partial bit error correction

Band	Band Edges (Hz)	Sampling Freq (Hz)	Min Step-Size (dB)	Bit Allocation	Kb/s
1	250–500	500	(Ref)	4	2.0
2	500–1000	1000	−1.9	3	3.0
3	1000–2000	2000	−6	2	4.0
4	2000–3000	2000	−10	2	4.0
SYNC AND ERROR CORRECTION					3.0
					16.0

16-kb/s coder with partial bit-error correction in the lowest band. The frequency response of this coder is shown in Fig. 18. In comparison to the 5-band coder, it is seen that this coder has a narrower overall bandwidth and an additional notch in its frequency response. Thus, the quality of this coder tends to be more reverberant than that of the 5-band design. The LPC distance measure for this coder, measured according to Fig. 8, was 0.82 compared to 0.58 for the 5-band coder. When the sub-band filters alone were measured, a distance of 0.69 was observed compared to 0.53 for the 5-band coder.

In trade for this reduced quality, the 4-band coder has 3 kb/s of remaining transmission rate or 18.75 percent of its total transmission rate which can be used for bit error protection. This is applied to the protection of the sign and most-significant magnitude bits of the 4-bit coder in the first sub-band.

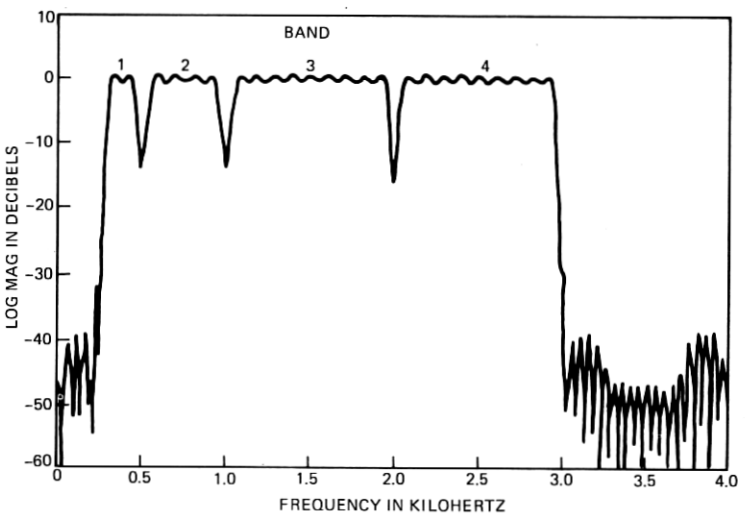


Fig. 18—Frequency response of the 4-band coder in Table II.

6.5 Overall performance of the sub-band coders with channel errors

In this section, we present the results of computer simulations of three different sub-band coder designs under conditions of channel errors. The simulations were made with random channel errors with error rates of up to 10^{-1}. The first coder, coder A, is the 5-band coder in Table I with no robust quantizer (i.e., $\beta = 1$). Coder B is the same 5-band design with a robust quantizer with $\beta = 31/32$. Coder C is the 4-band design, in Table II, for high channel errors. It has a robust quantizer, $\beta = 31/32$, and assumes ideal error protection of the sign and most-significant bit in its first sub-band (the 4-bit APCM coder).

Figure 19 shows the results of the s/n and SEG measurements for the three coders as a function of the bit error rate. Coders A and B, the 5-band designs, have the best performance and quality at very low error rates. The use of the robust quantizer does not significantly reduce the performance of Coder B (assuming the M values are scaled properly) at low error rates. The 4-band design has a somewhat lower quality at low error rates due to its reduced bandwidth and lower effective transmission rate.

As the bit error rate increases, the performance of the unprotected coder, coder A, drops rapidly. Channel error distortions are noticeable at error rates of 2×10^{-3}. At error rates of 5×10^{-3} and 10^{-2}, the quality drops rapidly and at error rates of 2×10^{-2} the coder is essentially unintelligible.

The use of the robust quantizer significantly improves the performance of the 5-band coder for moderate error rates. Coder B has noticeable degradations in quality at bit-error rates of about 10^{-2}. At error rates of 2×10^{-2}, this quality degrades rapidly and at error rates of 5×10^{-2} the coder starts to become unintelligible.

The 4-band coder, coder C, holds up well for error rates up to about 2×10^{-2} before the effect of channel errors becomes noticeable. Its quality, however, is slightly lower to begin with. At error rates of 10^{-1}, the quality degrades sharply although the coder still appears to be quite intelligible.

Figure 20 shows the results of the LPC distance measure on the three coders. These results do not appear to agree well with s/n and SEG measures nor do they agree well with our informal subjective observations. For example, at bit error rates of 10^{-2} the LPC distance of coder A is 0.35, indicating that the coder should have reasonably good quality. In fact, the subjective quality of the coder at this point was significantly degraded. Also, the LPC distance failed to sufficiently distinguish the differences in quality between coders B and C at high error rates.

To investigate this problem in more detail, we plotted the individual segmental LPC distances d_{1k} and d_{2k} defined in (7) and (10) as a function of time (measured in segments). Figure 21a shows these results for coder

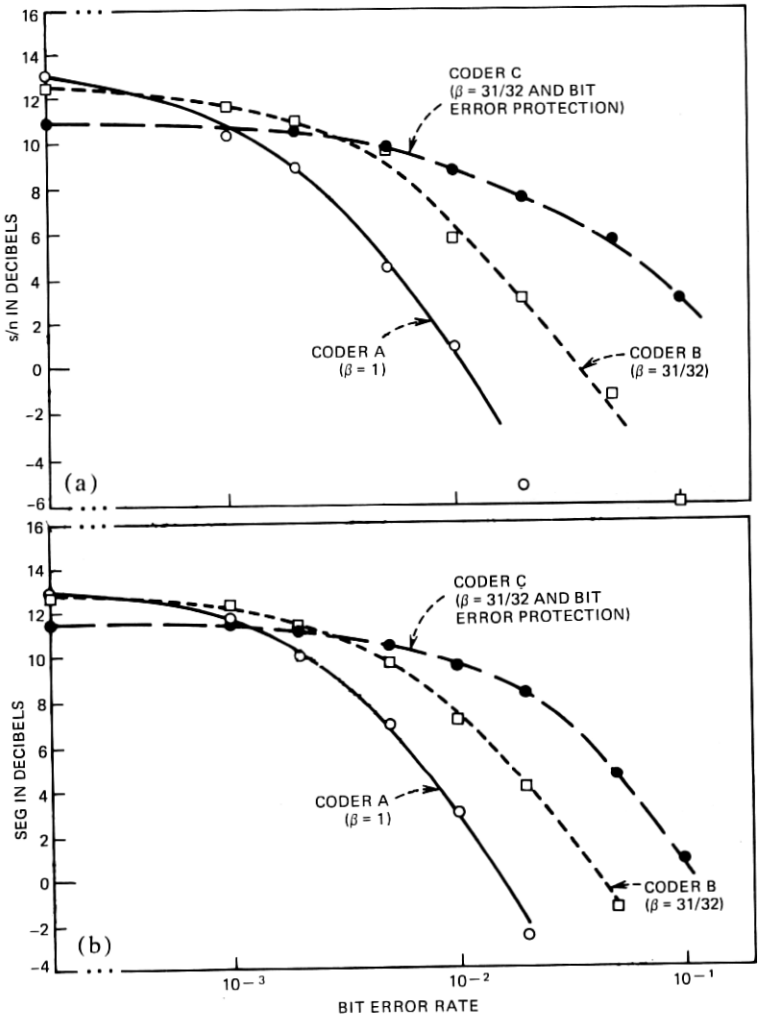


Fig. 19—Performance of the three sub-band coder designs under channel errors. (a) s/n measurements. (b) SEG measurements.

A at a bit error rate of 10^{-2} for two concatenated sentences. From this plot, it becomes clear as to what is happening. On the average, the coder performance is quite good. However, in about 10 or 12 isolated segments, severe distortions were observed where channel errors occurred in lower sub-bands. Because of these isolated errors, the entire sentence sounds poor in quality. Figure 21b shows similar results for coder A at error rates of 5×10^{-2}. Again, it is seen that there are numerous segments in which the distortions are intolerable; however, on the average, the distortion was not that bad; i.e., it was below 1. Subjectively, the presence of these large errors made the sentence virtually unintelligible.

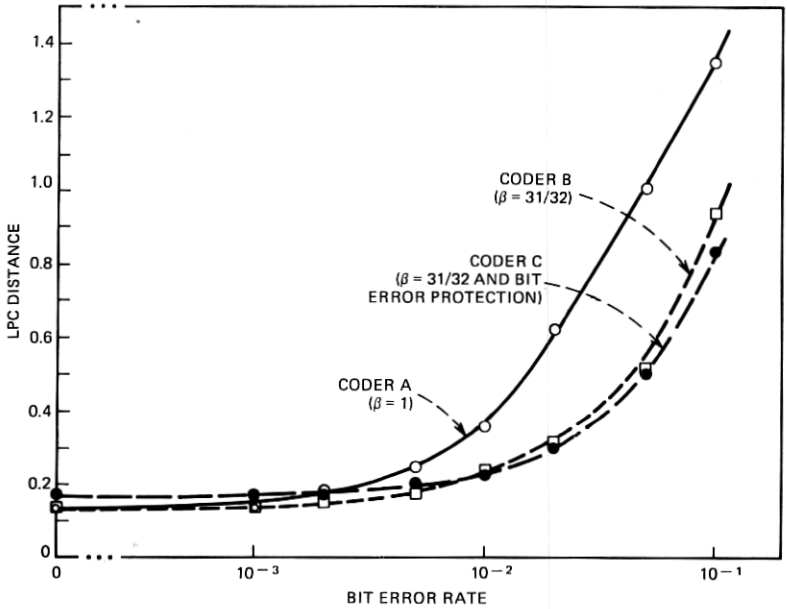


Fig. 20—LPC distance as a function of bit-error rate for the three sub-band coder designs.

From these observations, we conclude that the LPC distance, used properly, is in fact a good indicator of quality. However, when an overall measure of quality for an utterance is required, something more sophisticated than a simple mean of the segmental distances must be used. This is particularly important in the case of channel errors.

VII. CONCLUSIONS

In summary, a number of general conclusions can be drawn from the results of this work.

(*i*) When the maximum-to-minimum step-size ratios of the APCM coders is 128 and the dynamic range of sub-bands are properly aligned, the quality of the coder remains relatively constant over a range of input levels of about 30 dB. This range increases by about 6 dB per doubling of this step-size ratio. The idle channel noise performance of the coder can be improved by the use of a mid-rise/mid-tread switch on the quantizers in the APCM coders.

(*ii*) For tandem connections of sub-band coders with conversion to analog format between links, the signal-to-noise ratio drops by roughly 3 dB per doubling of the number of tandem coders. When linear phase FIR filters are used in the coders and they are connected by PCM links, the step sizes of the coders tend to synchronize, and the performance of the tandem connection improves.

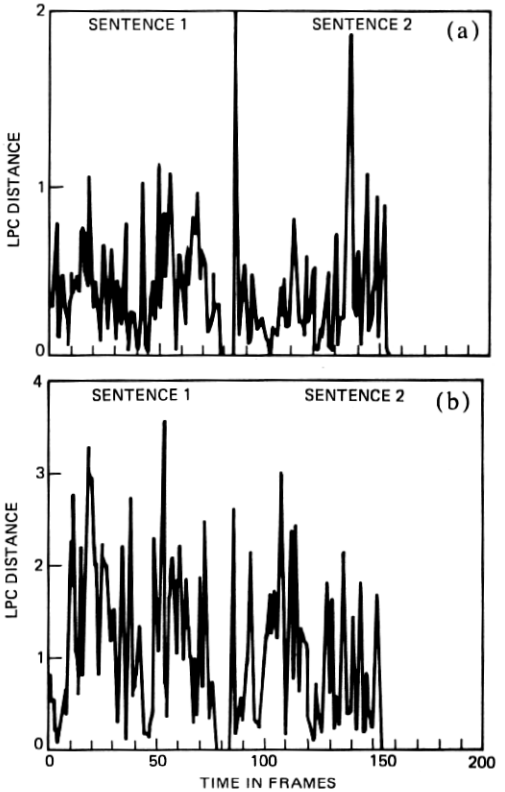


Fig. 21—LPC distance as a function of time for (a) coder A at a bit error rate of 10^{-2} and (b) coder A at a bit error rate of 5×10^{-2} (note a difference in scale).

(*iii*) The effects of channel errors in an unprotected sub-band coder are first observed at bit error rates of about 2×10^{-3}. At error rates of 2×10^{-2}, the quality of the coder is essentially unintelligible. When a robust quantizer algorithm is used, errors are first noticeable at bit error rates of about 10^{-2}, and at error rates above 5×10^{-2} the coder becomes unintelligible. When both a robust quantizer and partial bit-error protection is used in the lower sub-band(s), the effect of channel errors is not significant until error rates of about 2×10^{-2} are reached and the coder appears to be intelligible at error rates as high as 10^{-1}. The above results are based on the assumption that sufficient protection is provided for the synchronization and parity bits so that no loss of synchronization occurs between high channel errors and that nearly ideal error protection is possible for the coder bits which are protected in the partial bit-error protection scheme.

REFERENCES

1. R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Sub-Bands," B.S.T.J., *55,* No. 8 (October 1976), pp 1069–1085.
2. R. E. Crochiere, "On the Design of Sub-band Coders for Low Bit Rate Speech Communication," B.S.T.J., *56,* No. 5 (May-June 1977), pp. 747–770.
3. D. Esteban and C. Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes," 1977 IEEE Int'l. Conf. on Acoust., Speech and Sig. Proc. (May 1977), pp. 191–195.
4. N. S. Jayant, "Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantizers," Proc. IEEE, *62* (May 1974), pp. 611–632.
5. N. S. Jayant, (ed.) *Waveform Quantization and Coding,* New York: IEEE Press, 1976.
6. P. Noll, "Adaptive Quantization in Speech Coding Systems," Int. Zurich Seminar on Digital Communication (IEEE) (October 1976), pp. B3.1 to B3.6.
7. R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals," IEEE Trans. Acoust., Speech and Sig. Proc., *ASSP-25* (August 1977), pp. 299–309.
8. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust., Speech and Signal Processing, *ASSP-23,* (Feburary 1975), pp. 67–72.
9. M. R. Sambur and N. S. Jayant, "LPC Analysis/Synthesis from Speech Inputs Containing Quantizing Noise or Additive White Noise," IEEE Trans. Acoust., Speech and Sig. Proc., *ASSP-24* (December 1976), pp. 488–494.
10. L. R. Rabiner, private communication.
11. R. E. Crochiere, "A Mid-Rise/Mid-Tread Quantizer Switch for Improved Idle-Channel Performance in Adaptive Coders," B.S.T.J., this issue, pp. 2953–2955.
12. D. J. Goodman and R. M. Wilkinson, "A Robust Adaptive Quantizer," IEEE Trans. Commun. (November 1975), pp. 1362–1365.
13. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J., *52,* No. 7 (September 1973), pp. 1105–1118.
14. R. W. Hamming, "Error Detecting and Error Correcting Codes," B.S.T.J., *29,* No. 2 (April 1950), pp. 147–160.
15. E. R. Berlekamp (ed.), *Key Papers in the Development of Coding Theory,* New York: IEEE Press, 1974.
16. R. E. Crochiere, J. L. Flanagan, and S. A. Webber, "Digital Speech Communication System for Minimizing Quantizing Noise," U.S. Patent 4,048,443, September 13, 1977.