AT&T Technical Journal
Vol. 64, No. 6, July–August 1985
Printed in U.S.A.

Variable-Length Packetization of μ-Law PCM Speech*

By R. STEELE† and F. BENJAMIN‡

(Manuscript received October 31, 1984)

A variable-length packetization process is proposed for logarithmic Pulse Code Modulation (log-PCM)-encoded speech signals. Blocks of W log-PCM words are reduced in size by discarding words on the basis that the receiver can recover the speech with a signal-to-noise ratio (s/n) that is, in general, above a specified value. Specifically, blocks of two hundred fifty-six, 8-bit, μ-law PCM words are left unchanged or reduced to either 214, 205, 192, 171, or 128 words. These blocks are then formulated into packets. The discarded samples at the dispatching terminal are replaced at the receiver by means of adaptive interpolation. We found that by specifying the s/n of the decoded speech in each variable-length packet to be above 27 dB, the reduction in the transmitted speech data was 25 percent, while the recovered speech signal had negligible perceptual degradation.

I. INTRODUCTION

Speech signals are inherently bursty. Indeed, conversational speech is composed of speech segments sandwiched between variable durations of silences and interword pauses, as well as intraword gaps. Even during a speech utterance, the information rate is not constant and can exhibit surges. Stated simply, the statistical properties of speech are nonstationary, and our speech encoders are designed to exploit in

* Most of this work was conducted when the authors were members of AT&T Bell Laboratories.

† University of Southampton, England. ‡ Perkin-Elmer Corporation, Tinton Falls, New Jersey.

various degrees this nonstationarity. However, the majority of existing speech encoders[1-3] generate data at a fixed rate, and as the information content of the speech source is varying, so is the recovered speech quality. If we stipulate that the speech quality should be maintained above a specified level, then the variation in the information content of the speech signal causes the bit rate to vary. Earlier attempts at variable-bit-rate encoding have usually employed a buffer to convert the variable bit rate generated by the encoder to a fixed rate for transmission over the channel. Such arrangements suffer from long delays in the recovered speech signal and are often susceptible to transmission errors. However, when the transmission medium can accommodate many channels, variable-rate digitized speech can be used with advantage. By converting this variable data rate into packets of varying length, we are able to reduce the overall data rate for a given number of speech channels. These packets are amenable to statistical multiplexing, which is essentially a generalization of digital-speech interpolation. By employing packet switching,[4] we can conveniently combine statistical multiplexing with switching and routing procedures. Thus, the line of enquiry to be followed here is concerned with the formulation of packetized digital speech, and, in particular, variable-length packets that utilize the time-varying statistics of speech signals.

It is interesting to observe that there are many speech-encoding algorithms that operate on blocks of speech samples.[1-3] The resulting blocks of digital speech are eminently suited to packetization. All that is required to yield packets of digital speech is the addition of suitable headers and flags to the blocks. These additions enable the multiplexer and the switching and routing network to identify the beginning and termination of the packets, and give the receiver access to any decoding commands. Although the number of digital encoders conceived are legion, only a few types are employed in the networks. Foremost is logarithmic Pulse Code Modulation (log-PCM),[5] which is formidably entrenched in the international networks. Adaptive differential pulse code modulation is now being introduced[6] on trunk networks, and adaptive delta modulation[7] has been used in a limited way in subscriber loops and over satellite links. In this discourse we will focus on the variable-length packetization of log-PCM speech.

In Section II we review the adaptive interpolation procedures to be used in our variable-length packetized μ-law PCM-encoded speech system. The technique of determining the packet length is described in Section III. For each packet we ensure that the recovered speech quality at the receiver is, in general, above a minimum specified value. In Section IV we discuss our results. The final section deliberates on the properties and ramifications of the proposed packetization method.

II. APPLICATION OF ADAPTIVE INTERPOLATION TO REDUCE THE TRANSMITTED DATA RATE

Starting with blocks of W speech samples, it is possible to reject certain samples at the transmitter with the knowledge that the receiver is able to replace the discarded samples with virtually an imperceptible degradation. The replacement technique[8] employs adaptive interpolation. Thus, we commence with a block of W speech samples, discard some samples prior to transmission, and at the receiver reinsert the missing samples to yield speech that is acceptable according to a simple objective criterion. For a given recovered speech quality criterion, the number of samples discarded in any block is dependent on the correlative properties of its W speech samples. When the samples are highly correlated more samples are rejected prior to transmission, and vice versa when the samples are relatively uncorrelated. This is because of the nature of the interpolation procedure. In general, the W samples in each block are reduced in number. We may, therefore, think of packets having W speech samples converted to packets containing fewer speech samples. The length of the packet, i.e., the number of samples in the packet, depends upon the ability to discard samples in the confident knowledge that the receiver will reproduce W samples with a speech quality that is above a specified value. However, in our deliberations we will not be concerned with variable-length packetization of samples, but with log-PCM speech data. This matter will be dealt with in Section III.

Next we will briefly review the adaptive procedure used to replace the discarded samples. This procedure is cardinal in controlling the length of the transmitted packet.

The method is fully described in Ref. 8 and only a brief review will be presented here. Essentially, speech is sampled at or above the Nyquist rate to give the sequence $\{x_k\}$. Operating on a block containing W contiguous samples of this sequence, we reject every nth member in order to achieve a reduction in the speech data. The discarded samples are replaced by

$$\hat{z}_r; \qquad r = n, 2n, \ldots, W - n, W$$

to give the recovered sequence,

$$x_1, \ldots, x_{n-1}, \hat{z}_n, x_{n+1}, \ldots, x_{2n-1}, \hat{z}_{2n}, x_{2n+1}, \ldots, x_{W-1}\hat{z}_W.$$

The interpolated samples are

$$\hat{z}_r = \sum_{i=-(n-1)}^{n-1} a_i x_{r+i}, \tag{1}$$

where a_i are the interpolation coefficients, $a_0 = 0$. The set of interpo-

$$
\mathbf{A} = \begin{bmatrix}
1 & R(-5,-6) & R(-4,-6) & R(-3,-6) & R(-2,-6) & R(-1,-6) & R(1,-6) & R(2,-6) & R(3,-6) & R(4,-6) & R(5,-6) & R(6,-6) \\
R(-6,-5) & 1 & R(-4,-5) & R(-3,-5) & R(-2,-5) & R(-1,-5) & R(1,-5) & R(2,-5) & R(3,-5) & R(4,-5) & R(5,-5) & R(6,-5) \\
R(-6,-4) & R(-5,-4) & 1 & R(-3,-4) & R(-2,-4) & R(-1,-4) & R(1,-4) & R(2,-4) & R(3,-4) & R(4,-4) & R(5,-4) & R(6,-4) \\
R(-6,-3) & R(-5,-3) & R(-4,-3) & 1 & R(-2,-3) & R(-1,-3) & R(1,-3) & R(2,-3) & R(3,-3) & R(4,-3) & R(5,-3) & R(6,-3) \\
R(-6,-2) & R(-5,-2) & R(-4,-2) & R(-3,-2) & 1 & R(-1,-2) & R(1,-2) & R(2,-2) & R(3,-2) & R(4,-2) & R(5,-2) & R(6,-2) \\
R(-6,-1) & R(-5,-1) & R(-4,-1) & R(-3,-1) & R(-2,-1) & 1 & R(1,-1) & R(2,-1) & R(3,-1) & R(4,-1) & R(5,-1) & R(6,-1) \\
R(-6,1) & R(-5,1) & R(-4,1) & R(-3,1) & R(-2,1) & R(-1,1) & 1 & R(2,1) & R(3,1) & R(4,1) & R(5,1) & R(6,1) \\
R(-6,2) & R(-5,2) & R(-4,2) & R(-3,2) & R(-2,2) & R(-1,2) & R(1,2) & 1 & R(3,2) & R(4,2) & R(5,2) & R(6,2) \\
R(-6,3) & R(-5,3) & R(-4,3) & R(-3,3) & R(-2,3) & R(-1,3) & R(1,3) & R(2,3) & 1 & R(4,3) & R(5,3) & R(6,3) \\
R(-6,4) & R(-5,4) & R(-4,4) & R(-3,4) & R(-2,4) & R(-1,4) & R(1,4) & R(2,4) & R(3,4) & 1 & R(5,4) & R(6,4) \\
R(-6,5) & R(-5,5) & R(-4,5) & R(-3,5) & R(-2,5) & R(-1,5) & R(1,5) & R(2,5) & R(3,5) & R(4,5) & 1 & R(6,5) \\
R(-6,6) & R(-5,6) & R(-4,6) & R(-3,6) & R(-2,6) & R(-1,6) & R(1,6) & R(2,6) & R(3,6) & R(4,6) & R(5,6) & 1
\end{bmatrix}
$$

Fig. 1—Matrix **A** for $n = 2, 3, \ldots, 6$.

lation coefficients that minimizes the mean-square interpolation error

$$e_r = x_r - \hat{z}_r; \qquad r = n, 2n, \ldots, W, \tag{2}$$

is

$$\alpha = [a_{1-n}, a_{2-n}, \ldots, a_{-1}, a_1, \ldots, a_{n-2}, a_{n-1}]^T \tag{3}$$

and is found from

$$\alpha = \mathbf{A}^{-1}\mathbf{C}, \tag{4}$$

where

$$\mathbf{C} = [R(0, 1-n), R(0, 2-n), \ldots, R(0, -1), R(0, 1), \ldots, R(0, n-2), R(0, n-1)]^T. \tag{5}$$

The superscripts -1 and T represent inverse and transpose operations, respectively. The matrix **A** for $n = 2, 3, \ldots, 6$ is shown in Fig. 1, where the innermost dotted enclosure, the next dotted enclosure, and so on, and the complete matrix refer to $n = 2, 3, \ldots, 6$, respectively. We will not perform adaptive interpolation with $n > 6$ for the reasons stated in Section 3.1. The elements in the matrix **A**, and in the vector **C** given by eq. (5), are

$$R(k, j) = \frac{\sum_{r=n}^{W} x_{r+k}x_{r+j}}{\sum_{r=n}^{W} x_{r+j}^2}$$

$$k = \pm 1, \pm 2, \ldots, \pm n - 1$$

$$j = \pm 1, \pm 2, \ldots, \pm n - 1$$

$$r = n, 2n, \ldots, W. \tag{6}$$

Observe that

$$a_{-p} \neq a_p, \qquad p = 1, 2, \ldots, n-1 \tag{7}$$

and, therefore, that the number of interpolation coefficients to be computed is twice that required when the simple correlation coefficient

$$R(\tau) = \frac{\sum_{r=1}^{W-\tau} x_r x_{r+\tau}}{\sum_{r=1}^{W} x_r^2}, \qquad r = 1, 2, \ldots, W \tag{8}$$

is employed. The use of $R(\tau)$ yields significantly greater interpolation errors than does the use of $R(k, j)$.[8]

III. VARYING n TO DEMAND A BLOCK SIGNAL-TO-NOISE RATIO ABOVE A SPECIFIED LEVEL

In the previous subsection, we summarized (1) the procedure for removing every nth sample in a block of W speech samples and (2) the reinsertion of these discarded samples at the receiver by using adaptive interpolation based on the transmitted samples. Thus, the interpolated sample $\hat{z}_r$, given by eq. (1), is the weighted sum of $2(n-1)$ neighboring speech samples that were transmitted. In the original paper[8] the value of n was a constant, although the effect of varying n was studied. When the signal-to-noise ratio (s/n), computed over a block of length W samples, was plotted as a function of block number during a sentence of speech, it exhibited huge variations.

For example, for $W = 256$, $n = 4$, the s/n in a block could have been as small as 4 dB and as large as 58 dB, while the average of these block s/n's, the segmental s/n, was 35 dB. This implies that when the block s/n was low, the speech was relatively uncorrelated and n should have been significantly in excess of 4. By contrast, a high-block s/n meant that the speech correlation was so high that more samples could have been discarded, and a value of $n = 2$ might have still yielded an acceptable block s/n. To attempt to guarantee a minimum block s/n, we must be prepared to vary n, even allowing for n to be infinity, i.e., no samples rejected, when the speech is very uncorrelated. The value of n clearly determines the number of samples in the transmitted block, being as large as W when n is infinity and as small as $W/2$ when n is 2. As blocks are virtually synonymous with packets, we are led to the notion of variable-length packets.

3.1 Generation of variable-length packets

So far our discussion of adaptive interpolation has involved the processing of analog speech samples. We will not, however, be con-

cerned with the transmission of samples. Rather, we will describe the packetizing of μ-law PCM-encoded speech. The packets are formulated at a transmitting terminal, gathering the encoded speech words as they are produced. (Alternatively, the packetization may operate at a node in the network, converting a 64-kb/s data stream into a sequence of variable-length packets and thereby reducing the overall data rate.)

Figure 2 shows the system arrangement for the creation of variable-length packets when packets are created at the subscriber's packet network terminal or interface. Figure 3 is the similar system on the receiver side. The input speech sequence is μ-law PCM encoded, and the resulting words are stored in buffer 3. The μ-law PCM-encoded speech is locally decoded to give quantized speech samples that are directed into buffer 1. While this is in progress the sequence of previously decoded samples $\{x_k\}$ is removed from buffer 2 for processing. Buffers 1 and 2 are of length W samples. Buffer 3 holds the contents of buffers 1 and 2, but in the μ-law format. Starting with $n = 2$, every other sample in $\{x_k\}$ is discarded to yield the sequence $\{z_k\}$. The two interpolation coefficients a_{-1} and a_1 are computed from $\{x_k\}$ using eq. (4). These coefficients must be encoded for transmission, so they are scaled by a factor K to ensure that the largest coefficient will not exceed the range of the quantizer. The quantized coefficients are binary encoded and, if selected for transmission, are conveyed to the multiplexer (MUX). The coefficients are also locally descaled by $1/K$ to give the decoded coefficients $\hat{a}_{-1}$, $\hat{a}_1$ that a receiver would have to use, assuming they would be regenerated without error. Operating on $\{z_k\}$ and $\{\hat{a}_k\}$, the discarded samples are reinserted using the adaptive interpolation procedure described in Section II. The locally recovered sequence $\{\hat{z}_k\}$ and the corresponding block of input speech samples $\{\tilde{x}_k\}$ are then used to calculate the block s/n:

$$(s/n)_n = 10 \log_{10} \left[\frac{\sum_{k=1}^{W} \tilde{x}_k^2}{\sum_{k=1}^{W} (\tilde{x}_k - \hat{z}_k)^2} \right]. \tag{9}$$

This block s/n, $(s/n)_n$, is compared with a reference s/n, $(s/n)_{\text{ref}}$, and if

$$(s/n)_n \geq s/n_{\text{ref}}; \qquad n \text{ accepted}, \tag{10}$$

then the value of $n = 2$ is accepted. Otherwise, the process is repeated using the next higher value of n, namely, 3. Should this later condition prevail, the coefficients $\hat{a}_{-2}$, $\hat{a}_{-1}$, $\hat{a}_1$, $\hat{a}_2$, are calculated, and this time $\{z_k\}$ has every third sample missing. The new $\{\hat{z}_k\}$ is formulated, the $(s/n)_3$ computed, and inequality (10) tested. If this inequality is not

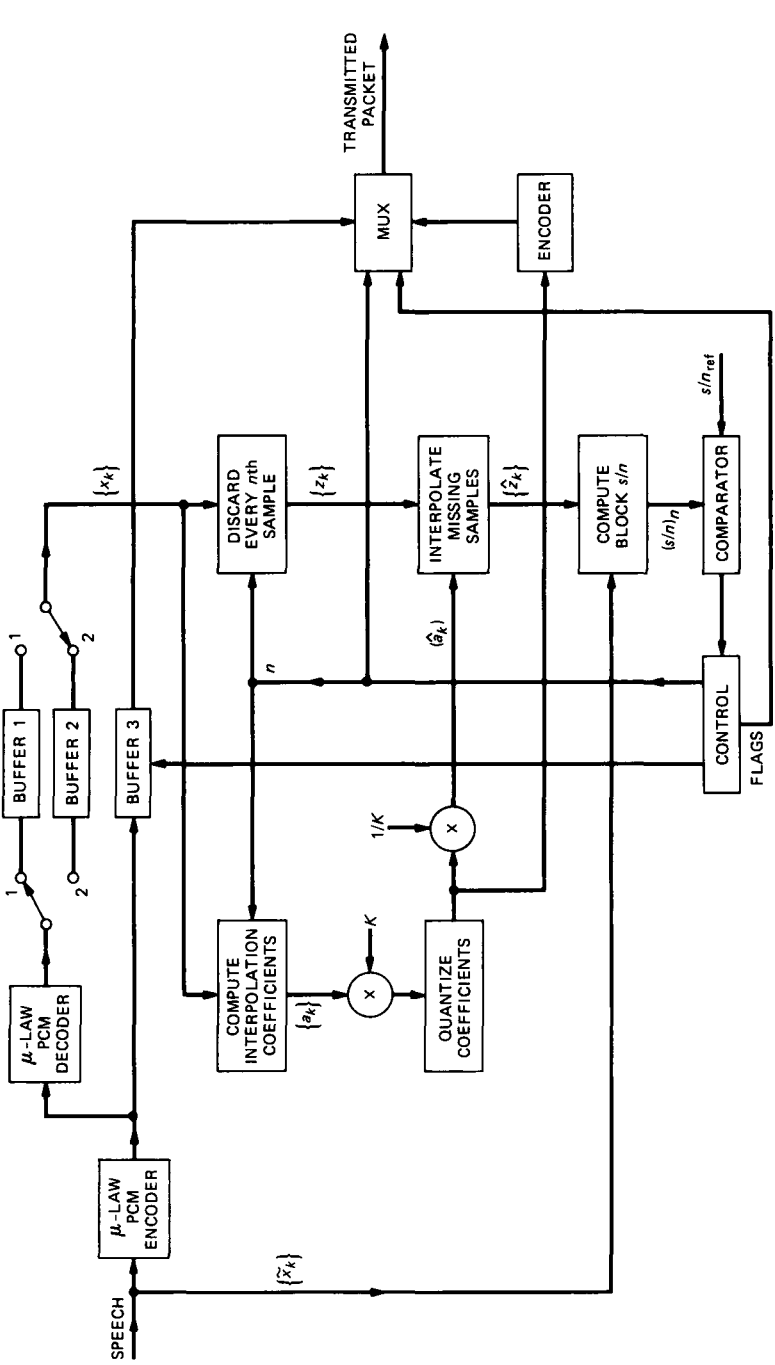

Fig. 2—System for generating variable-length packet speech.

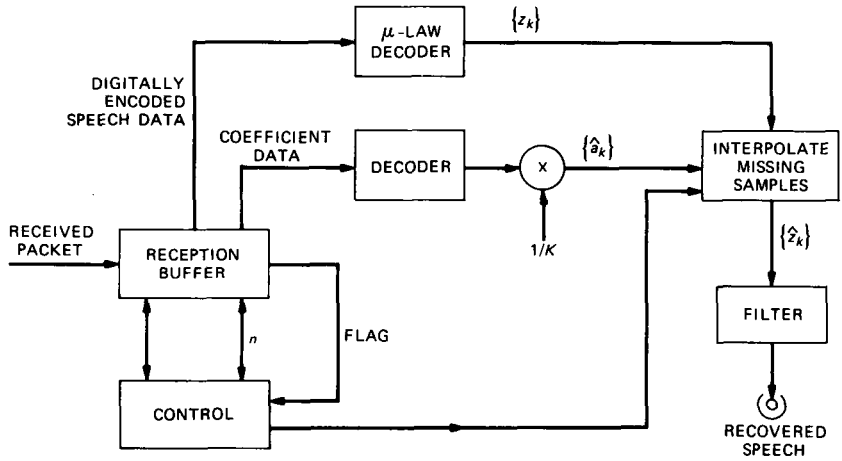

Fig. 3—System for decoding variable-length packet speech.

satisfied, n is increased to 4 and another iteration is performed, and so on. Should $n = 6$ fail to satisfy inequality (10), the iteration ceases because the packet size reduction with $n > 6$ is insufficient to justify the increase in processing time and system complexity. In this situation the entire block of μ-law PCM speech data is transmitted.

Observe that in computing $(s/n)_n$ we formulate the overall error [see the denominator of eq. (9)], which includes the effects of both quantization and interpolation. Thus, when inequality (10) is satisfied the overall s/n for the block of W speech samples is guaranteed to be above the specified minimum value of s/n_{ref}. However, we note that if s/n_{ref} is sufficiently high there will be some blocks where the quantization noise due to μ-law PCM encoding prevents inequality (10) from being satisfied, even though no samples are discarded. In this situation we do no worse than the performance of the μ-law PCM encoder, i.e., the s/n for the block is only dependent on the encoder noise, and the number of code words transmitted in the block remains unchanged at W.

The packetization scheme shown in Fig. 2 causes the original block of W speech samples, namely, $\{\tilde{x}_k\}$, to be processed to yield a data sequence where words have lengths of

$$L = \begin{cases} W\left(\dfrac{n-1}{n}\right); & n = 2, \ldots, 6 \\ W; & n = \infty. \end{cases} \tag{11}$$

This μ-law PCM data sequence is removed from buffer 3, with words discarded where appropriate, and conveyed to the multiplexer. These L μ-law PCM words are the digitized-speech data that are formulated

into the packet at the multiplexer. To these data must be added the header, which consists of a 3-bit binary number for n, followed by an 8-bit representation of each interpolation coefficient, and a system flag. The end of the packet is signified by another system flag. These flags are dependent on the switching arrangements for a particular network, and because of the myriad of possible switching systems, we will refrain from detailing the properties of the flags. The number of channel-protection bits for n and $\{\hat{a}_k\}$ is dependent on the bit error rate specified for the network. Again we will not concern ourselves with the numerous network scenarios. We can state that (1) if n is represented by an 8-bit word—i.e., it is equivalent in length to one μ-law PCM word—the combination of L data words and the part of the header needed by the receiver to decode the speech data is

$$L_r = \begin{cases} L + 1 + 2n; & n = 2, \ldots, 6 \\ L + 1; & n = \infty \end{cases} \tag{12}$$

because $2n$ words are required for the transmission of the interpolation coefficients when every nth word is discarded; and (2) when $n = \infty$ only one extra word is required to inform the receiver of the situation. It is not necessary that the end of packet be conveyed, because a knowledge of n specifies the length of the packet. By employing an end-of-packet flag, we are allowing for errors in the reception of n.

3.2 Decoding the variable-length packets

The received packet enters a buffer. The flag at the front end of the packet causes the control system to remove the data word representing n. Armed with this information, the coefficient data and digital-speech data are extracted from the buffer and subsequently decoded. The sequences $\{\hat{a}_k\}$ and $\{z_k\}$ are formed (see Fig. 3), and the samples that were rejected at the transmitter are reinserted by adaptive interpolation to yield $\{\hat{z}_k\}$. Speech is decoded in successive packets such that although there are L elements in $\{z_k\}$, where L is a function of n, there are always W samples in the recovered sequence $\{\hat{z}_k\}$. Thus, speech samples are presented to the final filter at the same rate at which they were originally generated.

IV. RESULTS

The sentences "Live wires should be kept covered," spoken by a male, and "To reach the end he needs much courage," enunciated by a female, were concatenated, bandlimited from 0.3 to 3.2 kHz, and sampled at 8 kHz to yield the input speech sequence. This sequence was 8-bit, μ-law PCM encoded, $\mu = 255$, to provide the input digital-speech sequence.

Figure 2 shows the block diagram of the variable-length packetization system. The μ-law PCM data stream was decoded, and the resulting samples were directed into either buffer 1 or 2, as described in Section 3.1. The selection of W was 256, a value that had been found to provide a reasonable recovered s/n for the range of n values considered here.[8] This choice of W also ensured that the amount of header information to convey the interpolation coefficients was insignificant compared with the number of data words in the packet and that the delay in the recovered speech signal at the receiving terminal due to the block size was not excessive. The duration of these blocks of 256 speech samples was 32 ms.

The coefficients $\{a_k\}$ were quantized using, for convenience, an 8-bit, μ-law quantizer, $\mu = 255$. The scaling factor K was determined as follows. The system of Fig. 2 was operated as described in Section 3.1, but without the interpolation coefficients $\{a_k\}$ being quantized. However, the value of K for each block was noted as

$$K_{\max} = \frac{V}{|a_{(\cdot)}|_{\max}}, \tag{13}$$

where V was the maximum range of the quantizer, namely, 4079 arbitrary units, and $|a_{(\cdot)}|_{\max}$ was the coefficient with the maximum magnitude. Although we considered quantizing each value of K with an 8-bit word and transmitting it as part of the header, we opted in favor of a fixed value of K. Accordingly, we set k to be the maximum value of K, namely, 7969, observed in the 152 blocks of input data used in our experiments. With this fixed K we commenced our packetization properly, quantizing the interpolation coefficients and employing them in our adaptive interpolation procedures. The same fixed K was used at the receiver.

The value of n used in our experiments was either 2, 3, 4, 5, 6, or infinity, depending on the minimum s/n, s/n_{ref}, stipulated. The block s/n for a particular value of n, namely, $(s/n)_n$, was computed using the input speech samples $\{\tilde{x}_k\}$ and $\{\hat{z}_k\}$. The latter sequence contained the original speech samples contaminated by both quantization and interpolation noise. Thus, the $(s/n)_n$ values employed in our simulations are indicative of the quality of the recovered speech. To visually emphasize, in Fig. 4, those occasions when n was infinity, we arranged for the block s/n to "hit the stops," namely, an arbitrary number of 50 dB. These large excursions in s/n dramatically underline those times when the system was unable to reject any samples and thereby effect a reduction in the transmitted data rate. In our calculations of segmental s/n[9] (shown in Fig. 5) we, of course, used the original input speech sequence and the recovered speech sequence at the output of the receiver. We note that those blocks for which n equaled infinity

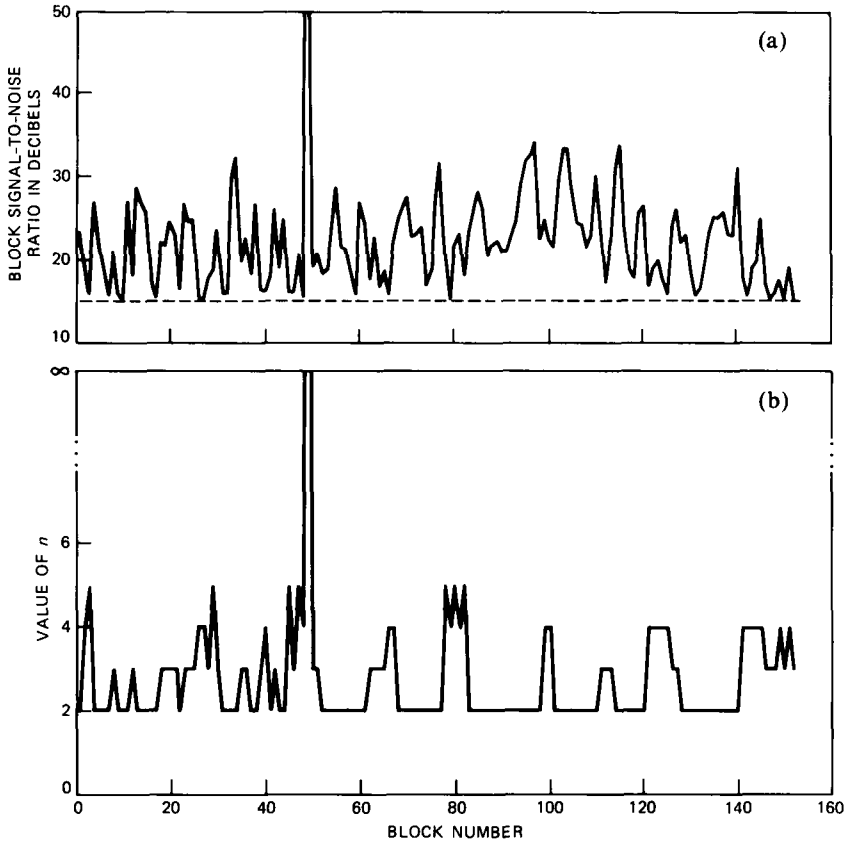

Fig. 4—Variation of block s/n and parameter n as a function of block number. (a) and (b) Block s/n and n, respectively, for s/n_{ref} of 15 dB. (Cont.)

had an s/n that was the signal-to-quantization noise ratio of the μ-law PCM speech.

Figures 4a, c, e, and g show the variation of block s/n, i.e., the $(s/n)_n$ for the final value of n used in the iteration, as a function of block number when s/n_{ref} was 15, 20, 25, and 30 dB, respectively. The corresponding profiles of the value of n used in the packetization process are displayed in Figs. 4b, d, f, and h, respectively. For s/n_{ref} of only 15 dB, $n = 2$ occurred most frequently, and only in one block was the system unable to reject any samples. By contrast, when $s/n_{ref} = 30$ dB, the guaranteed speech quality was so high that the most frequent decision was not to discard any speech samples in the block. This situation occurred because the block s/n of 8-bit, μ-law PCM was often below 30 dB.

Observe that for every 256 data words received, 128, 85, 64, 51, 42,

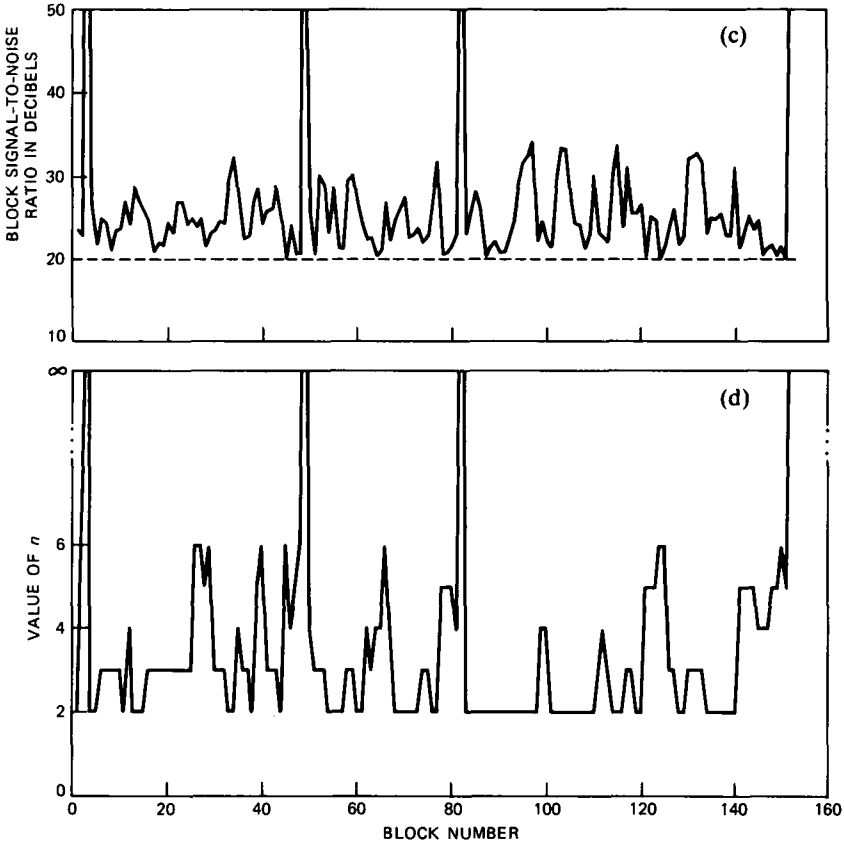

Fig. 4—(c) and (d) Block s/n and n, respectively, for s/n_{ref} of 20 dB. (Cont.)

and 0 words were discarded for transmitted packets associated with n of 2, 3, 4, 5, 6, and infinity, respectively. Clearly, the lower the value of n, the greater the data reduction. However, the value of n selected for a given block of W speech samples was dependent on s/n_{ref}, such that the lower the s/n_{ref}, the more probable the occurrence of a low value of n. Thus, as expected, the savings in the transmitted bit rate were at the expense of speech quality.

The histograms of n for different values of s/n_{ref} are displayed in Fig. 6. When a low recovered speech quality is acceptable, such as that obtained with s/n_{ref} = 15 dB, we found that $n = 2$ was the most frequently selected value of n and that $n > 4$ was rarely used. The reduction in the packetized data was found to be 41 percent. As the s/n_{ref} was increased, there was an increase in the frequency of occurrence of the higher values of n and a diminution in the rate at which lower n values occurred. The reduction in the data due to variable-

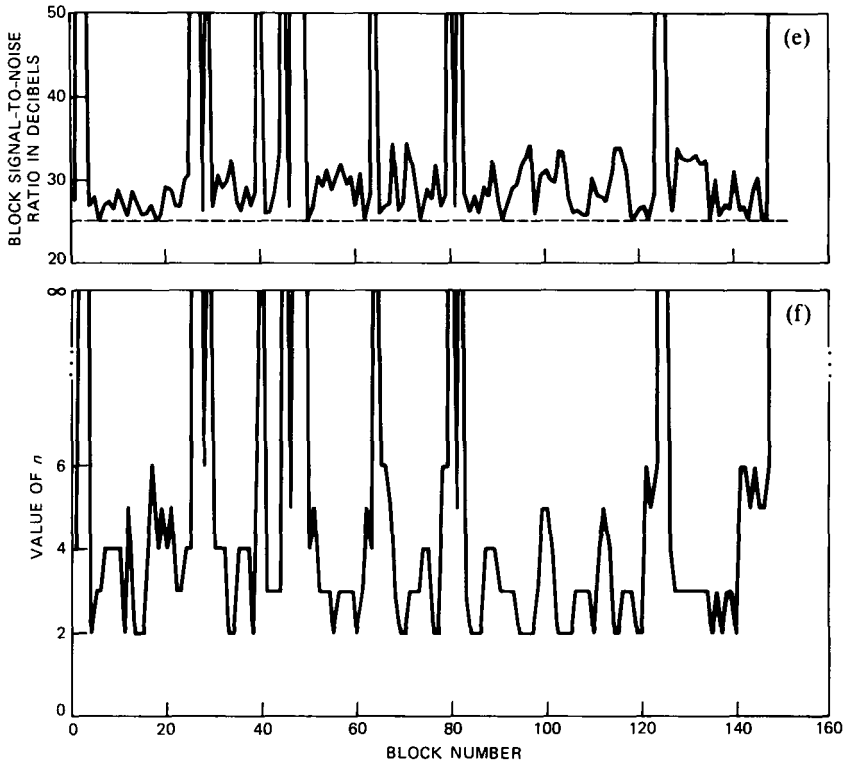

Fig. 4—(e) and (f) Block s/n and n, respectively for s/n_{ref} of 25 dB. (Cont.)

length packetizing was found to be 36, 28, and 20 percent for segmental s/n of 22.5, 25.5, and 33 dB, respectively. These reductions are displayed graphically in Fig. 5 as a function of our control parameter s/n_{ref}. Categorizing speech quality from segmental s/n is always contentious. However, when we included our informal listening experiences, we achieved a close approximation to toll quality speech with a data reduction of some 25 percent.

Figures 7a and b show a segment of the speech signal having a duration of 360 ms and its spectrogram. The corresponding error waveform and its spectrogram for 8-bit, μ-law PCM encoding are displayed in Figs. 7c and d. As expected with log-PCM, the quantization noise power was approximately proportional to the speech-signal power, and the error spectrum was relatively constant over the message band. When the variable-length packetization of the speech signal was performed, the error in the recovered speech signal was composed of the quantization noise and interpolation noise components. The error magnitudes were, therefore, increased as shown in Figs. 7e and f, where

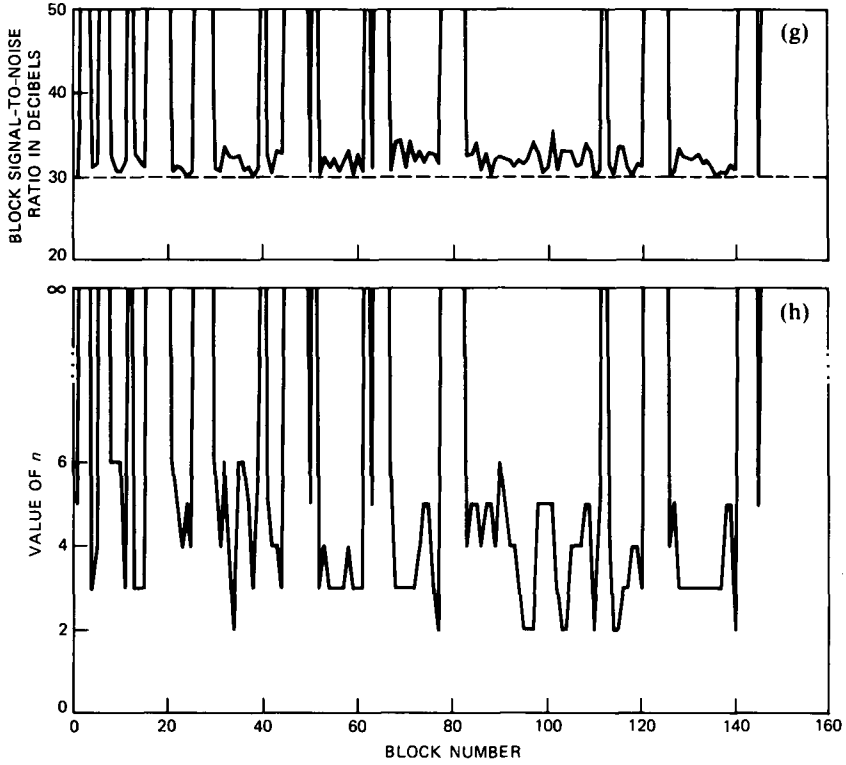

Fig. 4—(g) and (h) Block s/n and n, respectively, for s/n_{ref} of 30 dB.

s/n_{ref} = 25 dB. The spectral magnitudes in Figs. 7d and f have the same arbitrary units.

Figure 8a shows variations of the error signal for μ-law PCM over the entire speech signal, while the corresponding waveform for the recovered packetization speech is displayed in Fig. 8b. The waveforms of Figs. 8a and b are drawn to the same scale and provide a visual guide to the magnitude and location of the increase in error due to the packetization process. We observe that there is considerable correspondence between the variations in signal amplitudes in Figs. 8a and b. These variations also closely correspond to those in the original speech signal (not shown). This is to be expected, as the quantization noise is approximately proportional to the speech signal, and so is the interpolation noise because of the control exhibited by s/n_{ref}.

V. DISCUSSION

A variable-length packetization scheme has been proposed for 8-bit, μ-law PCM-encoded speech. By the aid of our control parameter

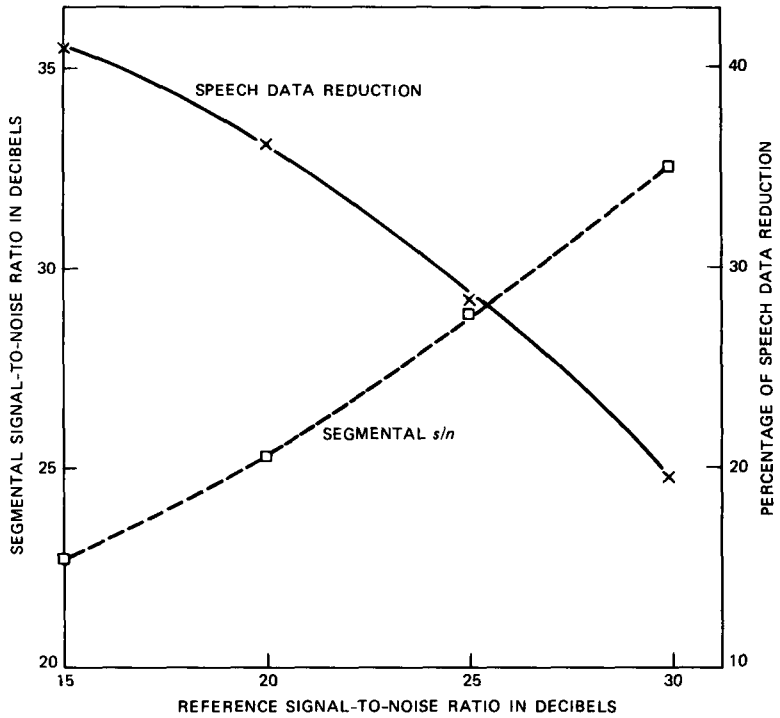

Fig. 5—Variation of segmental s/n, and percentage of speech data reduction, as a function of s/n_{ref}.

reference s/n we obtained a reduction in encoded speech data of 25 percent, while the segmental s/n exceeded 30 dB, i.e., the recovered speech displayed negligible perceptual impairments. An important feature of the system is that the s/n of the recovered speech in each packet is required to exceed an s/n_{ref}. Accepting s/n as a crude guide to quality, particularly for the relatively high s/n values employed here, our system endeavors to generate packets of variable length such that when they are decoded the speech will be maintained above a certain quality.

In the scheme proposed here the length of the packets varies. The number of μ-law PCM words in the packets could be either 256, 214, 205, 192, 171, or 128, where the original block of speech contained 256 words. While these packets can be transmitted as variable length with suitable headers and flags, they can also be transmitted as fixed-length packets, where the discarded speech data can be replaced with other types of data, such as that arising from computer traffic.

In our proposal the packet lengths are determined for a given speech signal by the selection of the s/n_{ref}, which acts as a quality control

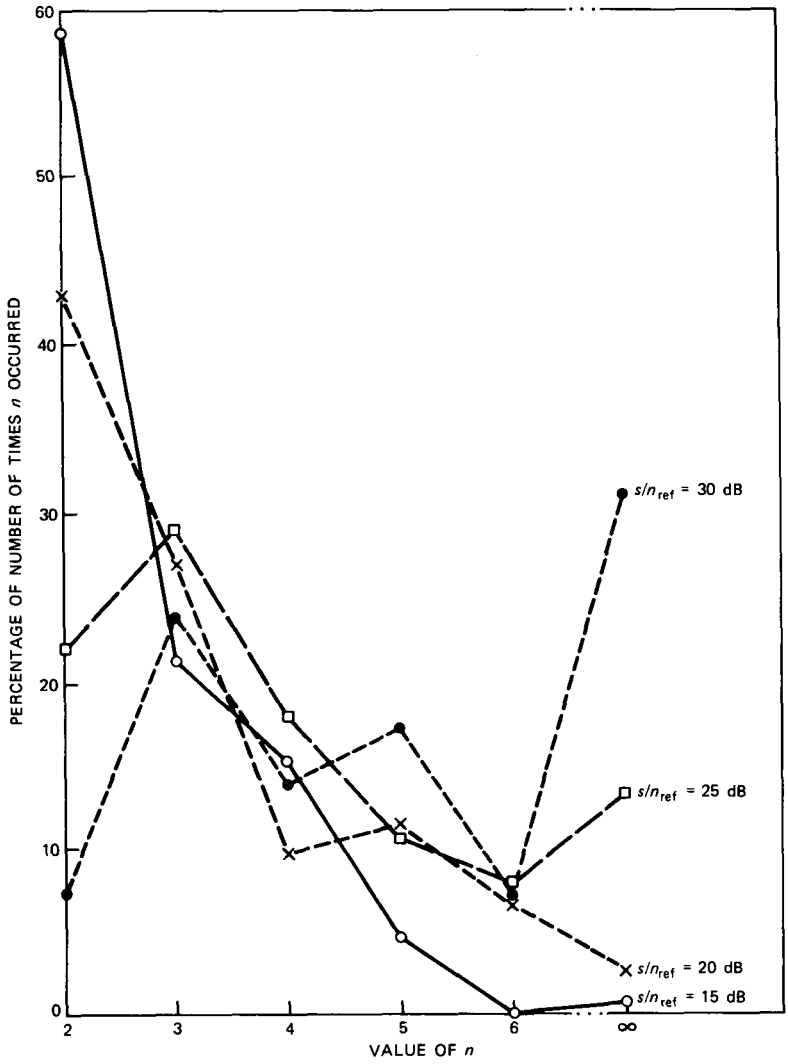

Fig. 6—Variation of the percentage of time the parameter n was used as a function of n for different s/n_{ref}.

mechanism. If the multiplexer is allowed to vary the s/n_{ref}, it can determine the length of the packets required for multiplexing. Thus, as more packets access the highways, the multiplexer control can discard more μ-law PCM words. Because the selection of those words to be discarded for a given s/n_{ref} is related to the interpolation algorithm, the degradation in recovered speech quality with increased user capacity is relatively smooth.

The packetization process need not be at the user's interface or

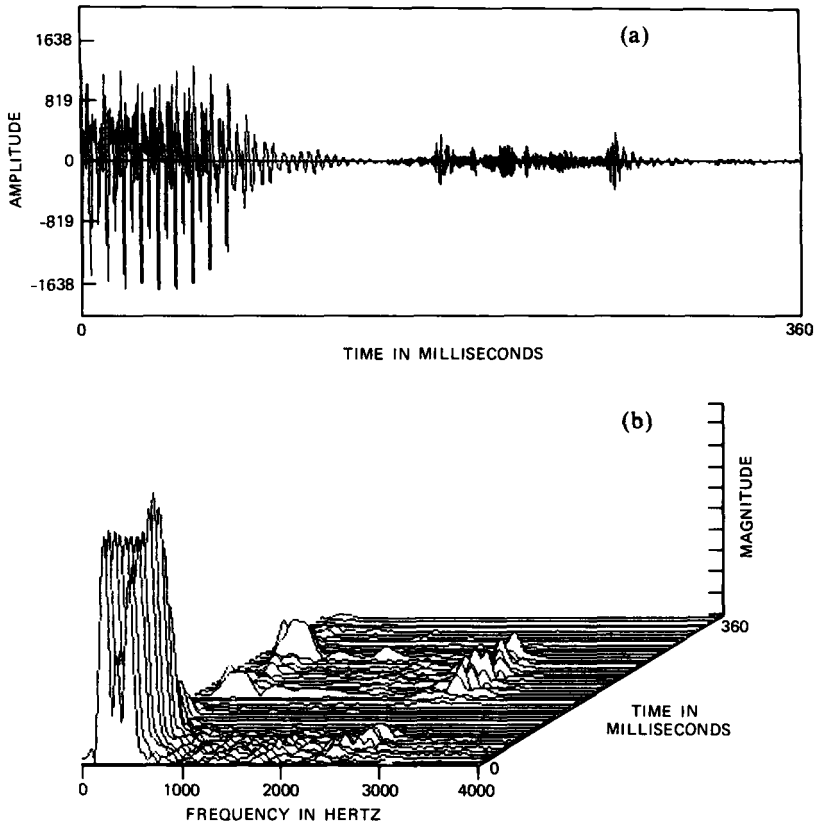

Fig. 7—Speech waveforms and spectrograms. (a) and (b) Segment of the speech signal and its spectrogram. (Cont.)

terminal; it can be located at a node in the network converting successive groups of W μ-law PCM words into variable-length packets. Thus, time-division-multiplexed μ-law PCM channels can be statistically multiplexed using the packetization scheme and thereby achieve a significant diminution in the data rate. Alternatively, a limited-capacity output port from the network node can accommodate more μ-law PCM traffic by employing the packetization procedure.

However, packetizing μ-law PCM at a node in the network requires a different s/n_{ref} compared with when the packetization occurs at the user's interface. In Fig. 2 we observe that the block s/n, namely $(s/n)_n$, is computed using $\{\hat{z}_k\}$ and $\{\tilde{x}_k\}$. The packetization equipment at a node in the network would not have access to the original speech sequence $\{\tilde{x}_k\}$ and would have to employ $\{x_k\}$. In this situation, when no samples are discarded in a block, i.e., n is infinity, $(s/n)_n$ would also be infinity because the sequences $\{\hat{z}_k\}$ and $\{x_k\}$ are identical. Thus,

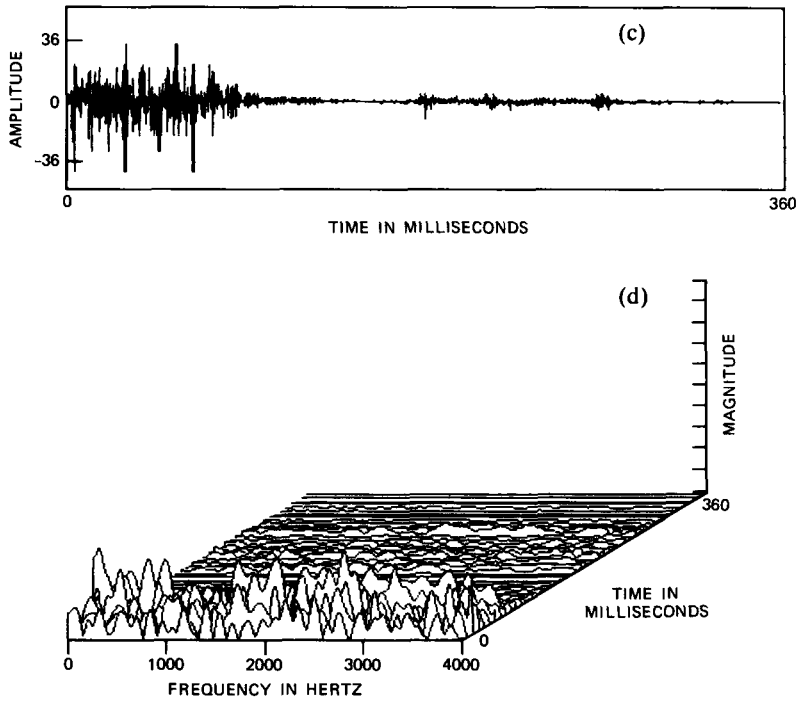

Fig. 7—(c) and (d) Corresponding error waveform and its spectrogram for 8-bit, μ-law PCM-encoded speech, $\mu = 255$. (Cont.)

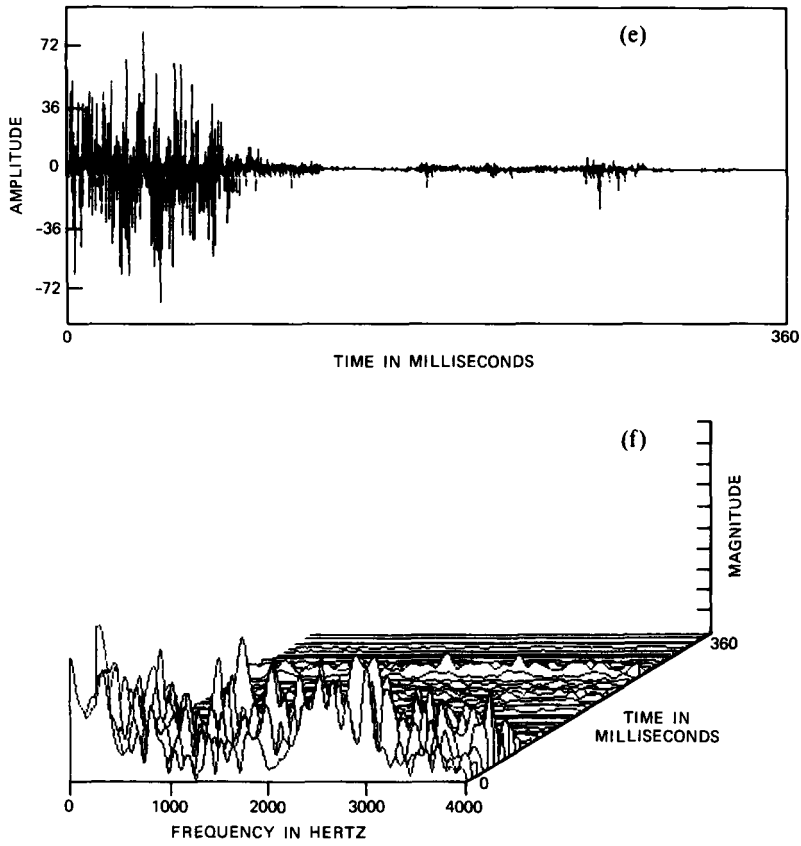

Fig. 7—(e) and (f) Corresponding error waveform and its spectrogram when the variable-length packetized speech system was employed.

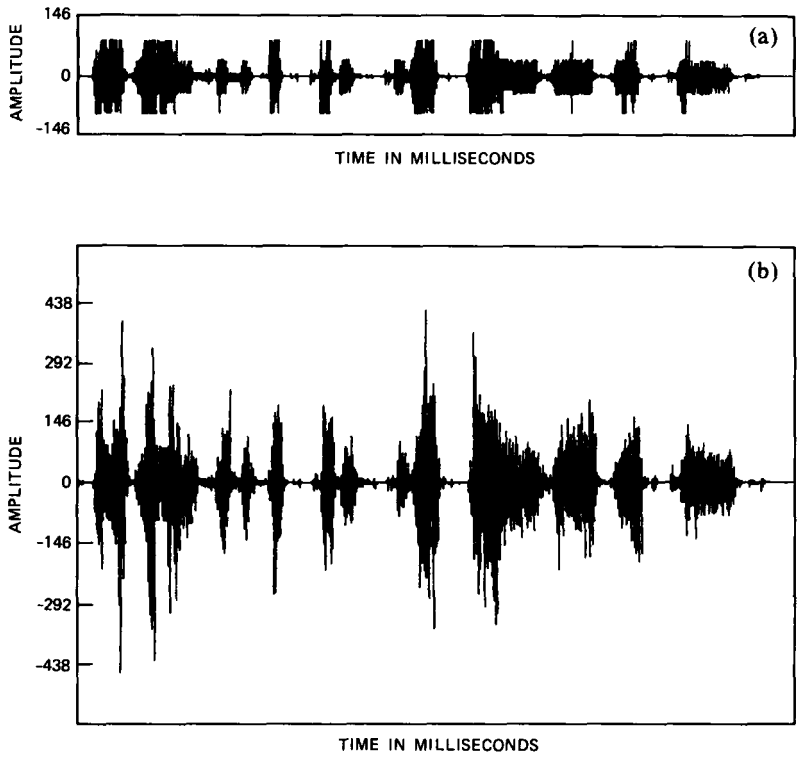

Fig. 8—Error waveforms for the speech signal. (a) Eight-bit, μ-law PCM speech. (b) Variable-length packetized speech.

$(s/n)_n$ computed at a node in the network is the signal-to-interpolation noise ratio, and for a given n this ratio is higher than that obtained for the arrangement shown in Fig. 2. Consequently, if s/n_{ref} is specified in terms of packetizing at the subscriber's interface, a higher value of s/n_{ref} is required when the packetization occurs at a node in the network. The important point to note, however, is that s/n_{ref} is a control parameter, and it is a simple procedure to increase it in order to achieve the required minimum block s/n performance, or a required average value of n.

Finally, we note that controlling packet length with spectral distance measures, and other measures more closely related to perception, should reduce the average transmitted bit rate per channel, but at the expense of added complexity.

VI. ACKNOWLEDGMENT

The authors are grateful to D. J. Goodman for his constructive criticism of this work.

REFERENCES

1. J. L. Flanagan, M. R. Schroeder, B. A. Atal, R. E. Crochiere, N. S. Jayant, and J. M. Tribolet, "Speech Coding," IEEE Trans. Commun., *COM-27* (April 1979), pp. 710–37.
2. B. G. Haskell and R. Steele, "Audio and Video Bit-Rate Reduction," Proc. IEEE, *69*, No. 2 (February 1981), pp. 252–62.
3. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Englewood Cliffs, N.J.: Prentice-Hall, 1984.
4. J. S. Turner and L. F. Wyatt, "A Packet Network Architecture for Integrated Services," IEEE Globecom '83, San Diego, Calif., November 18–December 1, 1983, pp. 45–50.
5. K. W. Cattermole, *Principles of Pulse Code Modulation*, London: Iliffe, 1969.
6. CCITT Study Group XVIII, Draft Recommendation G.7zz, "32 kbits s Adaptive Differential Pulse Code Modulation (ADPCM)," Geneva, November 21–25, 1983.
7. R. Steele, *Delta Modulation Systems*, London: Pentech Press, 1975.
8. R. Steele and F. Benjamin, "Sample Reduction and Subsequent Adaptive Interpolation of Speech Signals," B.S.T.J., *62*, No. 6 (July–August 1983), pp. 1365–98.
9. R. Steele and D. Vitello, "Simultaneous Transmission of Speech and Data Using Code-Breaking Techniques," B.S.T.J., *60*, No. 9 (November 1981), pp. 2081–105.

AUTHORS

Frank Benjamin, B.A. (Music Education), 1980; B.S. (Electronic Engineering), 1983, Valedictorian (both cum laude), Monmouth College, West Long Branch, N.J.; AT&T Bell Laboratories, 1981–1983; Perkin-Elmer Corp., 1983—. In 1980 he was appointed at Monmouth College as laboratory instructor in the Electronic Engineering Department and served as acoustical consultant to the Department of Physics on a publication for General Physics (Dr. Robert Smith, Chairman/author). Later in 1980 he joined United Telecontrol Electronics, Asbury Park, N.J., where he helped develop a missile guidance control system for the Navy. In 1981 he joined AT&T Bell Laboratories as a member of the Communications Methods Research Department, Crawford Hill, Holmdel, N.J., developing and writing software simulation systems for speech companding and interpolation techniques. In 1983 Mr. Benjamin joined Perkin-Elmer Corporation/Data Systems Group as a Member of Technical Staff hardware/software engineer in the Computer Processor/Memory Development Department, where he is currently designing and developing processor diagnostic equipment and computer graphics systems. President, college honor society Lambda Sigma Tau (1979–81). Member, national engineering society Eta Kappa Nu; national physics and mathematics honor societies. Nominee, Who's Who Among Students in American Colleges and Universities. New Jersey State Teacher's Certificate.

Raymond Steele, B.Sc. (Electrical Engineering), 1959, Durham University, Durham, England; Ph.D. (Delta Modulation Systems), 1975, and D.Sc. (Digital Encoding, Methods for Combatting Transmission Errors, and other Communication Techniques), 1983, Loughborough University of Technology, Loughborough, England. Prior to his enrollment at Durham University, he was an indentured apprenticed Radio Engineer. After research and development posts at E. K. Cole Ltd., Cossor Radar and Electronics Ltd., and The Marconi Company, all in Essex, England, he joined the lecturing staff at the Royal Naval College, Greenwich, London, England. Here he lectured in telecommunications to NATO and the External London University degree courses. His next post was as Senior Lecturer in the Electronic and Electrical Engineering Department of Loughborough University, Loughborough, Leics., England, where he directed a research group in digital encoding of speech and

television signals. In 1975 his book, *Delta Modulation Systems* (London: Pentech Press), was published. He was a consultant to the Acoustics Research Department at Bell Laboratories in the summers of 1975, 1977, and 1978, and in 1979 he joined the company's Communications Methods Research Department, Crawford Hill Laboratory, Holmdel, N.J. In 1983 he became Professor of Communications in the Electronics Department at the University of Southampton, England and nonexecutive Director of Plessey Electronic Systems Research Limited, England. Senior member, IEEE. Member, IEE.