THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 7, September 1979
Printed in U.S.A.

Statistical Block Protection Coding for DPCM-Encoded Speech

By R. STEELE, N. S. JAYANT, and C. E. SCHMIDT

(Manuscript received January 3, 1979)

Blocks of speech-carrying DPCM bits are protected from transmission errors by means of explicit communication of two block statistics—the maximum and the root-mean-square (rms) values of the adjacent-sample differences in the DPCM-quantized speech. At the receiver, the maximum value is used as a cue for error-detection, while the rms value is used for a partial waveform correction procedure that provides intelligible speech at bit error rates as high as 10 percent.

I. INTRODUCTION

Block protection coding, whereby a block of data words is protected by the addition of special code words or letters, is a common feature in communication systems for noisy channels. In algebraic error detection and correction, for example, the protection is derived from parity check bits. The number of parity checks, and hence the redundancy, increases with the number of data bits protected, but the resulting error-coding procedures are quite general, being applicable to any type of data, irrespective of its source. Nevertheless, with sources such as speech, where it is not crucial to recover every speech-carrying bit without error, it is meaningful to look for certain special, compact forms of *non-algebraic* block protection. The idea is to transmit a protection word that identifies some perceptually significant parameter of a speech-waveform segment; knowing the (correct) value of this parameter, the receiver can perform error-detecting and error-correcting operations, which may be only partial in an algebraic sense (due to the compactness of the protecting procedure) but nevertheless quite adequate from a speech-perception viewpoint.

In one recent investigation[1] along these lines, each block of differential PCM words was protected by a reference PCM word that signified the speech amplitude at the end of the block. Error detection was based on comparing the DPCM decoder amplitude at the end of the block with PCM reference. Procedures for locating (and correcting) errors within the block were simple for single errors, but fairly involved for multiple errors in a block. A very successful error-location procedure had however been noted in an earlier investigation;[2] this depended on the detection of a statistically unlikely change between adjacent samples in the corrupted speech signal, relative to the root-mean-square (rms) value of these differences measured over a suitably long block containing these samples. The rms parameter in Ref. 2 was obtained from the corrupted speech, and this affected the success of the procedure at high error rates (say, 5 percent or higher).

The scheme to be described in this paper recognizes and extends the statistical notions of Ref. 2 and incorporates them in a block protection system that is effective even at error rates as high as 10 percent. This statistical block protection coding (SBPC) system is discussed for the specific case of non-adaptive DPCM, but extension to an adaptive system should be possible, at least in cases where the (step size) adaptation is slow or syllabic.*

II. STATISTICAL BLOCK PROTECTION CODING (SBPC)

The SBPC system employs a simple protection code consisting of two words which represent:

(*i*) The maximum difference between adjacent locally decoded speech samples within the block of W samples.

(*ii*) The rms value of the differences between adjacent locally decoded speech samples within the block of W samples. Notice that the extremal statistic (*i*), together with the central statistic (*ii*), constitute a partial description of the PDF (probability density function) of first differences.

2.1 Transmitter

The arrangement of the DPCM encoder and the system for generating the protection code are shown in Fig. 1. Suppose that the mth block of speech samples is being processed. The input speech sample x_{mW+r}, corresponding to the rth instant in the mth block, is encoded into a quantized sample q_{mW+r} by a DPCM encoder using a uniform quantizer. The predictor is of first order, with a coefficient value of $LK < 1$. Z^{-1} represents a delay of one sample period.

Denoting the locally decoded speech sample by y_{mW+r}, the protection code words are defined in the form

* Recent studies have shown that our technique works quite well in conjunction with an adaptive procedure where the quantizer step size is constant within a block (several milliseconds or tens of milliseconds long) of samples, but is modified once at the beginning of each block in response to changing speech level.

$$d_{\max} = \max_{2 \le r < W} | y_{mW+r} - y_{mW+r-1} |_{Q2} \tag{1}$$

$$d_{\text{rms}} = \left[\frac{1}{W-1} \sum_{r=2}^{W} (y_{mW+r} - y_{mW+r-1})^2_{Q2} \right]^{1/2}. \tag{2}$$

The quantizer Q2 has the same number of levels as the DPCM quantizer Q1, but is arranged to quantize only positive samples. Thus, after multiplexing $d_{\max}$, d_{rms} and W DPCM words, the frame consists of $(W + 2)$ n-bit words. It is important, or at least very desirable, to protect the "protecting words," $d_{\max}$ and d_{rms}, by transmitting them in a redundant format. For example, one might transmit three versions

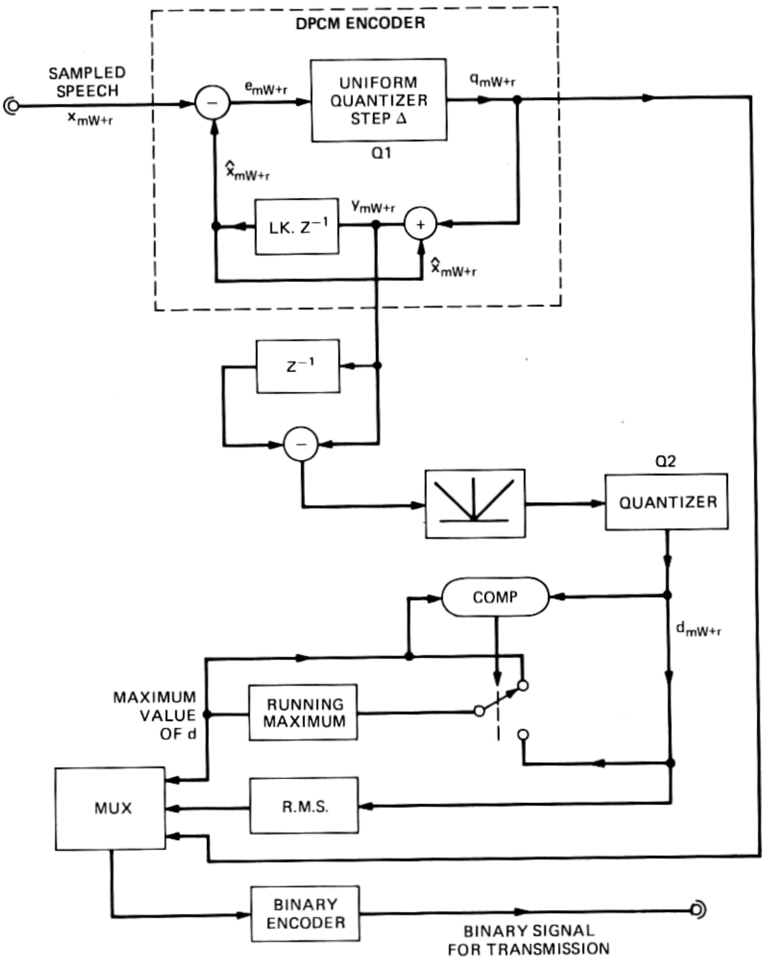

Fig. 1—SBPC encoder.

of each bit in the protecting word and decode the bit on the basis of a majority count. The overhead constituted by this protection arrangement would be $3 \times 2 = 6$ words, or 2.3 percent if $W = 256$; this overhead is much smaller than the redundancy required of an algebraic code that would correct some patterns we will discuss later in this paper.

The DPCM-encoded speech together with its simple protection code is transmitted through a channel which may cause some bits to be inverted. The probability of bit inversion is called the error rate *ER*.

2.2 Receiver

The receiver demultiplexes each frame into its data block and protection code. The DPCM sequence is decoded into Y_{mW+k}; $k = 1, 2, \cdots, W$. (Note that cap letters Y and D will be used to signify variables at the receiver.) Figure 2 shows the essential features of the SBPC

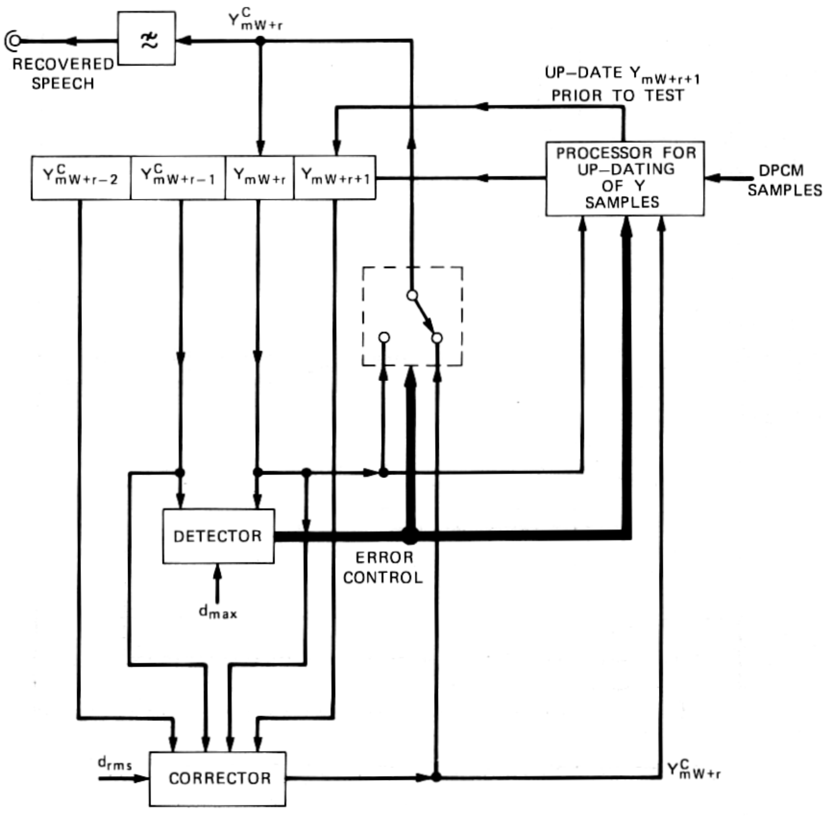

Fig. 2—SBPC decoder.

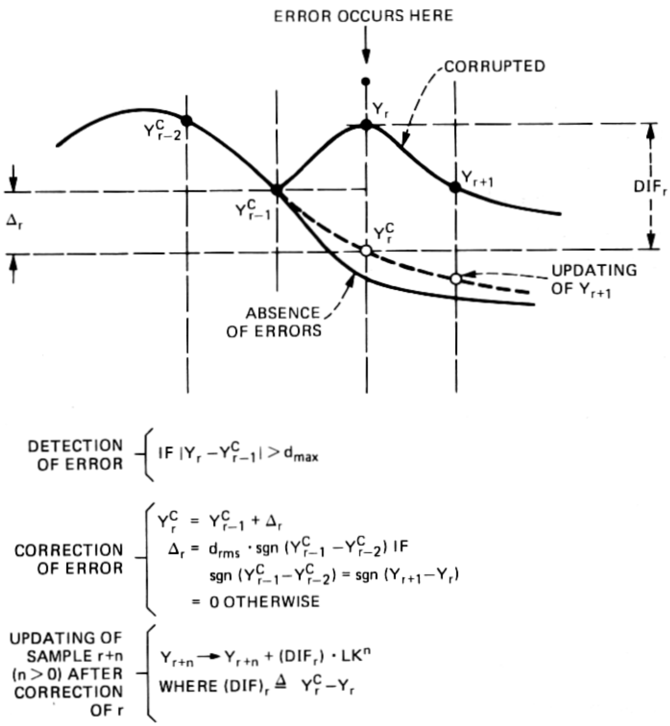

Fig. 3—Waveform correction at time r.

correction procedure. For simplicity, the demultiplexer decoders and control facilities have been omitted.

We suppose that samples $\cdots$, Y^C_{mW+r-3}, Y^C_{mW+r-2}, Y^C_{mW+r-1}, have been either considered correct and passed to the output, or deemed to be in error and partially corrected; hence the superscripts C. We now test sample Y_{mW+r}. We find the quantized magnitude difference D_{mW+r} between Y_{mW+r} and Y^C_{W+r-1}, and compare the difference with the maximum transmitted difference $d_{\max}$. Y_{mW+r} must be erroneous if

$$D_{mW+r} > d_{\max}. \tag{3}$$

If inequality (3) is satisfied, the correction must be switched into the circuit and the erroneous Y_{mW+r} replaced by a corrected value Y^C_{mW+r}.

The corrections are described by the algorithm (Fig. 3):

$$Y^C_{mW+r} = Y_{mW+r-1} + \Delta_{mW+r}, \tag{4}$$

where

$$\Delta_{mW+r} = d_{\text{rms}} \operatorname{sgn}\,(Y^C_{mW+r-1} - Y^C_{mW+r-2}) \tag{5}$$

$$\text{if } [\operatorname{sgn}(Y^C_{mW+r-1} - Y^C_{mW+r-2}) = \operatorname{sgn}(Y_{mW+r+1} - Y_{mW+r})]$$

$$= 0 \text{ otherwise.} \qquad (6)$$

Clearly, this correction is based on a "smooth" output waveform model where the sign of the slope at time r equals that at time $r - 1$ if the latter equals that at time $r + 1$ [eq. (5)], while, if the slopes at times $r - 1$ and $r + 1$ are opposite in sign, that at time r is given the average value of zero [eq. (6)]. Furthermore, in the former case, the magnitude of the slope at time r is set to the block-specific rms value d_{rms}. Strictly speaking, the optimum setting of this magnitude would take the form $J \cdot d_{\text{rms}}$, where J would be a constant depending on the shape of the first-difference PDF.

The correction algorithm has also been deployed[3] with d_{rms} being derived from the corrupted speech. With large values of the error rate ER, this would give rise to poor corrections. By explicitly transmitting the value of d_{rms}, the corrections are significantly improved.

2.3 Updating samples following a correction

Having made the correction to sample Y_{mW+r}, we remove the error from the subsequent samples before we continue testing the next sample Y_{mW+r+1} (Figs. 3 and 4). This is done as follows. Let

$$DIF_{mW+r} = Y_{mW+r} - Y^{C}_{mW+r}. \qquad (7)$$

As the propagation of the error is due to the integrator, and the integrator leakage factor is LK, the subsequent decoded samples are reduced to

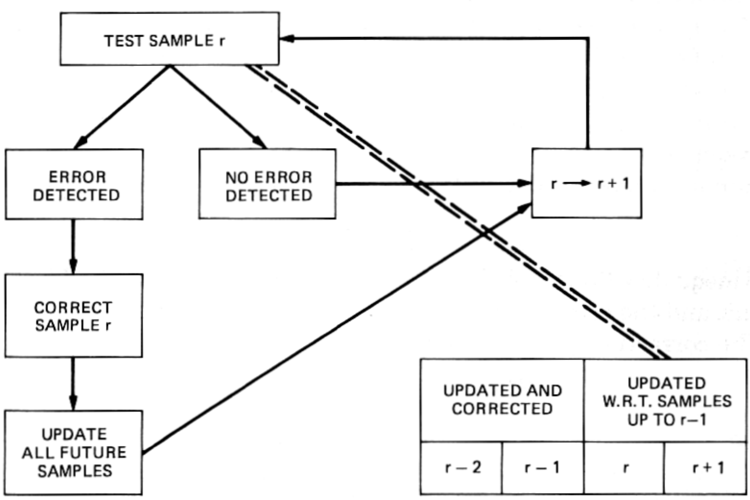

Fig. 4—Error detection, correction and sample updating.

$$Y_{mW+r+n} = Y_{mW+r+n} - (LK)^n DIF_{mW+r}$$
$$n = 1, 2, \cdots, W - r. \tag{8}$$

The value of the error at the end of the block is

$$E_{(m+1)W} = (LK)^{W-r} DIF_{mW+r}, \tag{9}$$

and this is stored.

For each correction, the propagation of the error is removed leaving a residue at the end of the block whose size depends on the position of the sample corrected, as shown by eq. (9). These residuals are summed to give the total residual propagation error $E^{(T)}_{(m+1)W}$. When the next block of DPCM samples are processed, each one is modified to remove the propagation effects from errors in the previous block:

$$Y_{W(m+1)+r} = Y_{W(m+1)+r} - (LK)^r E^{(T)}_{(m+1)W}$$
$$r = 1, 2, \cdots, W. \tag{10}$$

The detection and correction method epitomized by eqs. (3) and (8) are again used to process the DPCM samples in eq. (10).

2.4 Organization of the transmission block

The first part of the transmission frame contains the protection code. It is placed there in order for the detection and correction process to begin immediately. When testing Y_{mW+1}, samples Y_{mW} and Y_{mW-1} from the previous block must be available. When testing the last sample, $Y_{(m+1)W}$, the first sample $Y_{(m+1)W+1}$ from the next data block is required. Consequently, the total delay of the decoded speech is $(W + 1)$ sampling intervals.

The larger the value of W, the smaller the fractional increase in required channel capacity (due to the protection-word overhead), but the longer the decoding delay at the receiver output.

III. RESULTS AND DISCUSSION

The block protection scheme was simulated on a Data General Eclipse computer. The band-limited input signal, a single sentence spoken by a male, was sampled at 8 kHz prior to encoding by a uniform 7-bit DPCM encoder with predictor coefficient $LK = 0.9$. The coding of the quantizer output levels was such that an error in the most significant bit caused an error in the received sample equal to half the range of the quantizer.

The DPCM code words were assembled into blocks of W words with the protection code previously described. The DPCM code words were subjected to random errors, but the protection code words were left uncorrupted.

As a supplement to listening tests, the segmental signal-to-noise ratio,[4]

$$\text{SNRSEG} = \text{Average [short-time SNR(in dB)]}, \qquad (11)$$

was used, an objective performance criterion. The short-time SNR in (11) is a statistic computed over an interval typically 16 to 32 ms long. By performing the decibel operation prior to long-time averaging, the SNRSEG measure preserves information about how well the low-level segments of speech are reproduced.

Figure 5 shows the variation of SNR as a function of amplitude scaling AS of the imput speech signal. From the zero error rate curve, it can be seen that optimum loading occurs for $AS = 0.04$. When $ER = 4.2$ percent, the decoded signal is very corrupted and SNRSEG is reduced by 40 dB in the underloaded condition. However, the SBPC system dramatically improves the performance of the DPCM system, increasing the SNRSEG by 11 dB for $AS = 0.04$, and by 19 dB for $AS = 0.01$.

The unusual characteristic of the SBPC system is that, with large values of ER, the variation of SNRSEG is substantially independent of AS. This is a property found in adaptive DPCM. The reason for the nearly flat SNR characteristic is: In the presence of low level speech, d_{max} is a low number, and if many errors occur there will be numerous occasions when the differences between adjacent samples in the corrupted decoded signal exceed d_{max}. These erroneous differences are identified and will be partially corrected. Only those errors which

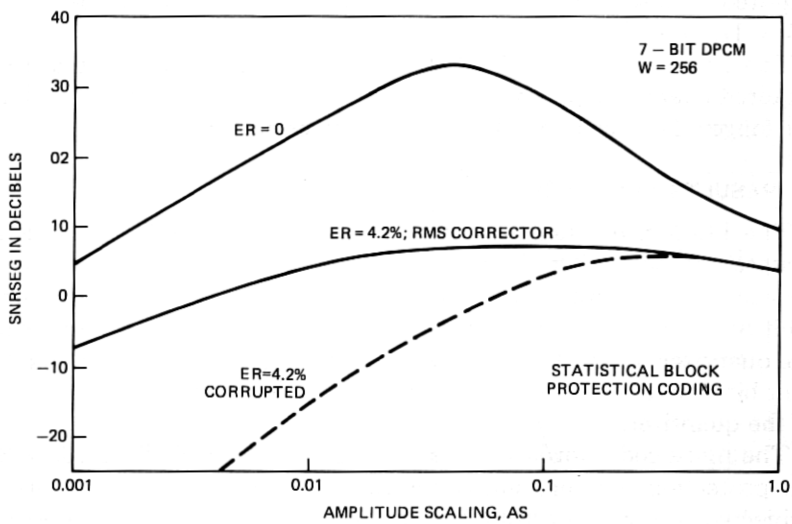

Fig. 5—SBPC gain as a function of input speech level AS.

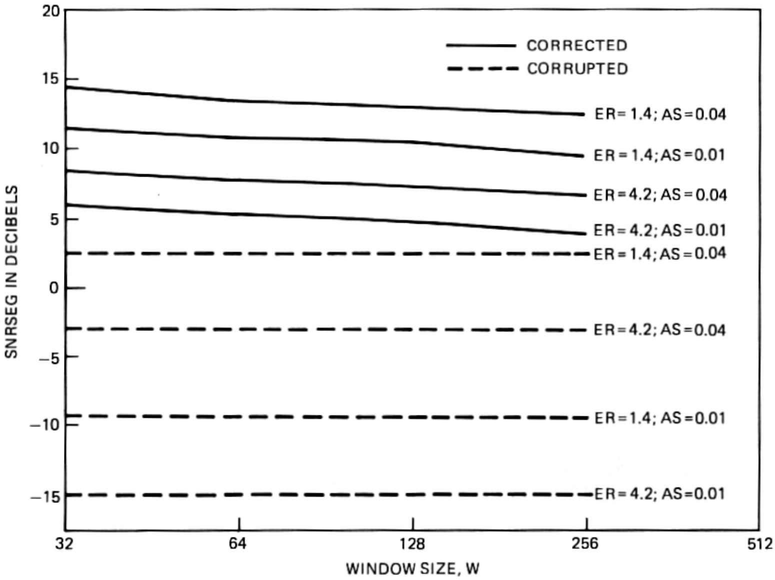

Fig. 6—SBPC gain as a function of block length W.

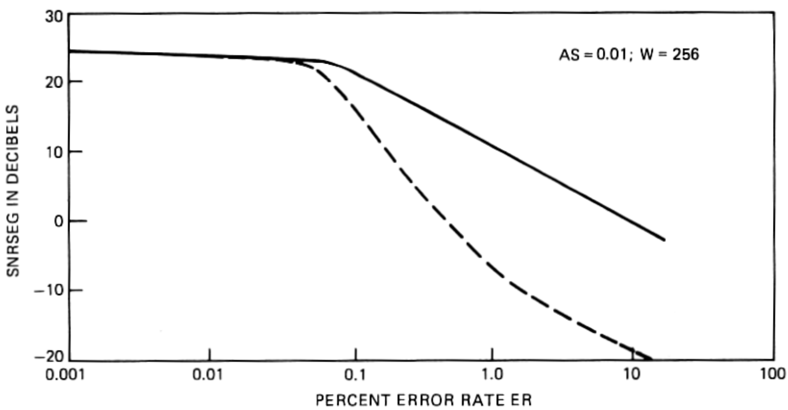

Fig. 7—SBPC gain as a function of error rate ER.

result in differences less than $d_{\max}$ will be missed. However, when the coder is occasionally experiencing some overloading ($AS > 0.04$, say), the maximum value $d_{\max}$ in some blocks will merely reflect quantizer saturation, rather than providing a cue for detecting transmission errors, and improvements are now gained only in purely unvoiced or silent intervals in speech.

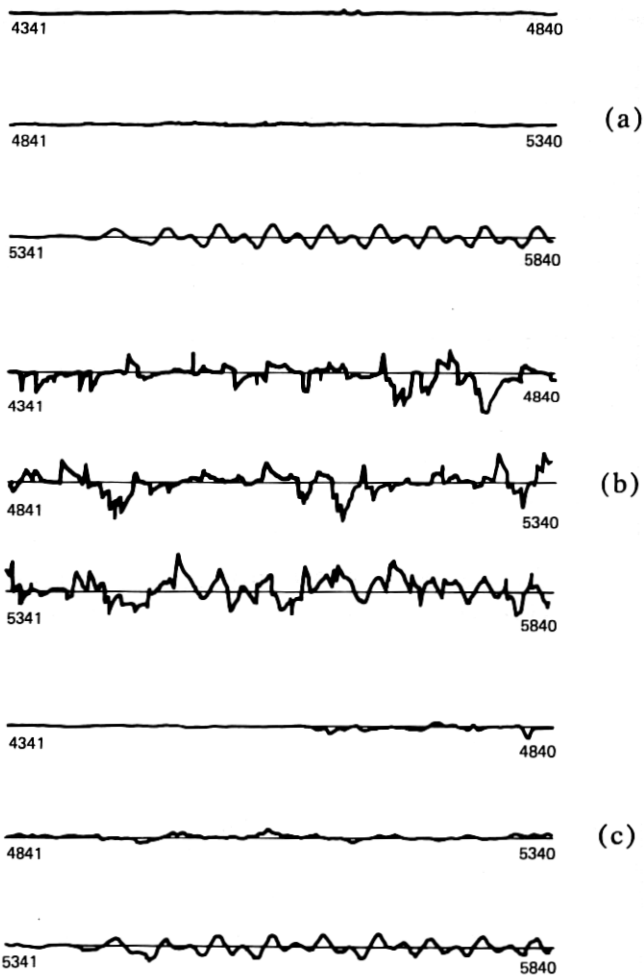

Fig. 8—Waveforms of (a) original, (b) corrupted, and (c) corrected speech (ER = 10%).

At AS = 0.01 and ER = 4.2 percent, the corrupted speech is perceptually very poor, and sounds almost like bandlimited white noise. By using the SBPC system, the speech is rendered intelligible, although of poor quality. The overall perceptual improvement is dramatic.

The variation of SNRSEG as a function of block size W is shown in Fig. 6. Increasing W from 32 to 256 results in a decrease in SNRSEG of less than 2 dB. The near-independence of SNR from W is perhaps related to the fact that none of the W values used is large enough to encompass a significantly nonstationary segment of speech.

The gain is SNRSEG as a function of ER is shown in Fig. 7 for $AS = 0.01$, $W = 256$. The objective gain is very slight for low error rates (say, $ER < 0.1$ percent), but significant for high error rates (say, $ER > 0.5$ percent). In particular, the gains are quite dramatic with $ER = 10$ percent. These objective gains are well reflected by the perceptual gains noticed in informal listening tests, and by the illustrative speech waveforms in Fig. 8.

REFERENCES

1. R. Steele, D. J. Goodman, and C. A. McGonegal, "DPCM with Forced Updating and Partial Correction of Transmission Errors," B.S.T.J., *58*, No. 3 (March 1979), pp. 721–728.
2. R. Steele and D. J. Goodman, "Detection and Selective Smoothing of Transmission Errors in Linear PCM," B.S.T.J., *56*, No. 3 (March 1977), p. 399.
3. R. Steele, D. J. Esdale, and D. J. Goodman, "Partial Correction of Transmission Errors in DPCM Without Recourse to Error Correction Coding," Elec. Lett., *13*, No. 12 (June 9, 1977), p. 351.
4. P. Noll, "Adaptive Quantizing in Speech Coding Systems," International Zurich Seminar, Zurich, Switzerland, April 1974.