THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 9, November 1975
*Printed in U.S.A.*

# An Autocorrelation Criterion for the Time-Diversity Reception of Speech Over Burst-Error Channels

By N. S. JAYANT

(Manuscript received March 25, 1975)

*This paper proposes a new approach to signal selection in time-diversity systems. Specifically, we consider the problem of digital speech transmission over a burst-error channel using two-channel time-diversity reception.*

*Let every speech segment (of length $W$) be transmitted twice so that at least one of the transmissions escapes an error burst, with a certain useful probability. Let the received speech segments be $Y_1$ and $Y_2$. We propose an autocorrelation-maximizing signal selection procedure of the following form.*

*Select $Y_1$ (or $Y_2$) as the "cleaner" speech segment according as $C(Y_1, W) \geqq$ (or $<$) $C(Y_2, W)$, where*

$$C(Y_u, W) = \sum_{r=2}^{W} (\text{sgn } Y_{ur} \cdot \text{sgn } Y_{u(r-1)})/(W - 1);\ u = 1, 2.$$

*$Y_{ur}$ is the speech amplitude at sample $r$ in $Y_u$, $W$ is a computational window that is typically a few milliseconds long, and sgn $Y_{ur}$ is a polarity function that is assumed to have zero mean and unit variance.*

*The use of sgn $Y_{ur}$ instead of $Y_{ur}$ leads to a simply implemented selection procedure, and computer simulations have demonstrated its practical utility. For example, in one study of three-bit DPCM coding, autocorrelation-based burst-error detection proved to be more useful than a procedure where DPCM samples were error-protected on a bit-by-bit basis, rather than in blocks.*

## I. THE BURST-ERROR PROBLEM

The research reported in this paper was motivated by the problem of digital speech communication over a mobile radio channel. Signal transmissions over such a channel are characterized by multipath fading. The fading is "slow" in the sense that a given fade (signal strength below a specified threshold) can last for several tens of milliseconds (which will typically involve several tens or several hundred speech bits). The end effect of these "slow" fades on digital transmissions is to introduce bursts of errors in the reception of speech-carrying bits.

The time statistics of these error bursts are illustrated by the distribution functions in Fig. 1. $D$ is the error-burst duration and $I$ the error-free interval between successive bursts. An error burst is defined to have a local error probability of 1. In other words, a burst of length $D_0$ implies that $D_0$ contiguous speech bits are in error. An isolated error, for example, is an error burst of length $D_0 = 1$. The curves refer to a subsegment from a bit-error sequence whose average error probability was 0.06. Note that the local error probability in Fig. 1 [the ratio of $D_{\text{average}}$ to $(D_{\text{average}} + I_{\text{average}})$] is 0.048. Notice also that $I_{\text{average}} \gg I_{\text{median}}$, suggesting a long tail in the interval distribution.

The error sequence was obtained from a fading simulator,[1] and it represents the impairment for a 24-kb/s signal-bit stream (the bit duration determines the number of bits affected by a fade) when the mobile radio link is characterized by two-branch diversity reception under the following (worse than average) conditions:

$$\begin{aligned} \text{Signal-to-interference ratio} &= 6\text{ dB} \\ \text{Signal-to-noise ratio} &= \infty \\ \frac{\text{vehicle speed}}{\text{radio wavelength}} = \frac{V}{\lambda} = \frac{29\text{ mi/h}}{0.353\text{ m}} &= 36.2\text{ Hz.} \end{aligned} \tag{1}$$

A companion paper[2] provides a somewhat more elaborate discussion of signal fades and error bursts.

## II. TIME-DIVERSITY CODING

The temporal structure of clustered errors can be exploited in redundant transmission schemes where message units are repeated with an appropriate time spacing. The optimum time spacing is, in general, a function of the error statistics. For example, the spacing can be designed to minimize the probability that both of two consecutive transmissions of a given message unit are affected by an error burst or bursts. The message unit can be a block of speech-amplitude samples, or a single bit from a digital speech code, and so on.

A recent proposal discusses the use of time diversity for three-bit DPCM transmissions over mobile radio.[2] Briefly, redundancy is introduced in the form of three transmissions of the most significant (sign) bit $B_1$ in a DPCM word and two transmissions of the second most significant (magnitude) bit $B_2$. The average redundancy is therefore 100 percent. The receiver decodes the sign bit $B_1$ on the basis of a majority

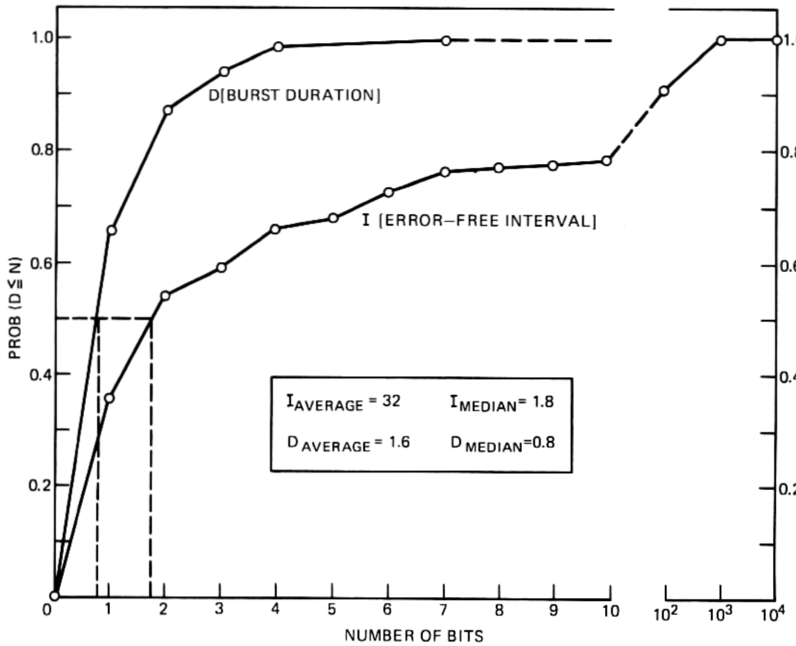


Fig. 1—Time statistics of burst errors [$P(E) = 0.06$].

count (over the three received versions). It also looks for unanimity between the two received versions of the significant magnitude bit $B_2$. If the unanimity does not exist, the DPCM word is forced to its minimum possible magnitude. When the spacings between the repetitions of DPCM bits are properly designed, the technique provides a significant advantage over nonredundant DPCM.[2] We comment again on this procedure at the conclusion of Section IV.

The purpose of this paper is to propose a different approach to time diversity. The method is based on error-burst detection (rather than single-error correction, as in a successful majority count); and the message unit that is error-protected is a block of contiguous speech amplitudes, rather than a basic speech-carrying DPCM bit or word. The idea of protecting message blocks using time diversity is not, in itself, claimed to be novel. What is interesting in our technique, however, is the method by which a high bit-error density is detected in a received speech segment (more strictly, in one of two segments in a diversity pair). The basis of such burst detection is a simple autocorrelation-type measurement of relative speech (or channel) quality, denoted by $C$. Unlike a signal-to-noise ratio (SNR), the quantity $C$ can be evaluated over a received segment without reference to the transmitted speech.

In fact, our channel evaluations, based on $C$, are somewhat reminiscent of eye-pattern-based channel assessments in digital data communications.

## III. AUTOCORRELATION C

The proposed measurement is the correlation

$$C(X, W) = \sum_{r=2}^{W} (\operatorname{sgn} X_r \cdot \operatorname{sgn} X_{r-1})/(W - 1), \qquad (2)$$

where $X_r$ represents a sampled speech amplitude, $W$ is a computational window that is typically a few milliseconds long, and sgn $X$ is a polarity function whose mean value and variance are assumed to be 0 and 1. We will also be interested in the correlations $C(XQ, W)$ and $C(Y, W)$, where the quantities $XQ$ and $Y$ refer to (unfiltered) staircase functions at the outputs of local and remote speech decoders (Fig. 2). $C(XQ, W)$ and $C(Y, W)$ are defined by operations similar to (2).

In simulating digital transmissions of speech over burst-error channels, we have found that clustered transmission errors tend to have the following type of effect on $C$: with a high probability (say, on the order of 0.9 or more), $C(Y, W) < C(X, W)$, where $X$ and $Y$ represent original and received speech segments. Qualitatively, the result is a consequence of increased zero-crossing activity in error-corrupted speech waveforms. Actual values of $C(Y, W)$ depend not only on the local error statistics, but also on the value of the corresponding $C(X, W)$, the nature of the quantization of $X$ (prior to transmission) as reflected in the value of $C(XQ, W)$, and the extent of channel error propagation in the received signal (if the quantization is differential). Because of these factors, the magnitude of $C(Y, W)$ cannot be used, as such, for very reliable burst-error detection.

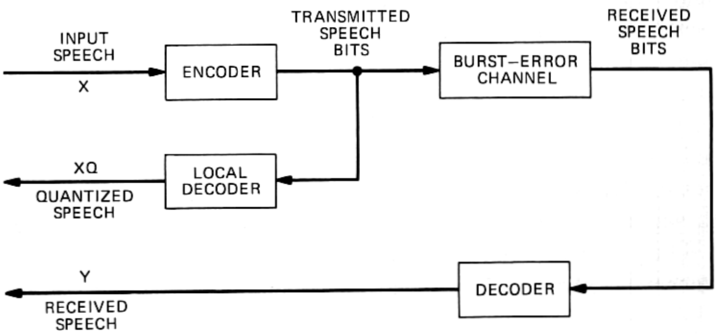


Fig. 2—Definition of $X$, $XQ$, and $Y$.

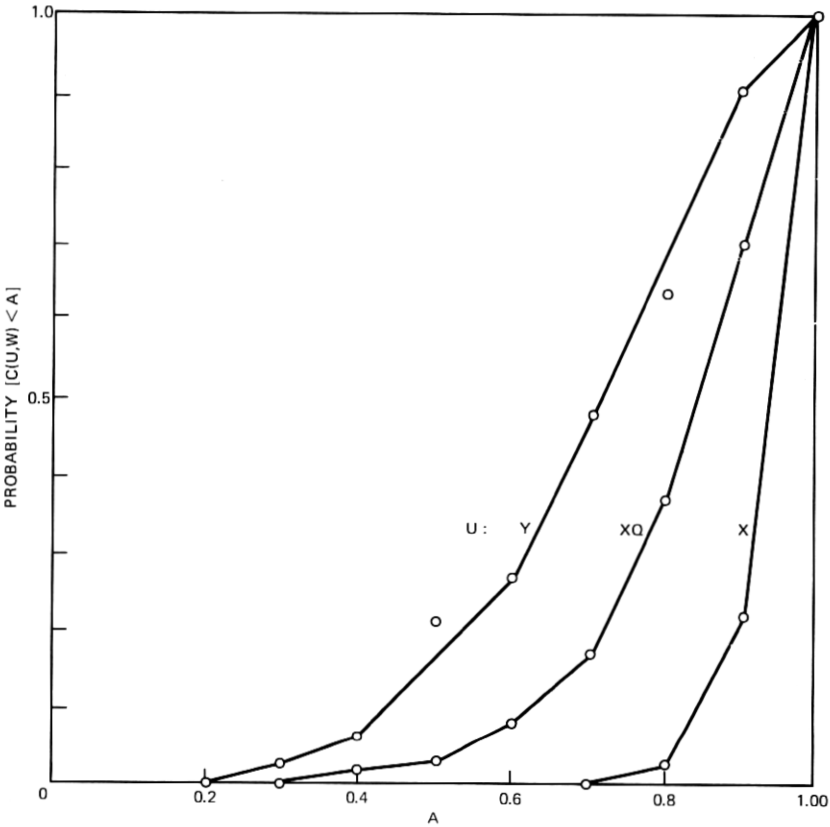


Fig. 3—Distributions of $C(X, W)$, $C(XQ, W)$, and $C(Y, W)$ in 24-kHz delta modulation [$P(E) = 0.06$, $W = 64$].

These points are demonstrated by the results in Fig. 3 and Table I. These results refer to the 24-kHz delta modulation[2] of the band-limited (200 to 3200 Hz) female speech utterance, "A lathe is a big tool." The delta-modulation bits were transmitted through a simulated burst-error channel whose time statistics are shown in Fig. 1. As mentioned earlier, the average bit-error probability $P(E)$ on this channel is 0.06. Local error probabilities, as measured over blocks of $W$ samples, will

Table I — Mean and median values of $C(X, W) - C(Y, W)$

| $W = 64$ | Median | Mean |
|---|---|---|
| $P(E, W) = 0.00$ | 0.11 | 0.14 |
| $P(E, W) = 0.14$ ave | 0.18 | 0.26 |

be denoted by $P(E, W)$. (In delta modulation, a "sample" is synonymous with a "bit." In $B$-bit PCM or differential PCM, a "sample" refers to an entire $B$-bit word.) The windows for Fig. 3 and Table I are $W = 64$ samples long.

Figure 3 shows the distributions of $C(X, 64)$, $C(XQ, 64)$, and $C(Y, 64)$: specifically, values of the probability that $C(U, W)$ is less than $A$ where $U = X$, $XQ$, or $Y$; $W = 64$; and $-1 \leqq A \leqq 1$. The results refer to a subset of samples characterized by nonzero values of $P(E, W)$, and an average error probability of 0.14. Notice how quantization errors, as well as transmission errors, tend to decrease the correlation $C$. Correlation losses due to noise and distortion are also demonstrated in Table I, which summarizes mean and median values of $[C(X, 64) - C(Y, 64)]$ for two channel conditions: the case of zero transmission errors [a subset of blocks where $P(E, W) = 0$] and the case of nonzero transmission errors [the subset of blocks where the average $P(E, W) = 0.14$]. Incidentally, both these subsets belong to the set of blocks whose average $P(E, W) = 0.06$. The top row in Table I measures the effect of quantization errors (plus, strictly speaking, the effect of error propagations in received speech), while the bottom row demonstrates the contributions of local transmission errors.

The distribution distances in Fig. 3 and the numbers in Table I both lead to the following conclusion: Although the channel quality $[P(E, W)]$ has a very clear effect on the autocorrelation $C$, the effect is not strong enough for $C(Y, W)$ to be employed, as such, as a reliable

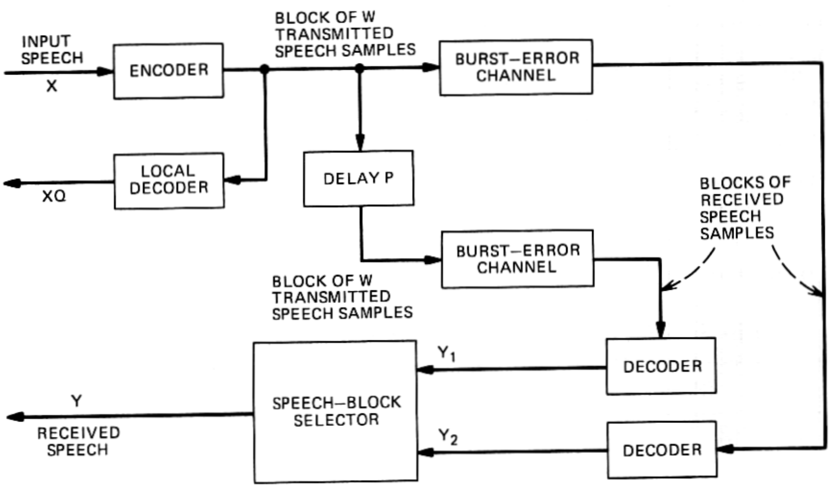


Fig. 4—Two-channel time diversity with block transmissions.

measure of speech or channel quality [$P(E)$]. To explain, a low value of $C(Y, W)$ is often indicative of a local transmission error burst. Occasionally, however, a poor autocorrelation may simply be a reflection of received waveform history and/or quantization noise, and/or an above-average high-frequency content in the local speech input.

A situation where channel information can be reliably extracted from $C$ is in time-diversity coding. Consider, for example, the two speech segments $Y_1$ and $Y_2$ of a time-diversity pair (Fig. 4). The channel-independent factors mentioned at the end of the previous paragraph are exactly the same for both $Y_1$ and $Y_2$. Consequently, any difference between $C(Y_1, W)$ and $C(Y_2, W)$ can be safely attributed to differences in the channel conditions affecting the receptions $Y_1$ and $Y_2$.

## IV. THE USE OF C IN TIME-DIVERSITY CODING

We propose that, for time-diversity reception, the autocorrelation $C$ be used as a criterion for speech segment selection at the receiver. For example, with two-channel time diversity (Fig. 4), we suggest the

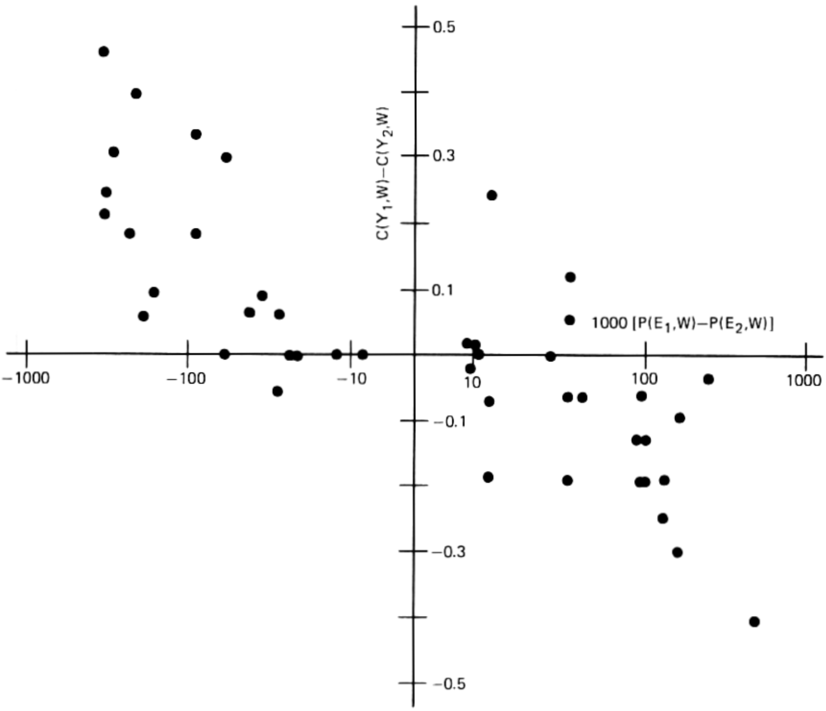


Fig. 5—Performance of $C(Y, W)$-based speech selector with three-bit DPCM [$W = P = 64$; $P(E) = 0.06$].

following reception rule:

$$Y = Y_1 \text{ (or } Y_2) \text{ according as } C(Y_1, W) \geqq \text{ (or } <) C(Y_2, W), \quad (3)$$

where

$$C(Y_u, W) = \sum_{r=2}^{W} (\text{sgn } Y_{ur} \cdot \text{sgn } Y_{u(r-1)})/(W - 1); \quad u = 1, 2.$$

The effect of (3) is to select the speech segment whose signum (polarity) function exhibits the higher autocorrelation. The rest of this section presents results that demonstrate the credibility of the above procedure. Specifically, we point out that very strong *negative correlations* exist between the following quantities:

$$\text{sgn } [C(Y_1, W) - C(Y_2, W)]$$

and (4)

$$\text{sgn } [P(E_1, W) - P(E_2, W)].$$

It is assumed that smaller $P(E)$ values imply better speech quality so

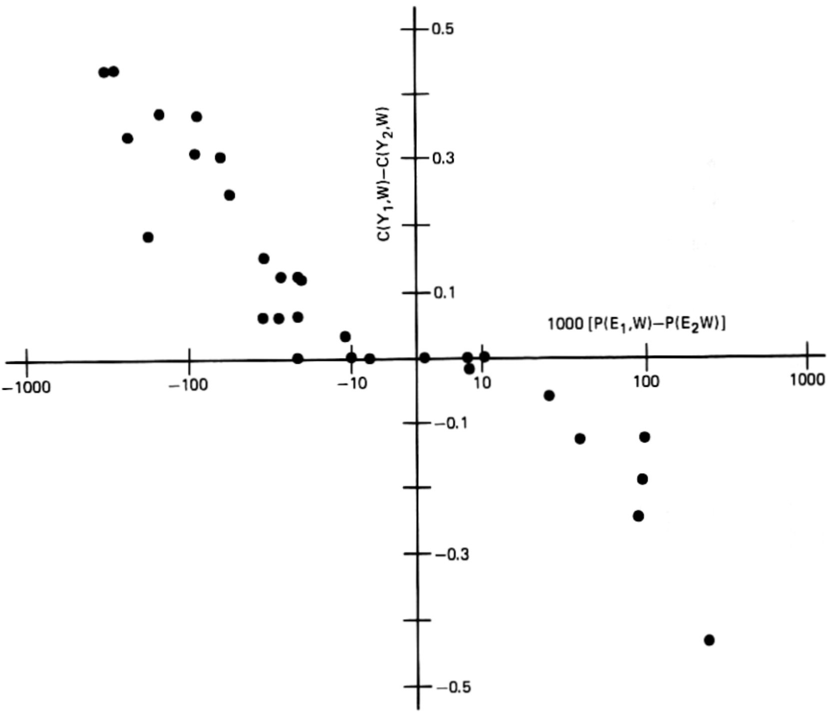


Fig. 6—Performance of $C(Y, W)$-based speech selector with three-bit PCM $[W = P = 64; P(E) = 0.06]$.

that negative correlations between the quantities in (4) are indeed indicative of an appropriate reception rule. Most of the following discussion refers to three-bit DPCM coding. This is an example of practical interest for the time-diversity coding of speech over burst-error channels.[2]

Figures 5, 6, and 7 are scatter plots of $[C(Y_1, W) - C(Y_2, W)]$ versus $[P(E_1, W) - P(E_2, W)]$ for illustrative speech codes (PCM, DPCM) and average transmission-error rates of 0.03 and 0.06. The speech input was the same as that used in Section III, and the scatter plots represent sample subsets of simulation results. The members of the subsets were equally spaced points that spanned the total speech duration of about 1.5 seconds. Notice the negative correlation between $[C(Y_1, W) - C(Y_2, W)]$ and $[P(E_1, W) - P(E_2, W)]$ in each of Figs. 5, 6, and 7. This negative correlation reflects the fact that (for a given speech input and quantization error pattern) a higher $C(Y, W)$ value implies a lower $P(E, W)$ value, i.e., a better speech quality. The very small I- and III-quadrant occupancies reflect a low probability

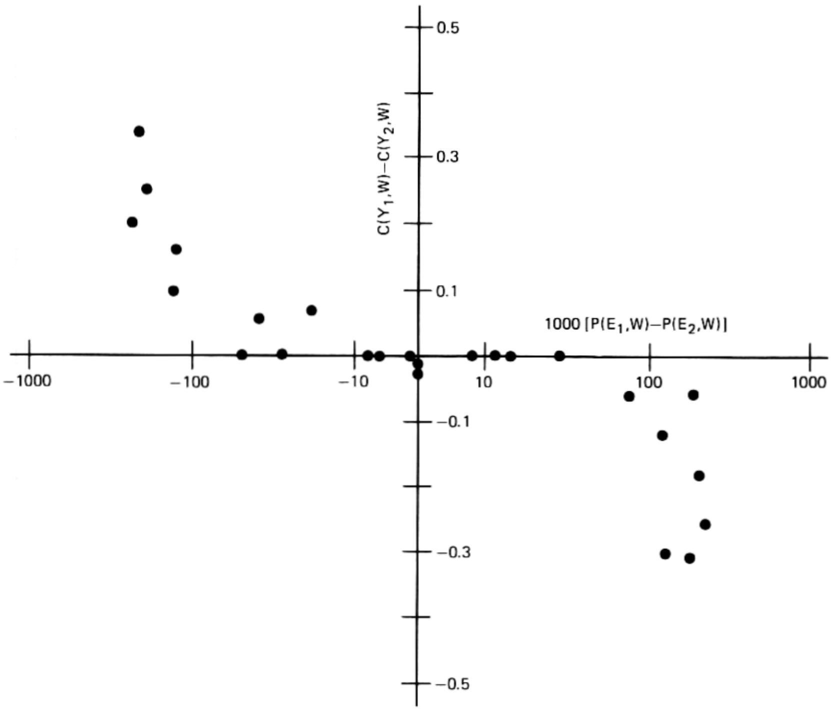


Fig. 7—Performance of $C(Y, W)$-based speech selector with three-bit DPCM $[W = P = 64; P(E) = 0.03]$.

of failure (wrong speech-segment selection for the $C(Y, W)$-based selector (3)).

We briefly discuss the effects of correlation window length $W$ and time-diversity spacing $P$ on received speech quality. The quantities SNRT and SNRR refer to signal-to-noise ratios measured over the duration of the entire speech utterance:

$$\begin{aligned} \text{SNRT} &= \sum X_r^2 / \sum (X_r - XQ_r)^2 \\ \text{SNRR} &= \sum X_r^2 / \sum (X_r - Y_r)^2. \end{aligned} \tag{5}$$

$T$ and $R$ refer to SNR values as measured at the local (transmitter-end) and remote (receiver-end) decoders (Fig. 2). We are interested in DPCM codes with a forward-adaptive quantizer: the step size is updated every 64 samples at the transmitter, and the step-size information communicated to the receiver in a special error-protected format.[2] Finally, the differential coding uses a time-invariant first-order predictor. The predictor coefficient was 0.6. This value was suggested by the need to dissipate the effects of channel errors in the reconstructed speech, as explained in the companion paper.[2]

Table II shows the effects of $W$ and $P$ on the received speech quality as measured by SNRR. It is seen that 100-percent redundancy, together with a good choice of $W$ and $P$, can buy a more-than-4-dB improvement over unprotected DPCM. Incidentally, the overall transmission rate is approximately 48 kb/s for the time-diversity codes and 24 kb/s for the nonredundant code. The lower error rate (0.03) used for the latter is a reflection of the lower transmission rate.[1,2]

Figure 8 elaborates on the performance of the optimal ($W = 64$, $P = 256$) time-diversity code, while Table III compares its performance with that of the bit-protecting scheme[2] mentioned in Section II.

The diversity systems are formally sketched in Fig. 9. The encoding delays ($P + W$ for block protection and $2P'$ for bit protection) are

Table II — Effect of $W$ and $P$ on SNRR [3-bit DPCM; $P(E) = 0.06$]

| $W$ (Number of 8 kHz-samples) | $P$ | SNRT (dB) | SNRR |
|---|---|---|---|
| 64 | 64 | 20.4 | 14.5 |
| 128 | 128 | 20.4 | 12.1 |
| 256 | 256 | 20.4 | 13.3 |
| 64 | 256 | 20.4 | 15.8 |
| Unprotected 3-bit DPCM with $P(E) = 0.03$ | | 20.4 | 11.5 |

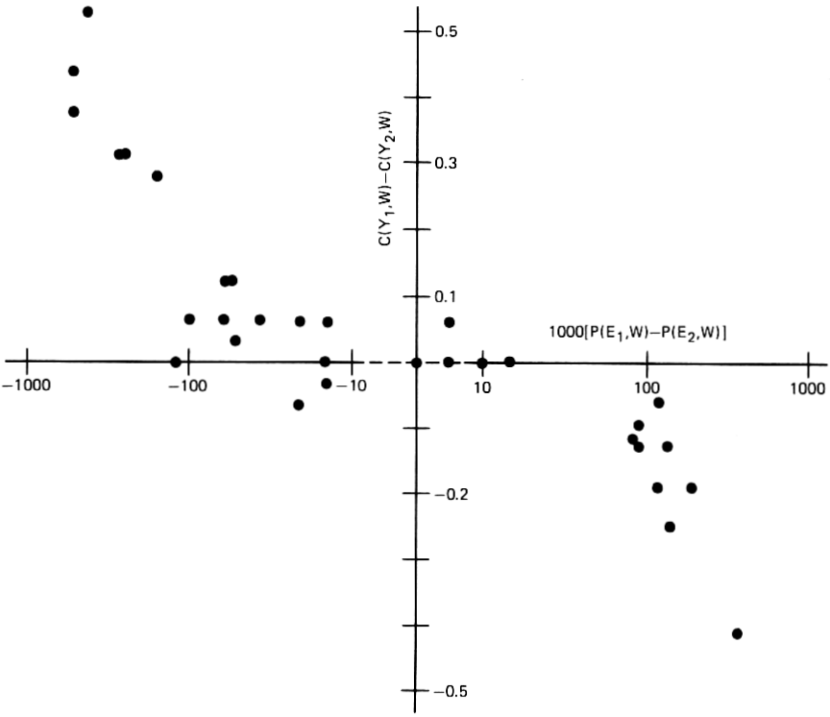


Fig. 8—Performance of $C(Y, W)$-based speech selector with three-bit DPCM [$W = 64$; $P = 256$; $P(E) = 0.06$].

chosen to be of the same order of magnitude. (Both the schemes are expected to perform slightly better with longer encoding delays.) Table III indicates a slight SNRR superiority for the block-protection technique, especially at the higher error rate. What is more significant than the SNRR advantage is a perceptual effect; the block-protected speech sounds considerably crisper. The companion paper[2] includes

Table III — SNRR values (dB) in block-protecting and bit-protecting schemes for time-diversity coding of three-bit DPCM speech

| $P(E)$ | Bit Protection | Block Protection |
|---|---|---|
| 0.000 | 20.4 | 20.4 |
| 0.024 | 17.0 | 17.4 |
| 0.054 | 14.5 | 15.8 |

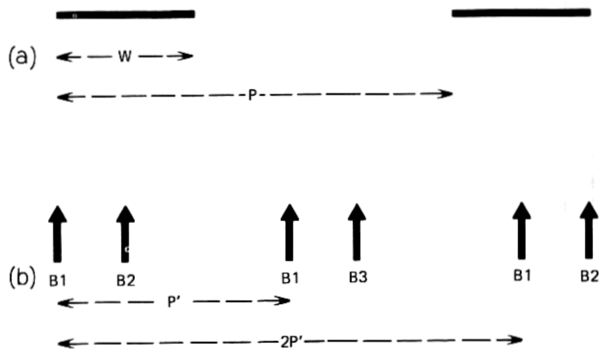


Fig. 9—Time-diversity coding based on (a) block protection ($W$ = 8 ms, $P$ = 32 ms); (b) bit protection ($2P'$ = 32 ms).

more observations on the speech quality resulting from error-protected DPCM.

## V. CONCLUSION

This paper has demonstrated the capabilities of a new technique for signal selection in time-diversity systems. The results of Table III are a good indication of the practical utility of the new technique. We believe, however, that the contribution of this paper consists not in the specific quality improvements (over bit-protecting systems) in Table III, but in the fact that the autocorrelation of the most significant bit (polarity function) is indeed a useful measure of relative signal quality over noisy channels. This is demonstrated mainly in the scatter plots in Figs. 5, 6, 7, and 8. The use of the most significant bit in evaluating signal quality leads obviously to simple implementations.

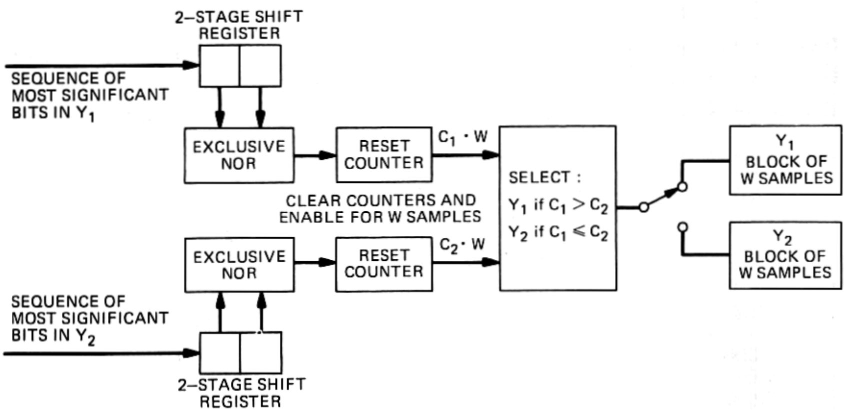


Fig. 10—Implementation of an autocorrelation-based block selector.

A possible configuration for an autocorrelation-maximizing signal selector is depicted in Fig. 10.

## VI. ACKNOWLEDGMENT

The author wishes to thank P. Cummiskey for comments on an earlier version of this paper.

## REFERENCES

1. V. H. McDonald and G. Arredondo, private communication.
2. N. S. Jayant, "Step-Size Transmitting Differential Coders for Mobile Telephony," B.S.T.J., this issue, pp. 1557–1581.