THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 9, November 1977
Printed in U.S.A.

Tandem Connections of Wideband and Narrowband Speech Communication Systems Part 2—Wideband-to-Narrowband Link

By L. R. RABINER, M. R. SAMBUR, R. E. CROCHIERE, and D. J. GOODMAN

(Manuscript received March 25, 1977)

In this paper the tandem link of a 16 kb/s Continuously Variable Slope Delta modulator (CVSD) waveform coder and a 2.4 kb/s Linear Predictive Coding (LPC) vocoder is studied. Of prime concern are the effects of the CVSD coder on the LPC vocoder analyzer. In particular the problems involved in making a reliable voiced-unvoiced decision, estimating pitch period, and estimating LPC coefficients from the coder output are studied. It is shown that LPC coefficient estimation from the CVSD output is highly inaccurate. An analytical distortion measure (an LPC distance) is used to show the magnitude of the distortion introduced by the coder as a function of the signal gain into the CVSD coder. Although the remainder of the LPC analysis (i.e., pitch detection, voiced-unvoiced decision, and gain calculation) can be performed reasonably accurately, the magnitude of the distortions in estimating the LPC coefficients is sufficiently large to make the vocoded speech barely intelligible and of poor quality.

I. OVERVIEW OF THE TANDEM LINK OF CVSD TO LPC

In the first part of this paper we discussed the effects of the narrowband system (the LPC vocoder operating at 2400 b/s) on the wideband system (the CVSD waveform coder).[1] There it was shown that one of the major issues was tailoring the signal characteristics of the vocoded speech to reduce the peak factor, thereby reducing the amount of slope overload noise generated in the CVSD. When we consider the tandem link of CVSD and LPC, far more serious problems are encountered since we must estimate the basic speech production parameters (i.e., pitch, voiced-unvoiced, LPC coefficients) from a severely degraded signal. Since speech parameter estimation is an imperfect process, even on high-quality speech, the effects of the CVSD coder, which include quantization noise

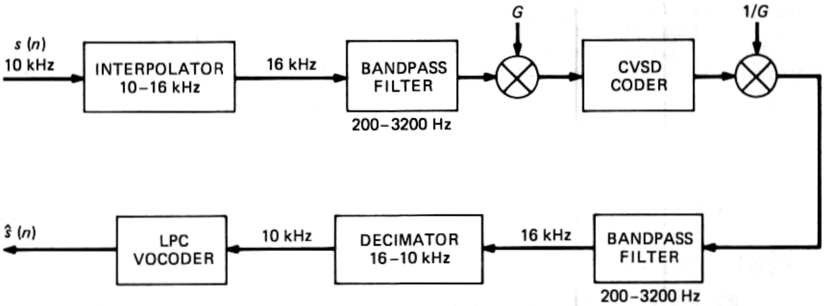

Fig. 1—Block diagram of signal processing operations in tandem link of a CVSD coder and an LPC vocoder.

as well as slope overload noise, could potentially make the tandem link totally unacceptable.

In this paper we discuss several aspects of a tandem link consisting of a CVSD waveform coder, and an LPC vocoder. Our purpose is to demonstrate the range of signal levels over which the LPC can operate reasonably well in tandem with the CVSD coder. Figure 1 shows a block diagram of the signal processing used in implementing and testing a CVSD-LPC tandem link. The speech signal $s(n)$ is assumed to be sampled at a 10-kHz rate. Thus the first block in Fig. 1 is an interpolator to raise the sampling rate of the signal to 16 kHz. The interpolator described in Part 1 of this paper was used here.[1] The 16-kHz signal was then sharply bandpass-filtered from 200 Hz to 3200 Hz using the 8th-order elliptic bandpass filter described in Part 1 of this paper.[1] To simulate variations in overall signal level into the CVSD coder, a variable gain G was applied to the filtered 16-kHz signal. The gain G was varied from 0.009375 to 2.5 in the simulations which gave about a 50-dB variation in signal level over which the system was studied. To compensate for the input scaling, a gain of $1/G$ was used at the output of the CVSD coder. The output of the coder was again sharply bandpass-filtered from 200 to 3200 Hz to remove the wideband quantization noise generated in the CVSD coder. For compatibility with the LPC system the signal was then decimated to a 10-kHz sampling rate using the decimator described in Part 1 of this paper.

Figure 2 shows a block diagram of the processing required for the LPC vocoder. The LPC analyzer estimates the following control parameters:

(*i*) Pitch period
(*ii*) Voiced-unvoiced decision
(*iii*) Signal gain
(*iv*) LPC parameters

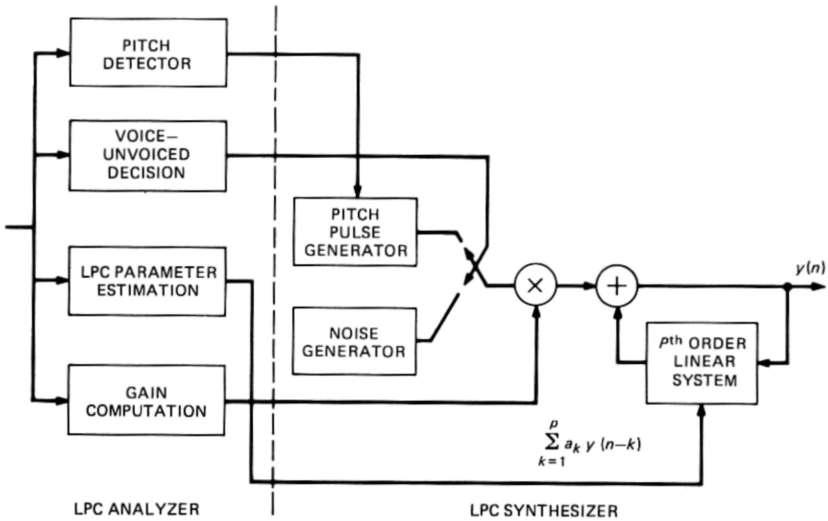

Fig. 2—Block diagram of LPC analyzer and synthesizer.

The LPC synthesizer uses the estimated parameters to recreate the speech in the manner shown in Fig. 2. The details of the analysis and synthesis methods are described in Part 1 of this paper.

Based on our knowledge of both the techniques used in LPC analysis and the degradations introduced by the CVSD coder, it was anticipated that the voiced-unvoiced decision and the LPC parameter estimation algorithms would be most affected by the CVSD coder. Thus, in the next two sections we discuss the specific algorithms used for voiced-unvoiced detection (along with pitch detection) and show results on how the algorithms performed in the tandem link as a function of the signal level into the CVSD coder. In Section IV we present results on the accuracy with which the LPC parameters were estimated from the coder output. For a measure of similarity between coder input and output, the LPC distance measure proposed by Itakura is used. Finally, in Section V we discuss the interactions between the CVSD coder and the LPC vocoder and suggest some possible ways to improve the performance of a tandem link of a wideband and a narrowband system.

II. PITCH DETECTOR AND VOICED-UNVOICED DETECTOR USED IN THE TANDEM LINK

As discussed in the preceding section, the choice of an appropriate pitch detector and voiced-unvoiced detector is critical to the proper operation of the LPC vocoder. Based on a series of intensive investigations into both objective and subjective rankings of a variety of pitch detectors,[2,3] it was shown that simple waveform pitch detectors would be in-

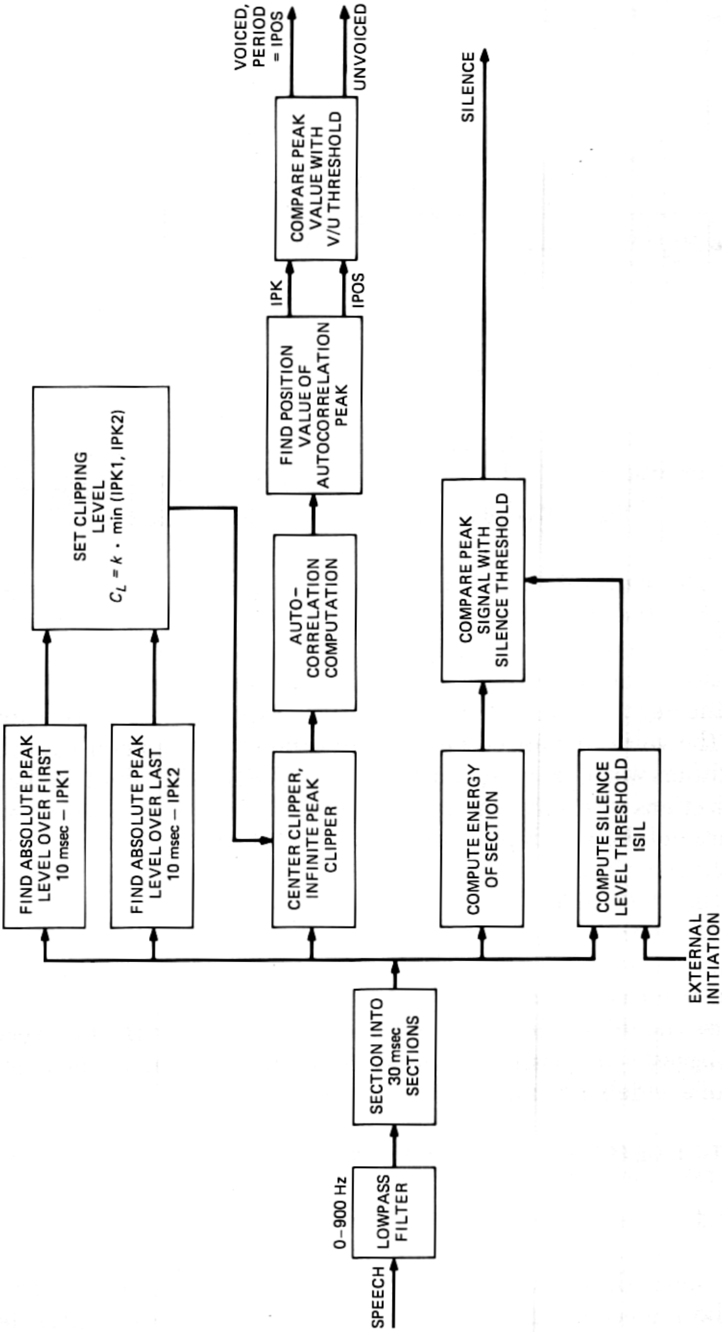

Fig. 3—Block diagram of modified autocorrelation pitch detector.

adequate for a severely degraded waveform such as obtained at the output of a CVSD coder. Thus either a sophisticated correlation-type pitch detector, or a spectral-type pitch detector is required for this application. From this class of pitch detectors both the AMDF[4] and AUTOC[5] pitch detectors were found to be moderately fast, and sufficiently robust over a wide variety of transmission conditions and pitch range of the speaker. Because of the familiarity of the authors with the AUTOC pitch detector, this method was finally selected.

Before the method of operation of this pitch detector is reviewed, some comments must be made about the selection of the voiced-unvoiced detector. Ideally one would prefer to make a voiced-unvoiced decision prior to, and independent of, the pitch detection. In this manner the role of the pitch detector is strictly to make the best estimate of pitch period, given a priori that the segment is accurately classified as voiced. For unvoiced segments, the pitch detector is not used at all. There have been at least three proposed methods for making a voiced-unvoiced decision prior to and independent of any pitch detection.[6–8] However, all three methods suffer from the necessity of having a training set of data that characterizes the signal classes. For CVSD coding, the variability of the signals due to variations in gain is exceedingly large—i.e., a 40-dB variation in input level can change the signal from one with a large amount of granular noise to one with a large amount of slope overload noise. Therefore, making a voiced-unvoiced decision accurately without a periodicity measurement (pitch detector) to aid the decision is extremely difficult. Thus, the voiced-unvoiced decision is combined with the pitch detection in the AUTOC method.

A block diagram of the AUTOC pitch detector is given in Fig. 3. The method requires that the speech be lowpass-filtered to 900 Hz. Thus a 99-point linear phase, FIR digital filter is used here.[9] The lowpass-filtered speech is sectioned into overlapping 30-msec (300 samples at 10 kHz) sections for processing. Since the pitch period computation for all pitch detectors is performed 100 times/second—i.e., every 10 msec—adjacent sections overlap by 20 msec or 200 samples.

The first stage of processing is the computation of a clipping level C_L for the current 30-ms section of speech. The clipping level is set at a value which is 64 percent of the smaller of peak absolute sample values in the first and last 10-ms portions of the section. Following the determination of the clipping level, the 30-ms section of speech is center clipped, and then infinite-peak-clipped, resulting in a signal which assumes one of three possible values, 1 if the sample exceeds the positive clipping level, -1 if the sample falls below the negative clipping level, and 0 otherwise.

Following clipping, the autocorrelation function for the 30-ms section is computed over a range of lags from 20 samples to 200 samples (i.e.,

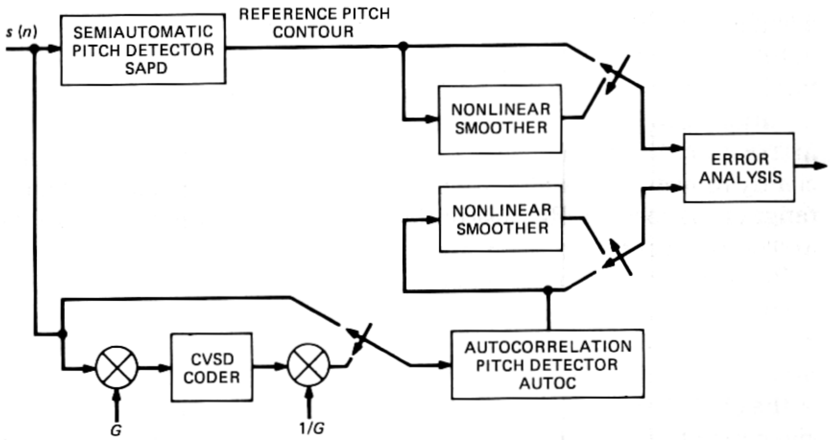

Fig. 4—Block diagram of system used to compare pitch contours from two pitch detectors and to perform an appropriate error analysis.

2-msec to 20-msec period). Additionally, the autocorrelation at 0 delay is computed for appropriate normalization purposes. The autocorrelation function is then searched for its maximum (normalized) value. If the maximum (normalized value) exceeds 0.25, the section is classified as voiced and the location of the maximum is the pitch period. Otherwise, the section is classified as unvoiced.

In addition to the voiced-unvoiced classification based on the autocorrelation function, a preliminary test is carried out on each section of speech to determine if the peak signal amplitude within the section is sufficiently large to warrant the pitch computation. If the peak signal level within the section is below a threshold computed from the background noise level, the section is classified as unvoiced (silence) and no pitch computations are made.

III. EFFECTS OF CVSD CODING ON PITCH DETECTION

To investigate the effects of CVSD coding on pitch detection, two sentences were used whose pitch contours were known extremely accurately.[9] Figure 4 shows a block diagram of the experimental arrangement used to show pitch detection errors in the tandem link. The speech, $s(n)$, is analyzed by the SAPD method[9] to give the reference pitch contour, $p_r(m)$, $m = 1, 2, \ldots, M$, where M is the number of 10-msec frames in the utterance, and $p_r(m) = 0$ if the frame is classified as unvoiced. Otherwise $p_r(m)$ is the estimated pitch period. Extensive tests have shown the SAPD method to be a reliable and robust procedure for obtaining the reference pitch contour.[9]

The test pitch contours are obtained by sending the speech either directly to the pitch detector, or first through the CVSD coder where the signal level is determined by the gain G. We denote the test pitch contour as $p_t(m)$, $m = 1, 2, \ldots, M$. The error analysis compares $p_r(m)$ and $p_t(m)$ over the utterance and makes the following measurements:

(*i*) Average pitch period error during voiced regions, $\overline{P}$, defined as

$$\overline{P} = \frac{1}{N_v} \sum_{\substack{m=1 \\ p_r(m) \neq 0 \\ p_t(m) \neq 0 \\ |p_t(m) - p_r(m)| \leqslant 10}}^{M} [p_r(m) - p_t(m)] \tag{1}$$

where N_v is the number of voiced regions satisfying the conditions that the reference pitch contour indicates a voiced region ($p_r(m) \neq 0$), the test pitch contour indicates a voiced region ($p_t(m) \neq 0$), and the difference in estimated pitch period is less than or equal to 10 samples ($|p_t(m) - p_r(m)| \leqslant 10$).

(*ii*) Standard deviation of the pitch period during voiced regions, σ_p, defined as

$$\sigma_p = \left[\frac{1}{N_v} \sum_{\substack{m=1 \\ p_r(m) \neq 0 \\ p_t(m) \neq 0 \\ |p_t(m) - p_r(m)| \leqslant 10}}^{M} (p_r(m) - p_t(m))^2 - \overline{P}^2\right]^{1/2} \tag{2}$$

(*iii*) Number of voiced-to-unvoiced errors, N_{vu}, defined as

$$N_{vu} = \sum_{m=1}^{M} g(p_r(m), p_t(m)) \tag{3}$$

where

$$g(x,y) = \begin{cases} 1 & \text{if } \; x > 0 \text{ and } \; y = 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

(*iv*) Number of unvoiced-to-voiced errors, N_{uv}, defined as

$$N_{uv} = \sum_{m=1}^{M} g(p_c(m), p_r(m)) \tag{5}$$

(*v*) Number of gross pitch period errors, N_G, defined as

$$N_G = \sum_{m=1}^{M} f(p_r(m), p_t(m)) \tag{6}$$

Table I — Error analysis for utterance "Every salt breeze comes from the sea"

(a) Analysis on raw pitch data

Signal	$\overline{P}$	σ_p	N_{vu}	N_{uv}	N_G
Original speech	0.142	0.786	8	7	1
CVSD-G = 0.009375	1.154	1.925	69	73	29
CVSD-G = 0.0395	0.221	0.901	22	18	6
CVSD-G = 0.158	0.252	0.874	7	8	4
CVSD-G = 0.316	0.288	0.961	6	8	5
CVSD-G = 0.632	0.294	0.952	3	12	4
CVSD-G = 1.264	0.397	1.037	5	23	4
CVSD-G = 2.528	0.397	1.159	4	37	10

(b) Analysis on nonlinearly smoothed pitch data

Signal	$\overline{P}$	σ_p	N_{vu}	N_{uv}	N_G
Original speech	0.156	0.756	7	1	0
CVSD-G = 0.009375	1.589	1.236	91	24	1
CVSD-G = 0.0395	0.556	1.029	15	2	0
CVSD-G = 0.158	0.426	1.073	7	0	0
CVSD-G = 0.316	0.282	0.922	6	0	0
CVSD-G = 0.632	0.356	0.920	2	1	0
CVSD-G = 1.264	0.367	1.155	1	4	0
CVSD-G = 2.528	0.490	1.253	0	6	1

where

$$f(p_r(m), p_t(m)) = 1 \quad \text{if} \quad p_r(m) \neq 0, p_t(m) \neq 0, \; |p_r(m) - p_t(m)| > 10$$
$$= 0 \quad \text{otherwise} \tag{7}$$

Since many of the errors made in pitch detection are easily corrected by a nonlinear median-type smoother,[10] the test arrangement in Fig. 4 also shows the capability of passing both the reference and test pitch contours through such a smoother prior to the error analysis. Results will be presented on both the raw data and the smoothed data.

Results obtained on two different sentences are presented in Tables I and II, and some of the key results are summarized in Figs. 5–8. Utterance 1 was the sentence "Every salt breeze comes from the sea" spoken by a low-pitched male and recorded off a conventional telephone line. The utterance had 256 frames (i.e., it was 2.56 seconds long), of which 108 were unvoiced and 148 were voiced. Table I shows values of $\overline{P}$, σ_p, N_{vu}, N_{uv}, and N_G as a function of the gain G, for both the raw data and the nonlinearly smoothed pitch contours. Figure 5 shows plots of N_{vu} versus G (plotted in dB on a normalized scale) for both the raw and smoothed data, and Fig. 6 shows plots of N_{uv} versus G. Results obtained on the original utterance (uncoded) are also presented as a means of comparison.

As seen in Table I, values of $\overline{P}$ for the coded speech were about 2 to 3 times larger than for the original speech (except for G = 0.009375).

Table II — Error analysis for utterance "I know when my lawyer is due"

(a) Analysis on raw pitch data

Signal	$\overline{P}$	σ_P	N_{vu}	N_{uv}	N_G
Original speech	0.304	0.796	1	3	0
CVSD-G = 0.009375	0.192	2.722	21	12	63
CVSD-G = 0.0395	0.304	0.738	17	2	7
CVSD-G = 0.158	0.193	0.660	10	1	2
CVSD-G = 0.316	0.209	0.639	10	1	4
CVSD-G = 0.632	0.228	0.812	9	2	4
CVSD-G = 1.264	0.225	0.922	6	3	5
CVSD-G = 2.528	0.221	0.993	8	4	9

(b) Analysis on nonlinearly smoothed pitch data

Signal	$\overline{P}$	σ_P	N_{vu}	N_{uv}	N_G
Original speech	0.323	0.617	1	2	0
CVSD-G = 0.009375	1.247	2.922	25	10	40
CVSD-G = 0.0395	0.382	0.656	18	1	0
CVSD-G = 0.158	0.172	0.573	11	1	0
CVSD-G = 0.316	0.213	0.549	12	1	0
CVSD-G = 0.632	0.252	0.711	11	1	0
CVSD-G = 1.264	0.257	0.823	10	0	0
CVSD-G = 2.528	0.329	0.985	10	0	0

However, values of $\overline{P}$ were all less than 0.5 samples (except for G = 0.009375) indicating that the average pitch period errors, due to the coder, were still relatively insignificant. For a gain of G = 0.009375 (large amounts of granular noise) the pitch detection process broke down en-

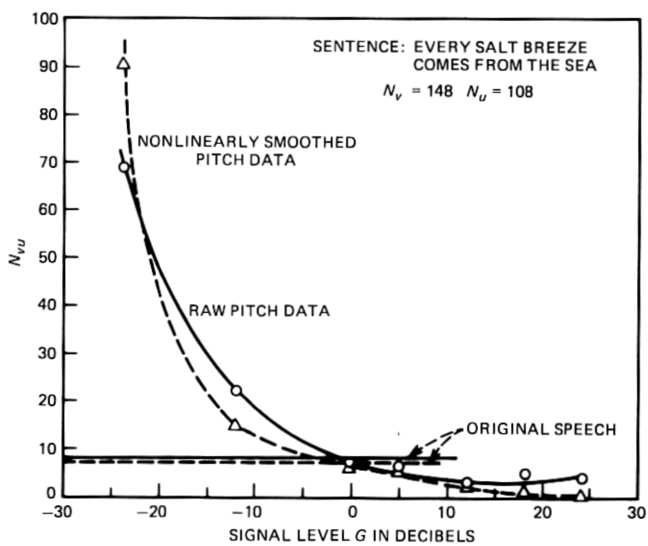

Fig. 5—Plot of number of voiced-to-unvoiced errors versus CVSD signal level for utterance "Every salt breeze comes from the sea."

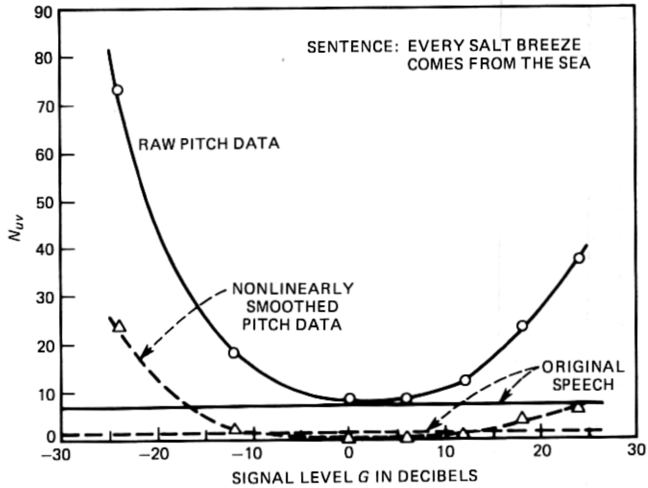

Fig. 6—Plot of number of unvoiced-to-voiced errors versus CVSD signal level for utterance "Every salt breeze comes from the sea."

tirely. Thus, at this extreme the LPC vocoder cannot possibly operate. However, as was shown previously, for this value of gain the CVSD coder produced unintelligible speech; hence we need not be concerned with this result.

Values for σ_p for the coded speech were essentially identical to those obtained for the original utterance. Also the number of gross pitch period errors was small for all values of G except $G = 2.528$ and $G = 0.009375$,

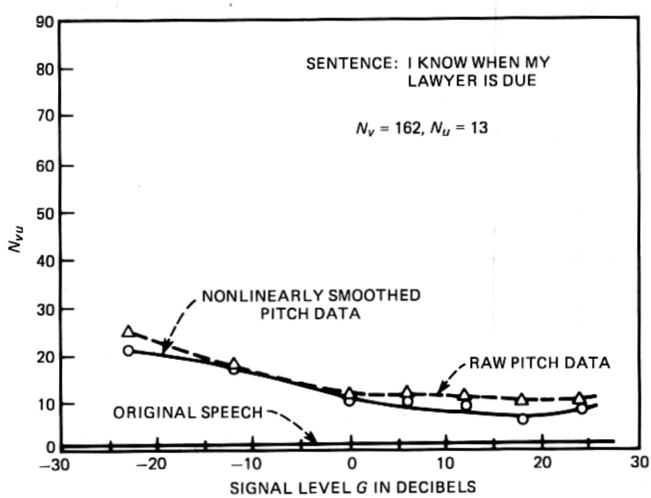

Fig. 7—Plot of number of voiced-to-unvoiced errors versus CVSD signal level for utterance "I know when my lawyer is due."

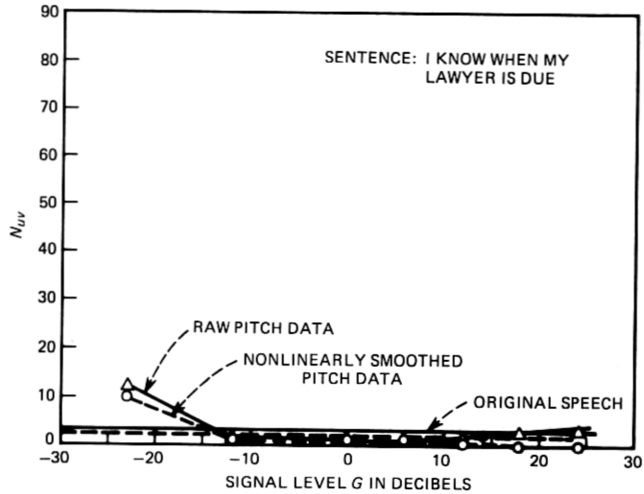

Fig. 8—Plot of number of unvoiced-to-voiced errors versus CVSD signal level for utterance "I know when my lawyer is due."

and all these errors were correctable by the nonlinear smoother, as shown in Table Ib. Thus, one can conclude that for cases in which both the reference and test pitch contours were classified as voiced, the coder did not impede accurate determination of the pitch period—i.e., pitch is well preserved in the CVSD output.

Now the major question is how well the voiced-unvoiced decision could be made on the coder output. An examination of Table I and Figs. 5 and 6 shows that, for several values of G, a substantial number of unvoiced-to-voiced errors occurred. However most of these errors were easily correctable by the nonlinear smoother since the estimated pitch periods (when such errors occur) are essentially random, and are automatically "smoothed" to zero (i.e., unvoiced). Also some of the voiced-to-unvoiced errors are corrected by the smoother.

For this sentence it is concluded that over a fairly large variation in coder input gain, the deterioration of the signal is not so large so as to make pitch detection unreliable.

A second set of results is given for the utterance "I know when my lawyer is due" spoken by another male speaker over a high-quality microphone. This sentence had 175 frames (1.75 seconds) of which only 13 were unvoiced and 162 were voiced. Thus this utterance was essentially all voiced. Results obtained on this utterance are given in Table II and Figs. 7 and 8. Again it is seen that, except for $G = 0.009375$, values of $\overline{P}$, σ_p and N_G (smoothed) are essentially the same for the coder output as for the original. Since there were very few unvoiced frames, the number of unvoiced-to-voiced errors is also the same for the coded speech as for the original. However, the number of voiced-to-unvoiced errors for the coded speech is much larger than for the original speech. Most of these errors occur in the region of the /z/ in "is due," and as such are not correctable by the nonlinear smoother. However, the errors in this low-intensity region are not very preceptible and therefore such errors are not overly crucial.

In summary we have shown that the CVSD coder preserves the pitch of the speech over a reasonably large signal range and that the voice-unvoiced decision can also be reliably made over a fairly large dynamic range of coder inputs.

IV. EFFECTS OF CVSD CODING OF ESTIMATION OF LPC COEFFICIENTS

The next issue to consider is the effects of the CVSD coder on the estimation of the LPC parameters. The LPC coefficients model the combined transfer function of the vocal tract, glottal source, and radiation load. Incorrect estimates of the coefficients can seriously perturb the frequency spectrum of the modeled speech signal and, hence, affect the intelligibility of the synthesized sound.[11]

4.1 Distance measure

To evaluate objectively the spectral distortion introduced by the CVSD coder, an LPC distance measure proposed by Itakura was employed.[12] The LPC distance measure is defined as

$$d_n = \log \left[\frac{a_n V a_n^t}{b_n V b_n^t} \right] \tag{8}$$

where

a_n = LPC coefficient vector $(1, a_1, \ldots, a_p)$ measured in the nth frame of the original uncoded speech signal.

b_n = LPC coefficient vector measured in the nth frame of the CVSD coded speech signal

and V is the speech correlation matrix with elements V_{ij} defined as

$$V_{ij} = v(|i-j|) = \sum_{n=1}^{N-|i-j|} x(n)x(n+|i-j|) \tag{9}$$

where $x(n)$ is the speech signal and N is the number of samples in the frame.

Figure 9 shows examples which illustrate how the measured d_n is useful in measuring the degree of spectral deviation of the coded sound from that of the original.* Although the measure d_n is not the only possible indicator of spectral distortion,[13] it has been shown to closely

* The quantitative significance of d_n is discussed in detail in Ref. 14.

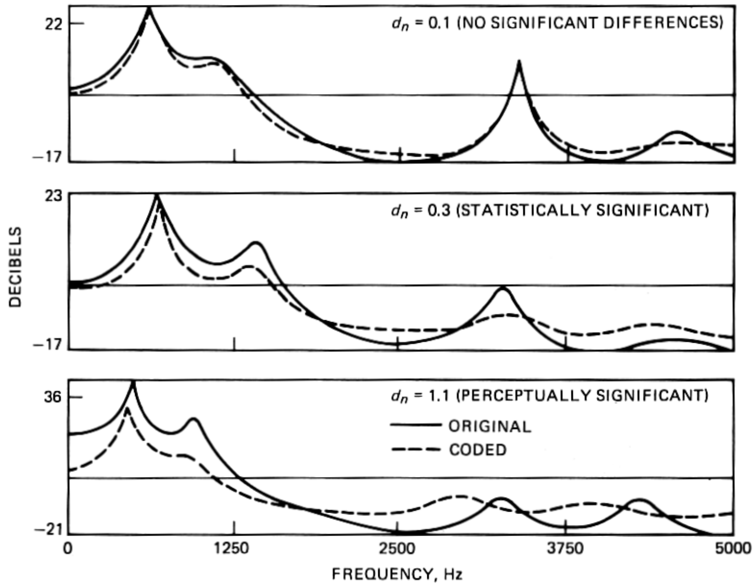

Fig. 9—Plots of typical spectra and the resulting values of d_n for three examples.

correspond to perceptual judgments.[14] In addition, the measure has been effectively applied in problems of speech recognition,[12] speaker recognition,[15] and variable frame rate synthesis.[16] Before discussing the results of the LPC distance evaluation of the CVSD coder, it is important to emphasize that d_n is not a perfect measure of perceptual changes in the character of the sound.[11,17] However, it is a good measure of spectral deviations, which is a useful indicator of intelligibility loss.[14]

4.2 Evaluation

The two sentences utilized in the investigation of pitch detection accuracy were also employed in the evaluation of the effects of CVSD distortion on the estimation of the LPC coefficients. For each sentence, the LPC coefficients for the uncoded, original speech are first calculated. The LPC parameters are calculated 50 times per second at a uniform rate using the autocorrelation method[18] with a 30-msec Hamming window. The speech is preemphasized using a first order digital network with transfer function

$$H(z) = 1 - 0.95z^{-1} \tag{10}$$

prior to LPC analysis in order to minimize the effects of performing the LPC analysis at a uniform rate (i.e., pitch asychronously).[19] The results of this analysis provide the reference LPC coefficients (the a_n's) for each 20-msec frame.

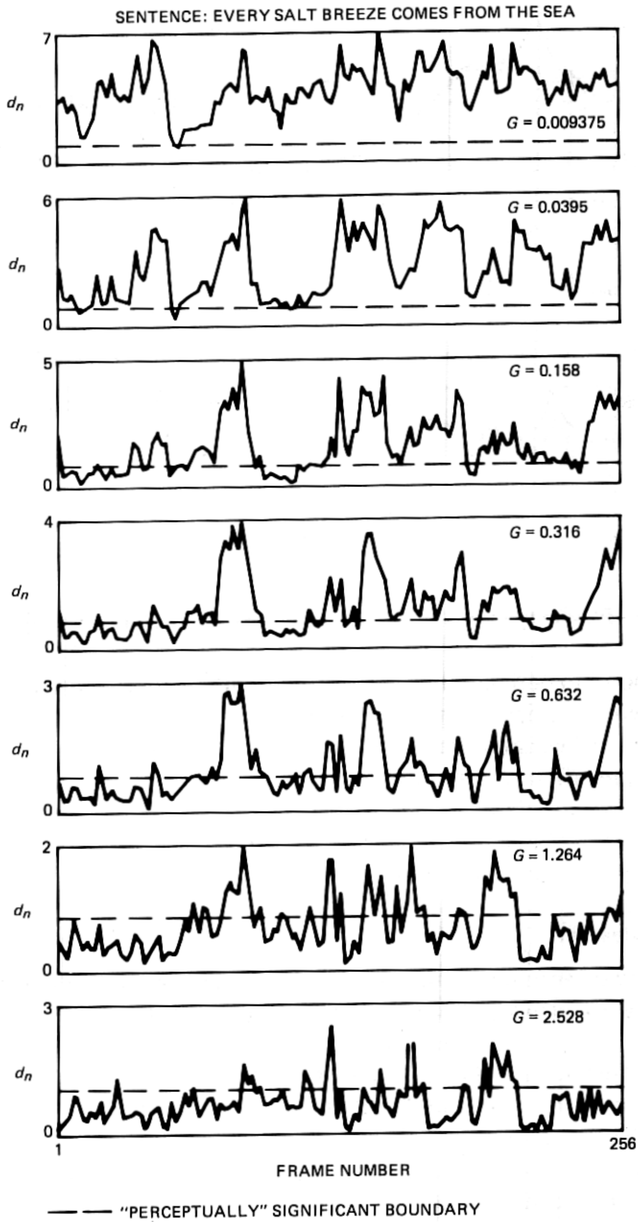

Fig. 10—Values of d_n versus frame number as a function of CVSD signal level for utterance "Every salt breeze comes from the sea."

A similar LPC analysis is performed for each of the various CVSD coded versions of the original sentences. These analyses provide the b_n's for use in the calculation of distance (d_n) between the original sentence and

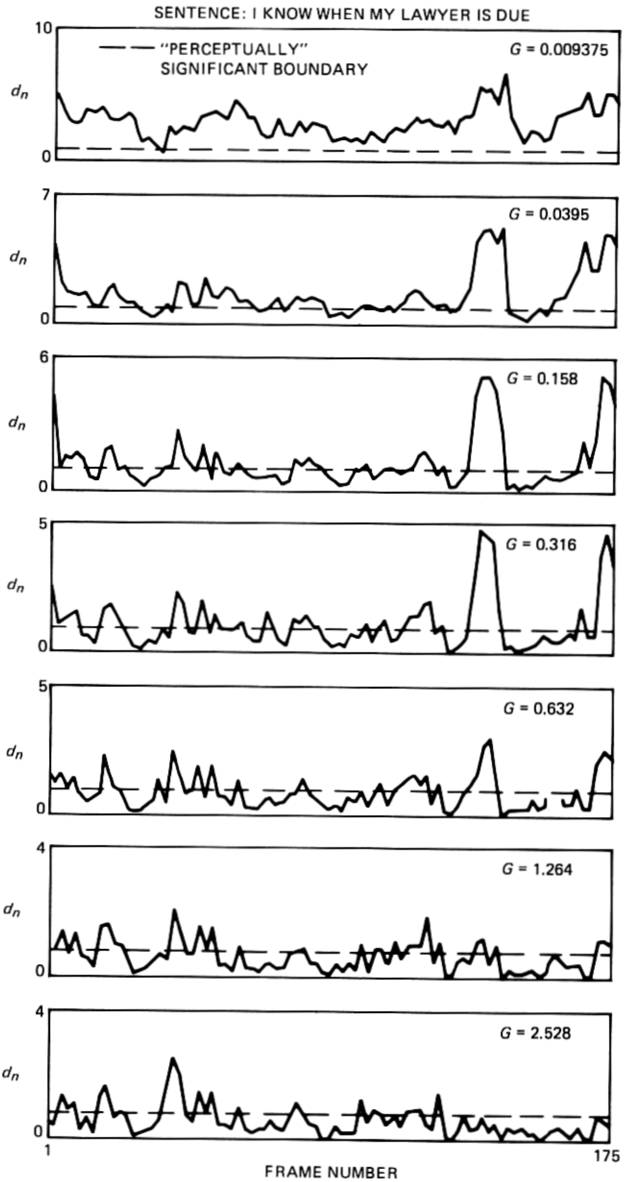

Fig. 11—Values of d_n versus frame number as a function of CVSD signal level for utterance "I know when my lawyer is due."

the particular CVSD-coded sentence. Figures 10 and 11 show the frame-by-frame LPC distance measured for each CVSD-coded version of the two original sentences. The dashed line in the figures refers to a suggested threshold of $d_n = 0.9$ for a just-perceptible difference.[14] Figure

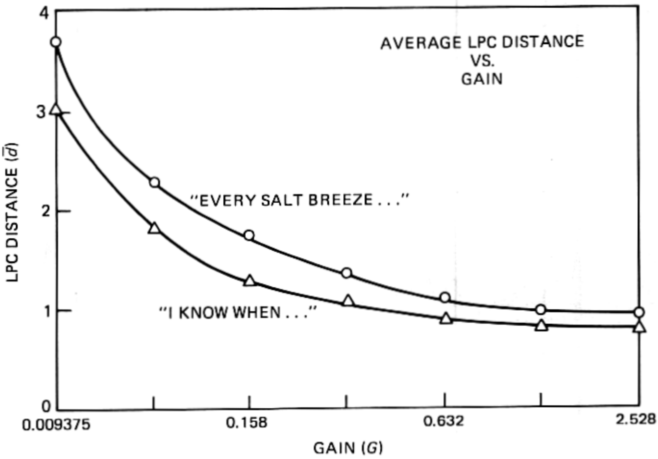

Fig. 12—Plots of average LPC distance ($\bar{d}$) as a function of CVSD signal level (G) for both test sentences.

12 shows the average LPC distance as a function of gain G. The average distance is defined as

$$\bar{d} = \frac{1}{M} \sum_{n=1}^{M} d_n \tag{11}$$

where M is the number of frames in the sentence.

The results of the LPC distance analyses are striking in that the distance uniformly decreases as the gain G increases. This result is in direct opposition to the SNR findings discussed in the first part of this paper.[1] According to the LPC distance measure, the CVSD-coded sentence is improving in quality (i.e., closer in distance to the original) as the gain increases. However, according to the SNR measurements, the similarity between the original and the CVSD-coded sentence is decreasing as the gain increases beyond $G = 0.158$. Although the dissimilarity between the waveforms of the original and the CVSD-coded version with $G = 1.264$ is apparent from Fig. 13, it is interesting to note that informal perceptual experiments indicate that the quality of the CVSD coder is actually improving as the gain G increases. Since the LPC distance measure is sensitive to spectral distortions, it is (in this case) a better measure of quality than SNR. The use of the LPC distance measure as an indication of speech quality has been suggested by other authors.[14]

V. COMPATIBILITY OF CVSD WITH LPC

As a final check on the performance of the entire system, an informal perceptual evaluation of the CVSD-LPC tandem link depicted in Fig. 1 was performed. The LPC vocoder was efficiently designed for a bit rate

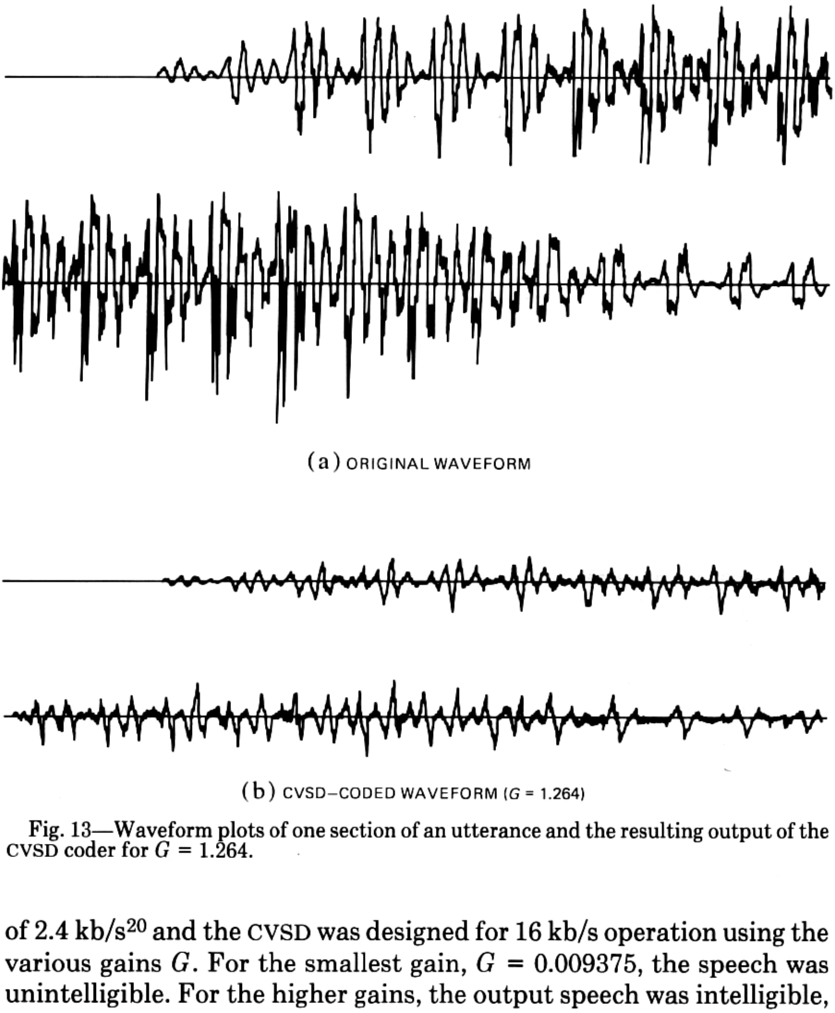

Fig. 13—Waveform plots of one section of an utterance and the resulting output of the CVSD coder for $G = 1.264$.

of 2.4 kb/s[20] and the CVSD was designed for 16 kb/s operation using the various gains G. For the smallest gain, $G = 0.009375$, the speech was unintelligible. For the higher gains, the output speech was intelligible, but the quality was significantly worse than the quality of the 2.4 kb/s LPC synthesis. The quality of the tandem link appeared to saturate (or even become slightly worse due to the poorer estimates of pitch and gain) for $G \geqslant 0.158$. Even for the best-quality output, the combination of CVSD noise and the parametric distortions of the LPC vocoder rendered the tandem a marginal communications link.

VI. SUMMARY

In the tandem link of a wideband and narrowband speech communication system in which the wideband system was a 16 kb/s CVSD coder and the narrowband system was a 2.4 kb/s LPC vocoder, the CVSD coder was shown to be the weak link. The major distortion introduced by the

CVSD coder was spectral distortion as measured using an appropriate LPC distance measure. This distortion was sufficiently severe to make the LPC output, although intelligible, of poor quality. It was further shown that the waveform distortion in the CVSD coder was not so severe so as to make pitch detection unreliable, and even a reliable voiced-unvoiced decision could be made on the CVSD-coded speech.

The major conclusion from this study is that alternative 16-kb/s coders be considered as the wideband communication system for such communication links. Possible alternatives include ADPCM systems,[21] sub-band coders,[22] and transform coders.[23]

REFERENCES

1. R. E. Crochiere, D. J. Goodman, L. R. Rabiner, and M. R. Sambur, "Tandem Connections of Wideband and Narrowband Speech Communications Systems: Part 1—Narrowband-to-Wideband Link," B.S.T.J., this issue, pp. 0000–0000.
2. L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, "A Comparative Performance Study of Several Pitch-Detection Algorithms," IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-24,* No. 5 (October 1976), pp. 399–418.
3. C. A. McGonegal, L. R. Rabiner, and A. E. Rosenberg, "A Subjective Evaluation of Pitch Detection Methods Using LPC Synthesized Speech," IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-25,* No. 3 (June 1977), pp. 221–229.
4. M. J. Ross et al., "Average Magnitude Difference Function Pitch Extractor," IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-22* (October 1974), pp. 353–362.
5. J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-Time Digital Hardware Pitch Detector," IEEE Trans. Acoustics, Speech Signal Proc., *ASSP-24* (February 1976), pp. 2–8.
6. B. S. Atal and L. R. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-24,* No. 3 (June 1976), pp. 201–121.
7. R. J. McAulay, "Optimum Classification of Voiced Speech, Unvoiced Speech and Silence in the Presence of Noise and Interference," Lincoln Laboratory Technical Note 1976–7, June 1976.
8. L. R. Rabiner and M. R. Sambur, "Application of an LPC Distance Metric to the Voiced-Unvoiced-Silence Detection Problem," submitted for publication.
9. C. A. McGonegal, L. R. Rabiner, and A. E. Rosenberg, "A Semi-Automatic Pitch Detector," IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-23* (December 1975), pp. 570–574.
10. L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, "Applications of a Nonlinear Smoothing Algorithm to Speech Processing," IEEE Trans. Acoustics, Speech, Signal Processing, *ASSP-23,* No. 6 (December 1975), pp. 552–557.
11. M. R. Sambur and N. S. Jayant, "Speech Encryption by Manipulations of LPC Parameters," B.S.T.J., *55,* No. 9 (November 1976), pp. 1373–1388.
12. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoustics, Speech, Signal Processing, *ASSP-23,* No. 1 (February 1975), pp. 67–72.
13. A. H. Gray Jr. and J. D. Markel, "Distance Measures for Speech Processing," IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-24* (October 1976), pp. 380–391.
14. M. R. Sambur and N. S. Jayant, "LPC Analysis/Synthesis From Speech Inputs Containing Noise or Additive White Noise," IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-24* (December 1976), No. 6.
15. H. Wakita, "On the Use of Linear Prediction Error Energy for Speech and Speaker Recognition," J. Acoust. Soc. Amer., *57,* Supplement No. 1, Spring 1975 (A).
16. D. T. Magill, "Adaptive Speech Compression for Packet Communication Systems," Telecommunications Conference Record, IEEE Publ. 73, CH0805-2, 29D 1-5.
17. J. R. Makhoul, L. Viswanathan, L. Cosel, and W. Russel, "Natural Communication with Computers: Speech Compression Research at BBN," BBN Report No. 2976, *II,* Bolt Beranek and Newman, Inc., Cambridge, Massachusetts, December 1974.

18. J. D. Markel and A. H. Gray Jr., *Linear Prediction of Speech,* Springer Verlag, 1976.
19. L. R. Rabiner, B. S. Atal, and M. R. Sambur, "LPC Prediction Error—Analysis of its Variation with the Position of the Analysis Frame," submitted for publication.
20. M. R. Sambur, "An Efficient Linear Prediction Vocoder," B.S.T.J., *54,* No. 10 (December 1975), pp. 1693–1723.
21. 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.
22. R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Subbands," B.S.T.J., *55,* No. 8 (October 1976), pp. 1069–1086.
23. R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals," to appear in IEEE Trans. Acoustics, Speech, Signal Proc., *ASSP-25,* No. 4 (August 1977), pp. 299–309.