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

Tandem Connections of Wideband and Narrowband Speech Communications Systems: Part 1—Narrowband-to-Wideband Link

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

(Manuscript received March 25, 1977)

The performance of a tandem connection of narrowband and wideband speech communication systems is evaluated. Specifically, the narrowband system consists of a conventional Linear Predictive Coding (LPC) vocoder operating at a bit-rate of 2.4 kb/s and the wideband system consists of a Continuously Variable Slope Delta modulator CVSD operating at a bit rate of 16 kb/s. In Part 1 of this paper the properties of the narrowband-to-wideband link are investigated and in Part 2 the properties of the wideband-to-narrowband link are investigated. In part 1 the SNR (signal-to-quantizing noise ratio) of the CVSD coder is analyzed over a 50-dB variation of the input signal levels and for a variety of source excitations for the LPC synthesizer. It is shown that SNR improvements in the CVSD coder of 2 to 2.5 dB are possible in the slope overload region of the coder by modifying the source excitation of the LPC synthesizer and by preprocessing the input signal to the coder with an allpass filter. Both methods aid in reducing the peak factor (peak-to-RMS level) of the input speech to the coder. Subjectively, however, only slight improvements in quality, if any, were observed with these modifications.

I. INTRODUCTION

Agencies of the United States government are currently formulating plans for an extensive digital secure voice communication network. In this network, a substantial fraction of the signals will be transmitted over "wideband" circuits at 16 kb/s. Owing to severe bandwidth constraints in some parts of the network, however, there will also be "narrowband" speech links in which the transmission rate is 2.4 kb/s. In preliminary plans, the wideband code format is CVSD (Continuously Variable Slope

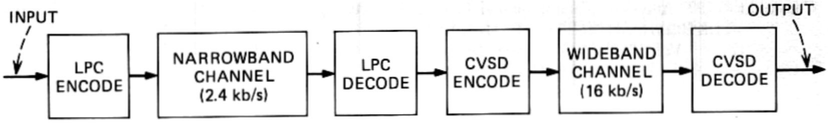


Fig. 1—Narrowband-to-wideband link.

Delta modulation) and the narrowband code format is LPC (Linear Predictive Coding).

Both of these coding methods have been studied extensively, and their performance over single transmission links (involving one encoding operation and one decoding operation) is now well understood.[1,2] In addition to creating single links, however, the proposed communication network will establish tandem connections containing both narrowband and wideband links. It is not clear a priori that two systems, each designed for single-link operation, will interact in tandem to provide acceptable overall quality. Existing knowledge of LPC and CVSD is of limited value in predicting tandem performance and yet the viability of the proposed network depends on adequate performance of tandem as well as single circuits. It is the purpose of this paper to describe the properties of CVSD and LPC that influence the performance of the narrowband-to-wideband connection shown in Fig. 1. A companion paper deals with the complementary wideband-to-narrowband connection.

Our study focuses on issues that arise in tandem links but not in individual circuits. In particular, in this paper we investigate the effects of the narrowband channel on CVSD signal-to-noise ratio (SNR). In doing so we have measured the SNR of the CVSD coder with an original speech input and compared it with the SNR when the CVSD input is LPC synthesized speech with a conventional (impulse) excitation source (during voiced intervals). With a view to improving the quality of tandem circuits, we have also investigated the effect of allpass filtering the LPC output and of using broadened excitation sources for voiced sounds in the LPC synthesizer.

The studies have been carried out by means of computer simulations on a Honeywell DDP 516 computer. In the narrowband-to-wideband tandem we have measured SNR as a function of CVSD input level for a variety of interface and LPC synthesizer source configurations. For each condition (i.e., a given input level, synthesizer source and interface) we have recorded two sentences transmitted through the tandem link. The SNR measurements as well as informal listening experience suggest that CVSD is a critical element in this tandem connection. It has been shown that combinations of interface filter and modified synthesizer source are effective (to some extent) in improving overall quality when the CVSD input level is high. In this case the delta modulator is subject to substantial slope overload. The overload is reduced both by prefiltering and

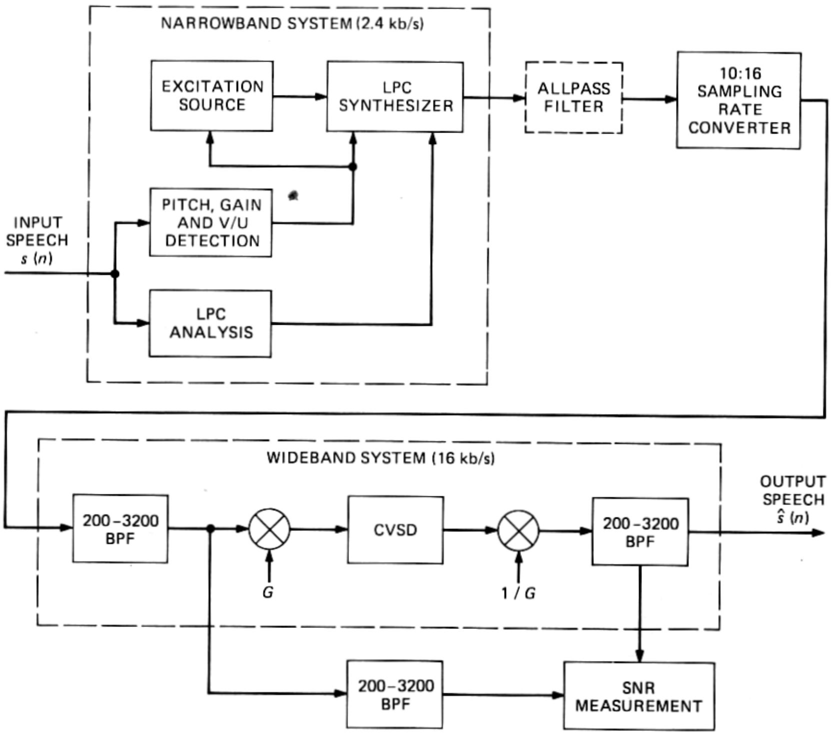


Fig. 2—Block diagram of LPC-to-CVSD link.

by broadening of the LPC synthesizer source because both of these methods reduce the ratio of peak-to-rms level of the signal at the CVSD input. Of the two methods, adjustment of the LPC excitation is more effective but does require some modification of the LPC link. All-pass filtering the LPC output signal has the advantage of being external to both CVSD and LPC.

II. OVERVIEW OF THE NARROWBAND TO WIDEBAND LINK

In this section we discuss the elements of the narrowband-to-wideband tandem connection. We will first review the basic operation of the LPC vocoder and the CVSD coder and will then discuss issues involved in connecting these two systems in a tandem link. After establishing a basic understanding of the various elements in this link, we will discuss factors which affect the performance of this connection and ways in which this performance can be improved.

Figure 2 shows a more detailed block diagram of the overall tandem connection. The narrowband system consists of an LPC analyzer and a pitch and voiced/unvoiced detector. The parameters from these two analyses are used by the LPC synthesizer to resynthesize the speech waveform. An allpass network may be used for further post processing of this waveform. The details of this network will be discussed in a later section.

As shown in Fig. 2, the wideband system consists of a bandpass filter to prevent aliasing, the CVSD coder and another bandpass filter that suppresses CVSD quantizing noise. Gains G and $1/G$ are used in measuring the dynamic range (i.e., variations in performance as a function of signal level) of the CVSD coder.

The basic sampling rate of the narrowband system is 10 kHz and the sampling rate of the wideband system is 16 kHz. In order to interface these two systems a sampling rate converter is used. The details of this conversion will also be discussed in this section as well as the conversion from 16 kHz to 10 kHz which is required in the wideband-to-narrowband tandem connection.

2.1 The wideband system (CVSD)

Figure 3a is a block diagram of the CVSD coding process. The input speech signal is called $x(t)$. An approximation signal $y(t)$ is generated in the encoder feedback loop and at the kth sampling instant ($t = kT$, $T = 1/16000$ sec), the transmitted signal is $b_k = 1$ if

$$x(kT) = x_k > y_k = y(kT) \tag{1}$$

Otherwise $b_k = -1$. A positive output causes $y(t)$ to increase during the next sampling interval making y_{k+1} attain the value

$$y_{k+1} = \alpha y_k + H(1 - \alpha)\Delta_k \tag{2a}$$

where α is the leakage of the approximation signal integrator and $\Delta_k = \Delta(kT)$ is the kth step size. A negative output, $b_k = -1$, results in

$$y_{k+1} = \alpha y_k - H(1 - \alpha)\Delta_k \tag{2b}$$

The step size is obtained from another integrator which processes the output of an overload detector. The overload detector has output V when the three previous CVSD outputs are identical (all 1 or all -1). Otherwise the output of the overload detector is 0. To ensure that the minimum step size is nonzero a small quantity V_1 is added to the output of the overload detector. Thus, the step size satisfies the relation

$$\Delta_{k+1} = \beta\Delta_k + (1 - \beta)(V + V_1) \tag{3}$$

when three previous outputs are identical where β is the leakage of the step size integrator. Otherwise,

$$\Delta_{k+1} = \beta\Delta_k + (1 - \beta)V_1 \tag{4}$$

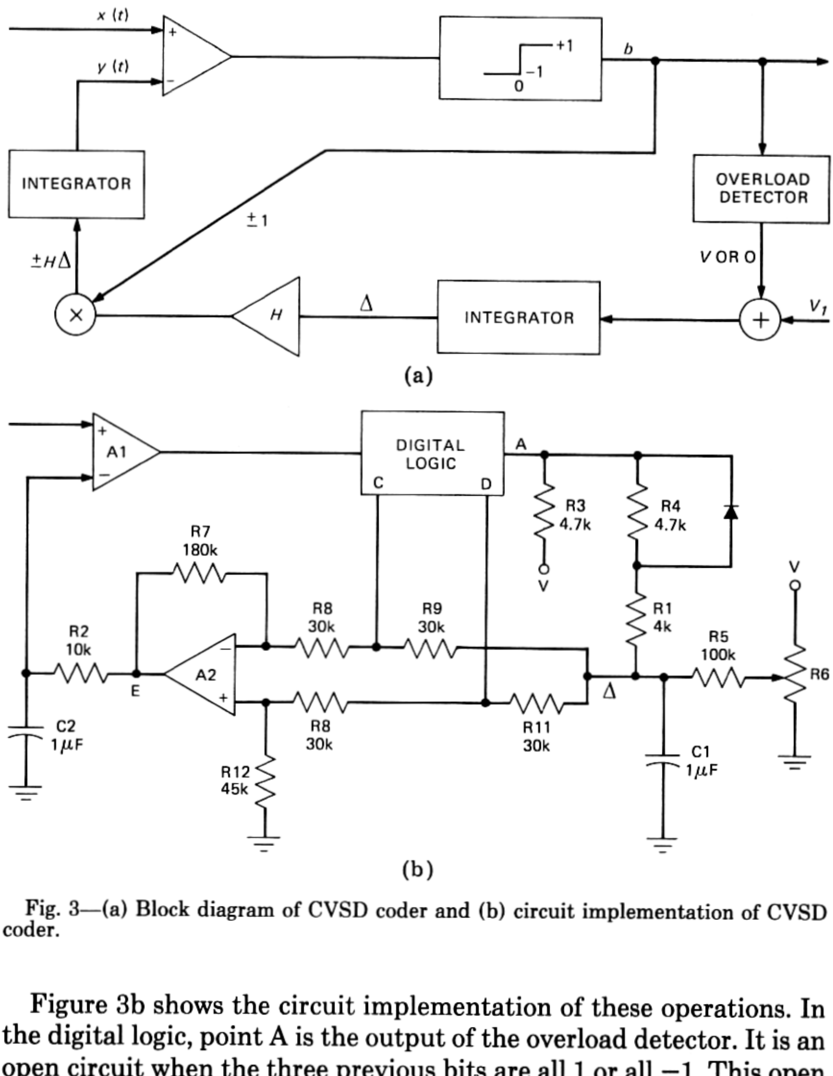


Fig. 3—(a) Block diagram of CVSD coder and (b) circuit implementation of CVSD coder.

Figure 3b shows the circuit implementation of these operations. In the digital logic, point A is the output of the overload detector. It is an open circuit when the three previous bits are all 1 or all −1. This open circuit condition allows the 1 μF capacitor to charge toward $+V$ through R1 and R3. When the last 3 bits are not identical, point A is grounded and the capacitor discharges to ground through R1 and R4. The potentiometer R6 establishes V_1 the minimum voltage on C1.

When $b_k = 1$, point C is grounded and point D is an open circuit. The gain of amplifier A2 is $H = 3$ and the voltage at point E is 3Δ. When $b_k = -1$, C is open circuited and D is grounded causing the voltage at E to be −3Δ. Thus the integrator R2-C2 charges toward ±3 times the voltage on capacitor C1, depending on whether $b_k = \pm 1$.

The time constant of the step size integrator is 5.69 ms (1 μF × R1 +

R3 or R4 in parallel with R5 and with the input impedance of A2 which is 20 kΩ). The step size coefficient is therefore

$$\beta = \exp\left(-\frac{1}{16{,}000}\Big/5.69 \times 10^{-3}\right) = .99 \tag{5}$$

The time constant of the approximation signal integrator is 1 ms which gives

$$\alpha = \exp\left(-\frac{1}{16{,}000}\Big/10^{-3}\right) = .94 \tag{6}$$

In the computer simulation, speech is represented as a 16-bit integer between −32,768 and 32,767, so that a value of $V = 32{,}767$ is equivalent in hardware to a peak speech input equal to the supply voltage. In our studies we have provided for a wide dynamic range of step sizes, $V/V_1 = 200$ so that $V_1 = 164$.

Thus eqs. (3) and (4) are, numerically,

$$\Delta_{k+1} = .99\Delta_k + 329 \tag{7}$$

when three outputs are identical and

$$\Delta_{k+1} = .99\Delta_k + 1.64 \tag{8}$$

otherwise. Similarly eqs. (2a) and (2b) are,

$$y_{k+1} = .94y_k + .18\Delta_k \tag{9}$$

if $x_k > y_k$ and

$$= .94y_k - .18\Delta_k \tag{10}$$

otherwise.

2.2 The narrowband system (LPC)

The narrowband system consists of a Linear Predictive Coding (LPC) system based on an all-pole model of the speech production mechanism. The all-pole model implies that within a frame of speech, the output speech sequence is given by

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + Gu_n \tag{11}$$

where p is the number of poles, u_n is the appropriate input, G is the gain, and the a_k's are the LPC coefficients that represent the spectral characteristics of the speech frame. For a voiced speech segment, u_n is a sequence of pulses separated by the pitch period. If the segment is unvoiced, pseudorandom noise is used as input.

In our study, the LPC coefficients were calculated by the autocorrelation method with $p = 12$ (Ref. 2). The analysis was performed every

20 msec (50 times/sec) across overlapping 300 sample (30 msec) Hamming windowed speech frames. The pitch detection and V/U (voiced/unvoiced) decision is based on the modified autocorrelation method.[3] The effects of pitch and V/U analysis do not in general influence the performance of the narrowband-to-wideband link. In the reverse link (wideband to narrowband), however, the pitch and V/U analysis is strongly affected by the performance of the wideband system. Therefore we will discuss the pitch and V/U analysis in the accompanying paper on the wideband-to-narrowband link.

Since the stability and characterization of the LPC synthesizer is extremely sensitive to small perturbations in the LPC coefficients, it is not possible to achieve low-bit-rate coding by transmitting the LPC coefficients.[2] However, by transmitting either the log area coefficients or the parcor coefficients, a 2.4-kb/s vocoder is readily achieved.[4] The log area coefficients are related to the LPC coefficients by

$$g_i = \log \frac{1 + k_i}{1 - k_i} \tag{12}$$

where the k_i's are termed the parcor coefficients.[2] If we denote $a_i{}^{(j)}$ as the ith linear prediction coefficient for a jth-order linear-prediction model then

$$k_i = a_i{}^{(i)} \tag{13}$$

The parcor coefficients have the very important property that if

$$|k_i| < 1 \qquad i = 1, \ldots, p \tag{14}$$

then it is guaranteed that the linear prediction synthesizer is stable.[2] Thus, small perturbations in the parcor coefficients or log area coefficients will not affect the stability of the synthesizer, and morever these small perturbations will not seriously alter the spectral characterization of the speech segment.[5] Since the log area coefficients are slightly less sensitive to quantization error[5] they were transmitted in the narrowband system.

The quantization of the LPC control signals (pitch, gain, and the g_i's) was accomplished by ADPCM (Adaptive Differential PCM) techniques.[6] In this scheme, the value of a particular control parameter in the nth frame is initially estimated as equal to the transmitted values of the parameter in the (n − 1)st frame. The difference between this predicted value and the actual parameter value is then quantized using a gamma or laplace quantizer with an adaptive step size.[4,6] Complete details of the adaptation scheme and the quantization method are given in ref. 4.

The bit allotment in the narrowband link is as follows. The pitch and gain information is encoded with 3 bits/sample each. The first six log area ratios $g_1, g_2, \ldots, g_6$ are each encoded with 4 bits/sample and $g_7, g_8, \ldots, g_{12}$ are encoded with 2 bits/sample. One bit/frame is used to transmit the V/U decision. This gives a total of 43 bits/frame or a transmission rate of 2.15 kb/s. Another 5 bits/frame are used for transmission of initialization information and frame synchronizing information giving a total of 48 bits/frame or a total transmission rate of 2.4 kb/s for the narrowband system.

2.3 Bandpass filters

The bandpass filters in Fig. 2 are all identical and are used to limit the bandwidth of the signal to the range 200 Hz to 3200 Hz. The third bandpass filter below the block diagram of the wideband system is used for compensating the group delay of the input signal of the CVSD in order to make meaningful SNR measurements on the CVSD.

The bandpass filters are eighth-order recursive elliptic filters with a passband ripple of 0.25 dB and a stopband attenuation greater than 35 dB. The average group delay of the filters is 0.325 msec in the passband and it peaks to 7 msec in the lower transition band. Figure 4 shows the log magnitude response (dB), group delay, and impulse response of these filters.

2.4 Sampling rate conversion

In the tandem connections it is necessary to convert the sampling rate of the signal from 10 kHz to 16 kHz and from 16 kHz to 10 kHz (in the opposite connection). One way of achieving this conversion is to convert the signal to analog form and then resample it at the new sampling rate. This approach is susceptible to electronic noise in the analog circuitry and is limited by the dynamic range of the analog components.

A more attractive approach to the sampling rate conversion process is to do a direct digital-to-digital conversion of the sampling rate. This conversion can be done as accurately as desired and is not prone to extraneous noise from electronic components. The digital-to-digital conversion is accomplished with the aid of a linear phase FIR digital interpolating filter whose output sample values are computed at a different sampling rate than the incoming samples.[7]

Figure 5a shows the frequency response of a 119-tap FIR lowpass filter which was used in the 10 kHz to 16 kHz conversion. Although the length of the filter is 119 samples, only 15 multiplications and additions per output sample are required in the conversion process because only a subset of the filter coefficients are needed in computing each output sample.[7] Similarly, Fig. 5b shows the frequency response of a 127-tap linear phase FIR filter used in the 16 kHz to 10 kHz sampling rate conversion. In this case 26 multiplies and adds are used in computing each output sample.

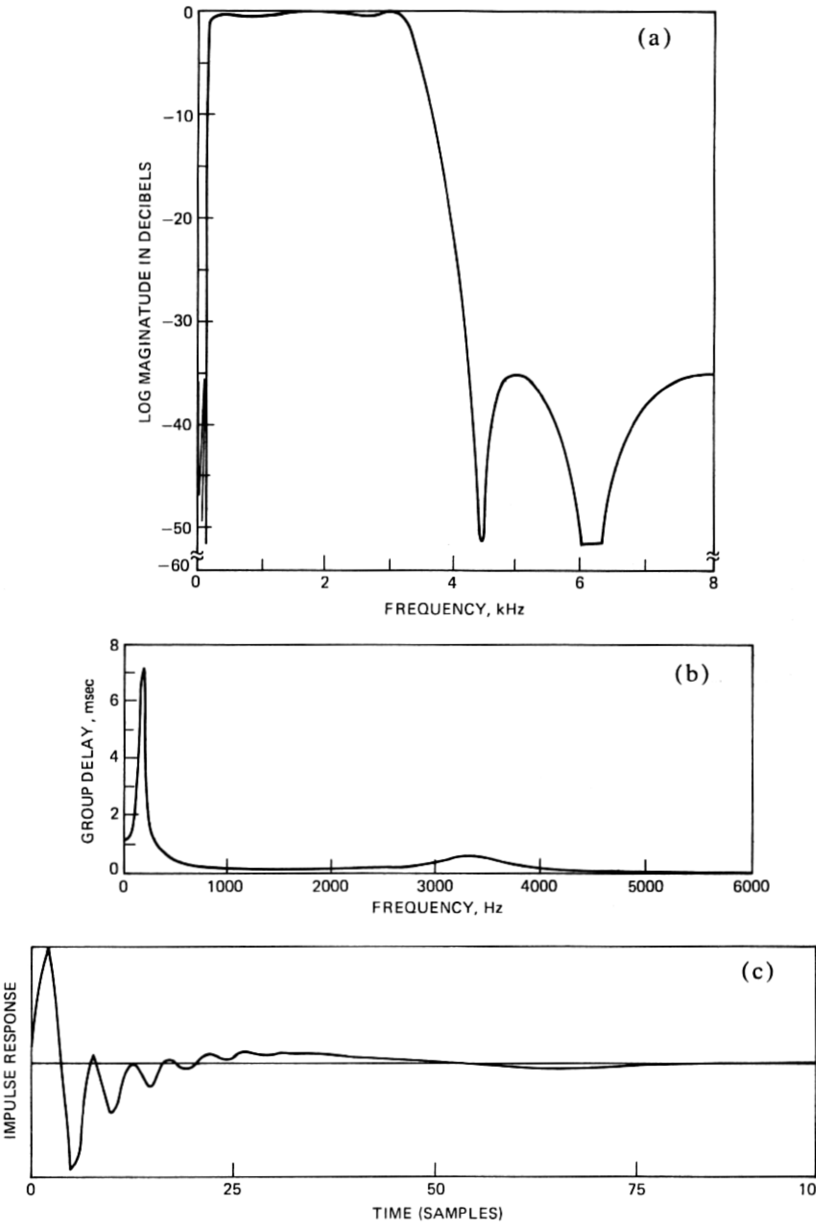


Fig. 4—(a) Log magnitude response. (b) group delay, and (c) impulse response of bandpass filters.

III. FACTORS AFFECTING THE TANDEM LINK

The performance of the LPC to CVSD link is affected by several parameters. Since the LPC vocoder analyzes and then synthesizes the

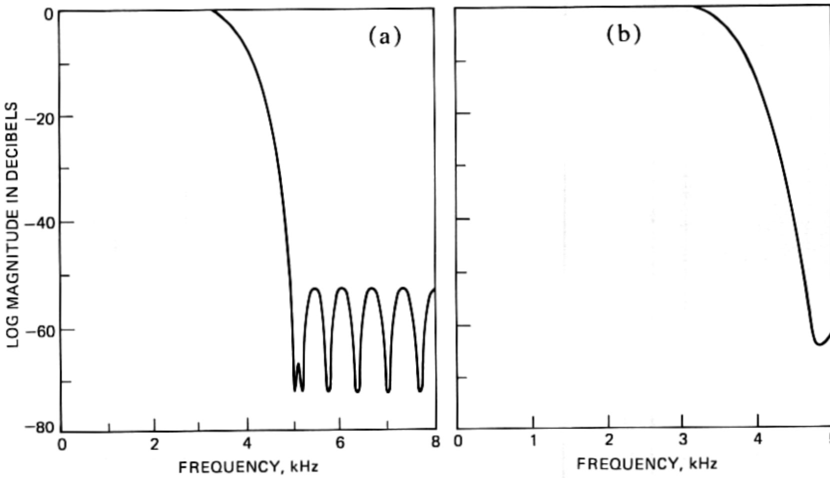
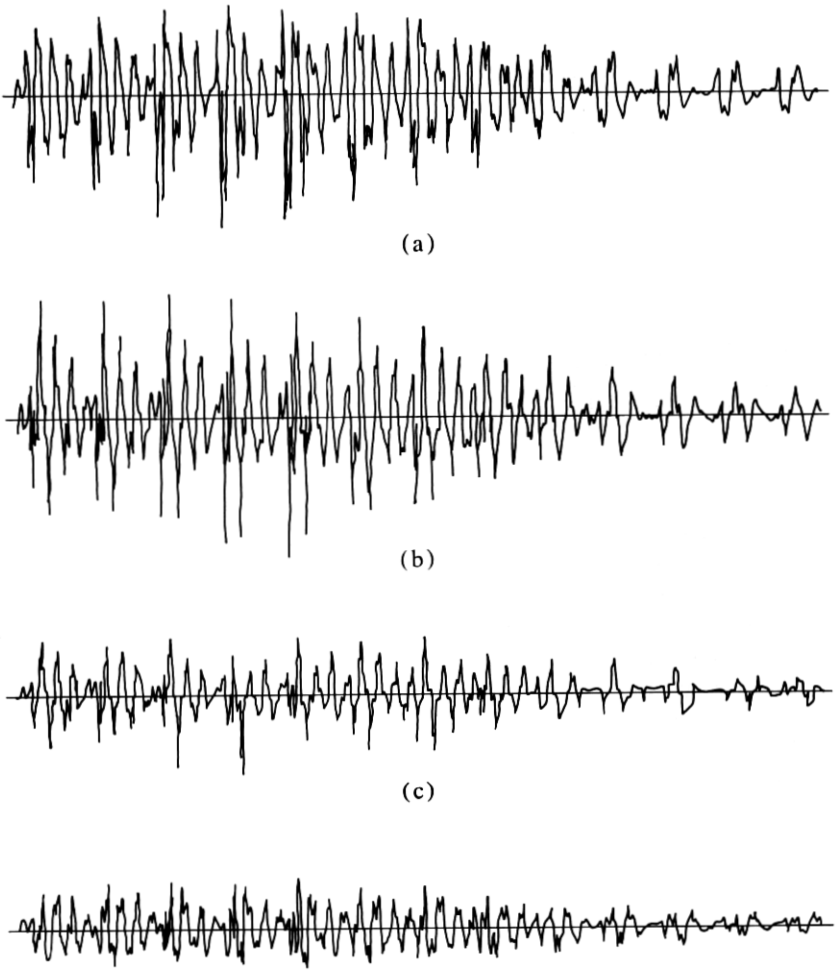


Fig. 5—Frequency response of (a) 119-tap FIR filter for 10:16 sampling rate conversion and (b) 127-tap FIR filter for 16:10 sampling rate conversion.

speech signal, the performance of the CVSD coder will be affected by the manner in which the speech waveform is synthesized. Primarily the performance of the CVSD coder can be affected by factors such as the input level of the speech and the peak factor (ratio of peak-to-RMS value) of the synthesized waveform. Alternatively, parameters in the narrowband system relating to the pitch and the coefficients of the all-pole filter in the synthesis model have little bearing on the performance of the CVSD coder. Therefore, our investigation of the LPC to CVSD link concentrates primarily on the first effects (input level and peak factor).

The input level of the speech waveform determines the operating mode of the CVSD coder. If the input level is too low the coder will be operating in the region in which its performance is determined primarily by granular noise. If the input level is too high the coder will operate in a slope overload condition. Typical waveforms for these coder conditions are shown in Figs. 6–8. Figure 6 shows a complete sequence of waveforms for the wideband system in Fig. 2 under normal (maximum SNR) operating conditions. Figure 6a shows 100 msec of speech appearing at the output of the 10 kHz to 16 kHz sampling rate converter. In Fig. 6b the speech waveform has passed through the first bandpass filter (see Fig. 2) and the effects of bandlimiting and phase distortion can be observed. Figure 6c shows the output waveform of the CVSD coder with the gain $G = 0.158$ which results in maximum SNR. The effects of quantization are clearly noticeable. Finally, Fig. 6d shows the CVSD coder output after bandpass filtering (i.e., the output of the wideband system). Figure 7 shows waveforms for the coder operating in the granular noise region ($G < 0.158$). In Fig. 7a and 7b, waveforms of the unfiltered and band-



(a)

(b)

(c)

(d)

Fig. 6—Speech waveforms for the wideband system. (a) Waveform after 10:16 sampling rate conversion. (b) Waveform after first BPF (input to CVSD). (c) CVSD output waveform (G = 0.158). (d) CVSD output waveform after BP filtering (output of wideband system).

pass-filtered CVSD coder output are shown for a gain setting of G = .009375 or about 25 dB below the maximum SNR operating point. The effects of severe distortion are clearly visible and the speech was completely unintelligible at this point. In Fig. 7c and 7d, waveforms are shown for unfiltered and filtered coder outputs with G = .0395 or about 12 dB below the maximum SNR operating point. Figure 8 shows examples of waveforms for the coder operating in the slope overload region (G > 0.158). In Fig. 8a and b the unfiltered and bandpass-filtered CVSD coder

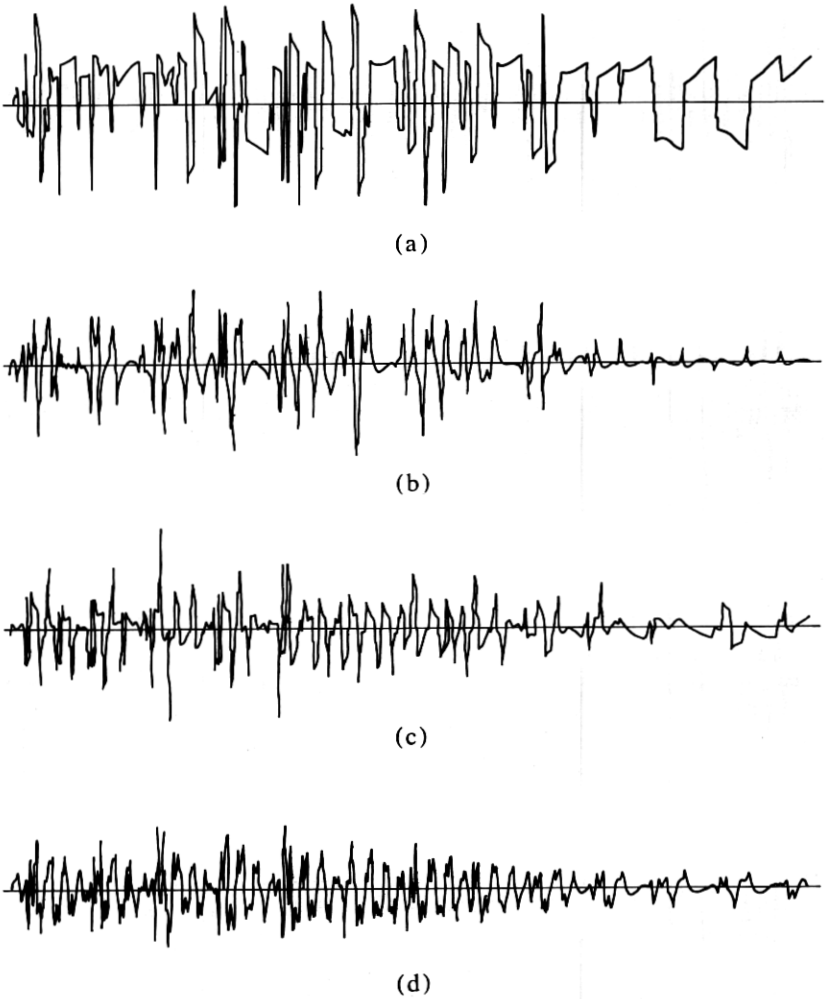
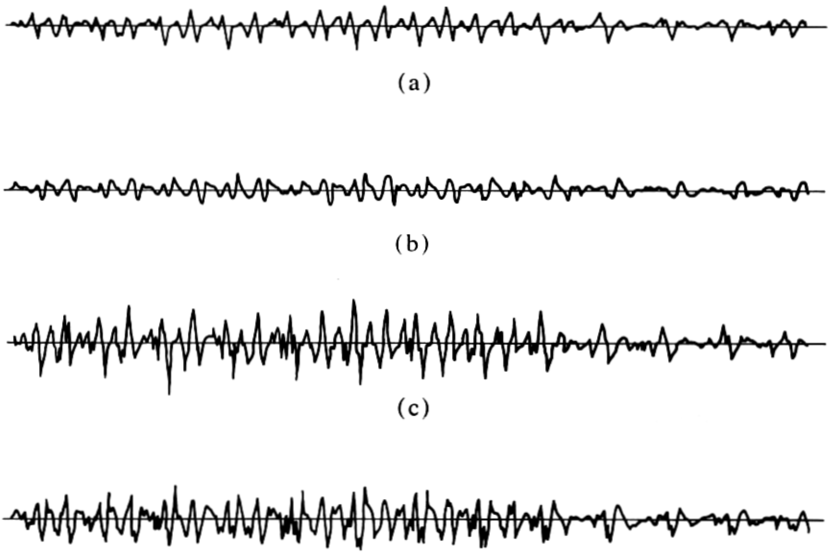


Fig. 7—Output waveforms of the CVSD coder in the granular noise region. (a) Coder output for $G = 0.009375$ and (b) the output after BP filtering. (c) Coder output for $G = 0.0395$ and (d) the same output after BP filtering.

output is shown for $G = 2.528$ or about 24 dB above the maximum SNR operating point. Although the effects of severe slope overload are apparent, the intelligibility of the coder in the slope overload region is not greatly reduced from that at the maximum SNR. Finally, Fig. 8c and Fig. d show unfiltered and filtered output waveforms of the CVSD coder for $G = 0.632$ or about 12 dB above the maximum SNR operating point.

One measure of coder performance is signal to quantizing noise ratio (SNR). The range of input signal level over which the coder maintains an acceptable SNR is often used as a measure of the dynamic range of



(a)

(b)

(c)

(d)

Fig. 8—Output waveforms of the CVSD coder in the slope overload region. (a) Coder output for $G = 2.528$ and (b) the output after BP filtering. (c) Coder output for $G = 0.632$ and (d) BP filtered output.

the coder. The point of optimum SNR is achieved when the coder is on the verge of slope overload.[6] Unfortunately this operating point is not the same as the optimum operating point observed on the basis of subjective performance.[6] Subjectively the noise due to slope overload is less objectionable than the granular noise. Therefore, SNR by itself is not a reliable means for determining the optimum operating region of the coder. More will be said about this in the next section, and in Part 2 (the accompanying paper) another measure of coder performance is proposed which correlates better with subjective performance than the SNR measure.

An important factor affecting the performance of the CVSD, at least in terms of its SNR, is the peak factor of the LPC synthesized speech. The step size of the coder tends to track the RMS level of the input and, if the speech waveform has a large peak-to-RMS ratio, slope overload will cause the peaks to be clipped giving the speech a hoarse sound. If the clipping is severe, intelligibility is degraded.

The peak factor of the synthesized speech can be reduced in several ways to make it more amenable to waveform coding. In one technique the standard pitch source excitation to the LPC synthesizer (an impulse), is modified to spread the energy of the pitch pulse over a larger portion of the pitch period.[8] A pulse which is spread over about 7 percent of the pitch period has been found to be effective for this purpose.[9] Two pulse

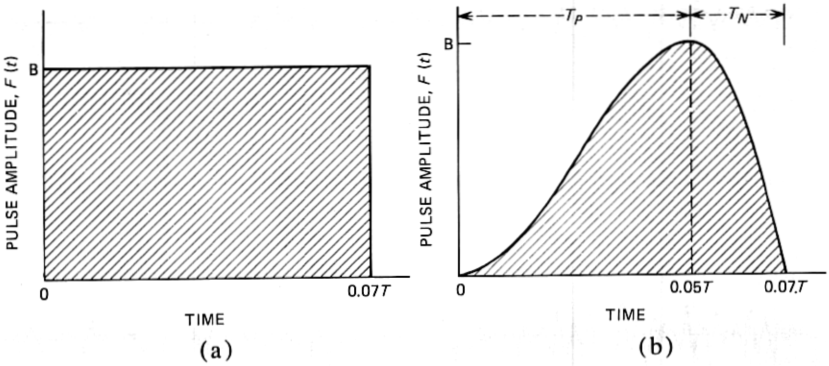


Fig. 9—Pulse excitation sources used for the LPC synthesizer. (a) Rectangular pulse shape. (b) Rounded pulse shape.

shapes were tried in this experiment—a rounded pulse shape and a rectangular pulse shape. The rectangular pulse shape is shown in Fig. 9a. The energy in the pulse is normalized to that of an equivalent impulse. The second pulse shape, shown in Fig. 9b, is a rounded shape proposed by Rosenberg[8] (pulse shape B) to approximate the shape of an actual glottal pulse. T_p is defined as the opening time and T_N is defined as the closing time of the pulse. The pulse shape is then defined by the relation

$$F(t) = B\left[3\left(\frac{t}{T_p}\right)^2 - 2\left(\frac{t}{T_p}\right)^3\right] \text{ for } 0 \leqslant t \leqslant T_p$$

$$F(t) = B\left[1 - \left(\frac{t - T_p}{T_N}\right)^2\right] \text{ for } T_p \leqslant t \leqslant T_p + T_N \tag{15}$$

where $F(t)$ is the height of the pulse and B is its peak amplitude. Values of T_p and T_N used in the experiment are $T_p = 0.05T$ and $T_N = 0.02\mathrm{T}$ where T is the pitch period. The width of the pulse therefore expands or contracts dynamically with the pitch period. The rounded pulse shape was found to give the most natural sound for the LPC synthesized speech.[9]

A second technique that can be used to reduce the peak factor of the LPC sythesized speech is to filter the speech with an allpass filter which disperses the energy of pitch peaks in the waveform. One approach to designing such an allpass filter has been proposed by Rabiner and Crochiere[10] in which the parameters of an allpass filter were optimized to spread the energy of an impulse signal under the limitations of a maximum peak amplitude. This allpass network has been effective in reducing the peak factor of the LPC synthesized speech.

The allpass filter which was used in our experiments was an eighth-order filter which was cascaded three times to give a total allpass filtering

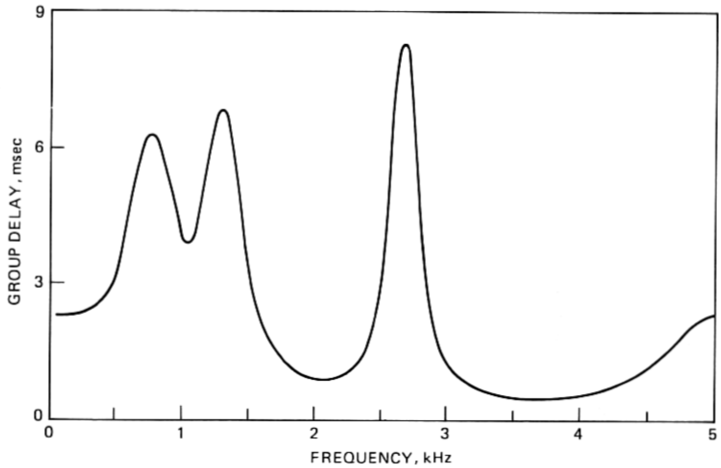


Fig. 10—Group delay of the allpass filter used for preprocessing the CVSD coder input.

equivalent to that of a 24th order filter. The z-transform of each eighth-order filter is of the form

$$H(z) = \prod_{i=1}^{4} H_i(z) \tag{16}$$

where

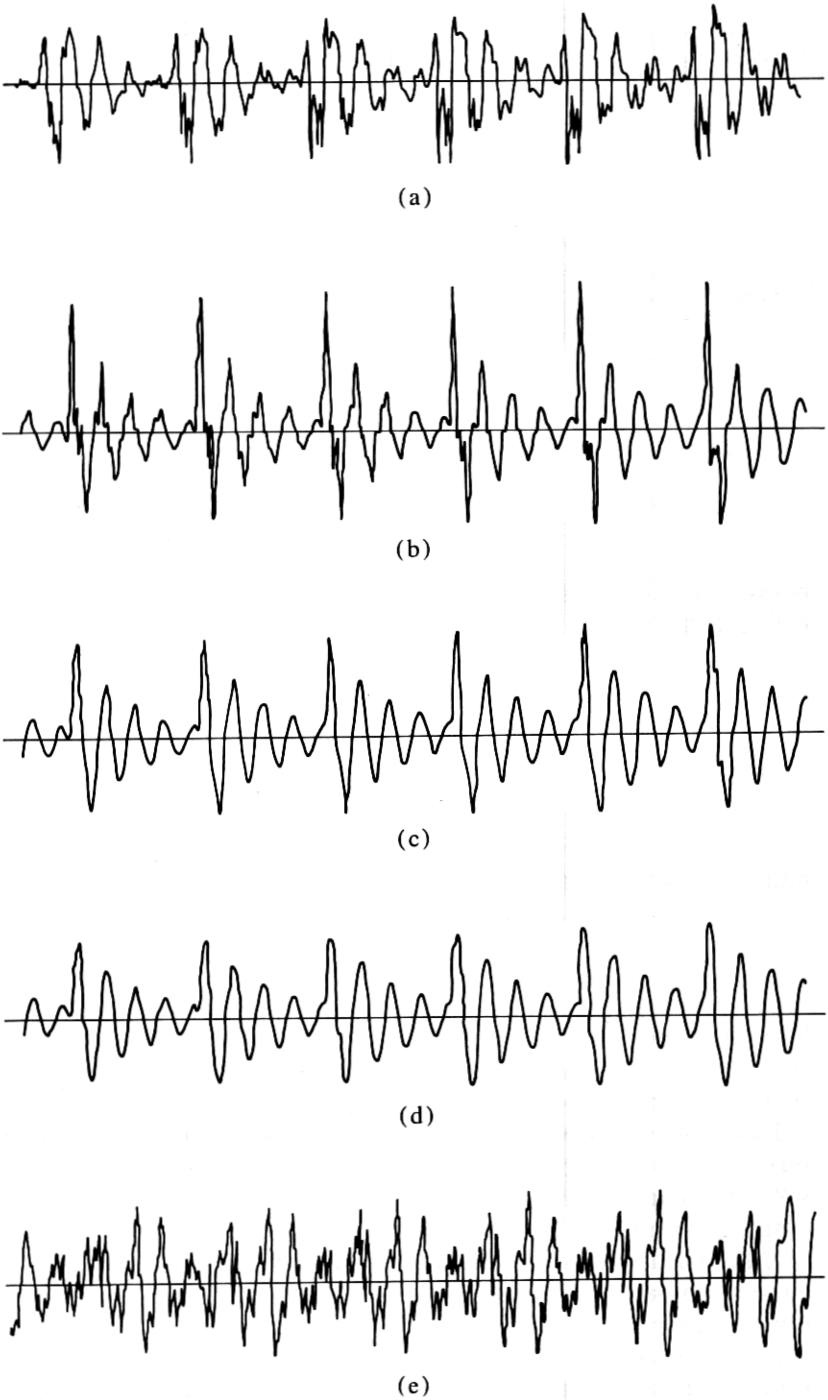
$$H_i(z) = \frac{b_i - c_i z^{-1} + z^{-2}}{1 - c_i z^{-1} + b_i z^{-2}} \tag{17}$$

and the coefficients are

$$\begin{aligned} b_1 &= 0.8149 & c_1 &= 1.2308 \\ b_2 &= -0.4970 & c_2 &= -0.1060 \\ b_3 &= 0.8621 & c_3 &= -0.2135 \\ b_4 &= 0.7870 & c_4 &= 1.5727 \end{aligned}$$

The total group delay of the 24th order all-pass filter is given in Fig. 10. It is seen that the group delay is dispersed between 5 and 90 samples (0.5 to 9 msec) across the frequency band (0 to 5 kHz).

Fig. 11 shows the effects of pitch pulse modifications and allpass filtering on a voiced region of speech. Figure 11a shows the natural speech waveform and Fig. 11b shows an equivalent section of LPC synthesized speech using an impulse excitation. In Fig. 11c and d waveforms are given for LPC synthesized speech with the rectangular and rounded pulse excitations respetively. Figure 11e shows the waveform for the LPC impulse excited speech which was allpass filtered. Finally, Fig. 11f and g show the combination of both allpass filtering and rectangular and rounded source excitations respectively. It is seen that the rectangular or rounded


(a)
(b)
(c)
(d)
(e)

(f)

(g)

Fig. 11—Waveforms of the LPC synthesized speech. (a) Natural speech input. (b) LPC synthesized speech with impulse excitation. (c) LPC synthesized speech with rounded source excitation. (d) LPC synthesized speech with rectangular source excitation. (e) Allpass filtered waveform of LPC speech with impulse excitation. (*f*) Allpass filtered waveform of LPC speech with rounded source excitation. (g) Allpass filtered waveform of LPC speech with rectangular source excitation.

excitation source modifications do improve the peak factor of the speech as does the allpass filtering. The combination gives a further improvement. In the next sections we investigate the effects of these modifications on the performance of the CVSD system.

IV. SNR MEASUREMENTS OF THE CVSD SYSTEM

In this section we report on the performance of the CVSD coder in the tandem link as a function of the signal gain and the modifications of the peak factor of the LPC synthesized speech. Computer simulations were made for the system shown in Fig. 2. Two sentences were used for the simulations. The first sentence, "Every salt breeze comes from the sea," was spoken by a low-pitched male and was recorded off a conventional telephone line. The second sentence, "I know when my lawyer is due," was spoken by another male into a high-quality microphone.

The signal-to-quantizing noise ratio (SNR) of the CVSD coder was measured across the entire sentence. The CVSD noise was obtained by subtracting the filtered output from the CVSD input (also filtered) as shown in Fig. 2. The gain G of the signal was varied from 0.009375 to 2.528 or over a range of approximately 50 dB.

Table I shows the resulting SNRs for the first sentence, "Every salt breeze . . . ." Column 1 corresponds to results for natural speech input to the CVSD coder. Columns 2, 3, and 4 are for LPC synthesized speech using an impulse source, a rounded source, and a rectangular source excitation respectively. Table II gives corresponding SNR's measured with the all-pass filter preceding the CVSD. Tables III and IV pertain

Table I — SNR of CVSD coder vs. gain and source excitation

		Coder SNR* (dB) LPC synthesized speech		
Gain G	Original speech	Impulse source	Rounded source	Rectangular source
0.009375	2.00	1.84	1.77	1.78
0.0395	7.30	7.35	7.89	
0.158	9.29	8.89	10.47	10.62
0.316	7.29	7.23	9.00	
0.632	5.06	5.13	6.39	6.76
1.264	3.31	3.34	4.17	4.43

* For sentence "Every salt breeze comes from the sea."

Table II — SNR of CVSD coder vs. gain and source excitation for allpass filtered inputs

		Coder SNR* (dB) LPC synthesized speech		
Gain G	Original speech	Impulse source	Rounded source	Rectangular source
0.009375	1.57	1.63	1.74	1.84
0.0395	7.53	7.51	7.91	7.94
0.158	9.26	9.67	10.33	10.79
0.316	7.95	8.37	9.41	9.68
0.632	5.83	6.10	7.18	7.53
1.264	3.60	3.84	4.68	4.92
2.528	2.00	2.15	2.63	2.76

* For sentence "Every salt breeze comes from the sea."

Table III — SNR of CVSD coder vs. gain and source excitation

		Coder SNR* (dB) LPC synthesized speech		
Gain G	Original speech	Impulse source	Rounded source	Rectangular source
0.009375	2.52	2.37	2.28	2.31
0.0395	8.93	8.80	8.85	9.06
0.158	11.14	10.77	11.61	12.01
0.316	9.48	8.90	10.01	10.46
0.632	7.07	6.61	7.54	7.81
1.264	4.50	4.38	4.96	5.10
2.528	2.52	2.64	2.95	3.03

* For sentence "I know when my lawyer is due."

to the sentence "I know when . . ." and show measurements corresponding to those in Tables I and II, respectively.

The data indicate that, with or without the allpass filter, CVSD SNR with natural speech input is quite similar to SNR with speech derived from an LPC synthesizer with impulse excitation. (In all four tables the greatest difference between an entry in Column 1 and the corresponding

Table IV — SNR of CVSD coder vs. gain and source excitation for allpass filtered inputs

Gain G	Original speech	Coder SNR* (dB) LPC synthesized speech Impulse source	Rounded source	Rectangular source
0.009375	1.98	1.42	1.69	1.67
0.0395	8.33	8.48	8.91	9.05
0.158	10.59	10.44	11.67	11.49
0.316	9.53	9.02	10.42	10.16
0.632	7.22	6.89	7.84	7.86
1.264	4.59	4.76	5.44	5.55
2.528	2.54	2.75	3.20	3.31

* For sentence "I know when my lawyer is due."

entry in Column 2 is 0.6 dB; most differences are less than 0.4 dB.) Comparing Columns 3 and 4 with Column 2 in the tables we see that broadened pitch pulses lead to 1–2 dB improvements in measured CVSD performance in the slope overload region. As a rule the rectangular pulses result in a slightly higher SNR than rounded ones.

The benefits of allpass filtering are less pronounced than the benefits of broadened pitch pulses. Comparing Column 2 entries (impulse excitation) in Table I and Table II, we see that the allpass filter offers improvements of about 1 dB in SNR at high levels for one sentence. Tables III and IV show virtually no improvement with the other sentence. When the synthesizer uses broadened pitch pulses (Columns 3 and 4) the allpass filter adds 0.5 to 1 dB to CVSD performance with the first sentence and little or nothing to the SNR of the second sentence.

Figure 12 displays the range of possible improvements in CVSD SNR relative to the conventional tandem configuration which includes an LPC synthesizer with an impulse source and no allpass filter at the narrowband-wideband interface. The lower curve in Fig. 12a and b shows CVSD for this configuration for the two sentences in our study. The upper curve in Fig. 12a pertains to the most successful modification of the sentence recorded form a telephone line. This modification involves rectangular pitch pulses and an allpass filter. With the sentence recorded from a high-quality microphone, the best SNR performance, plotted in Fig. 12b, was obtained with the rectangular excitation and no allpass filter.

V. SPEECH QUALITY

Informal judgments of the processed speech suggest that the predominant distortions of tandem circuits are those of CVSD. However, the quality of a vocoder such as LPC depends on speaker and utterance while a waveform coder such as CVSD is relatively insensitive to speech material. Although the utterances used in this work were amenable to

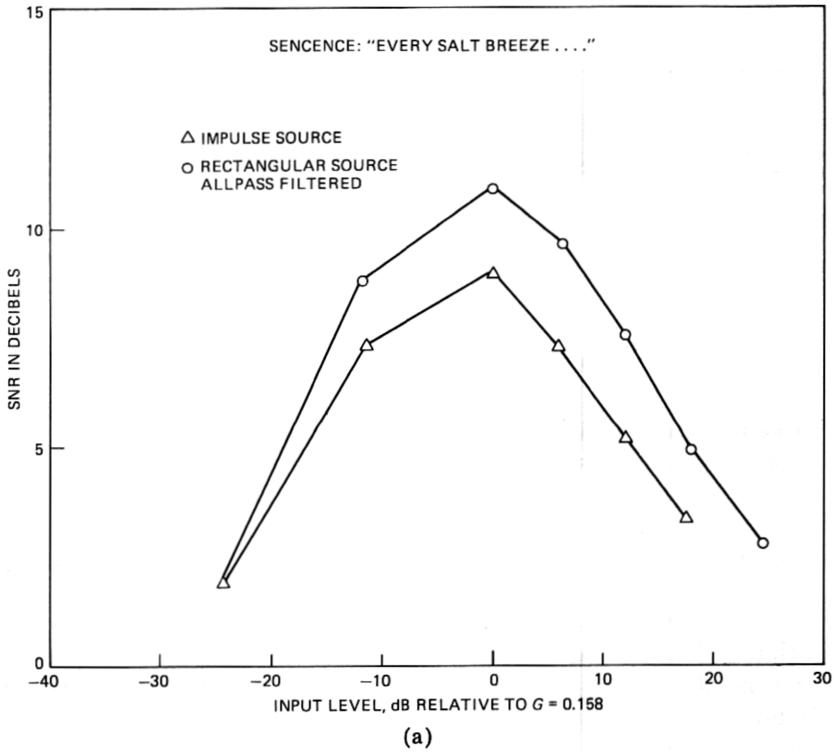


Fig. 12—Summary of the results of the SNR measurements of the CVSD coder, for sentences (a) "Every salt breeze comes from the sea" and (b) "I know when my lawyer is due".

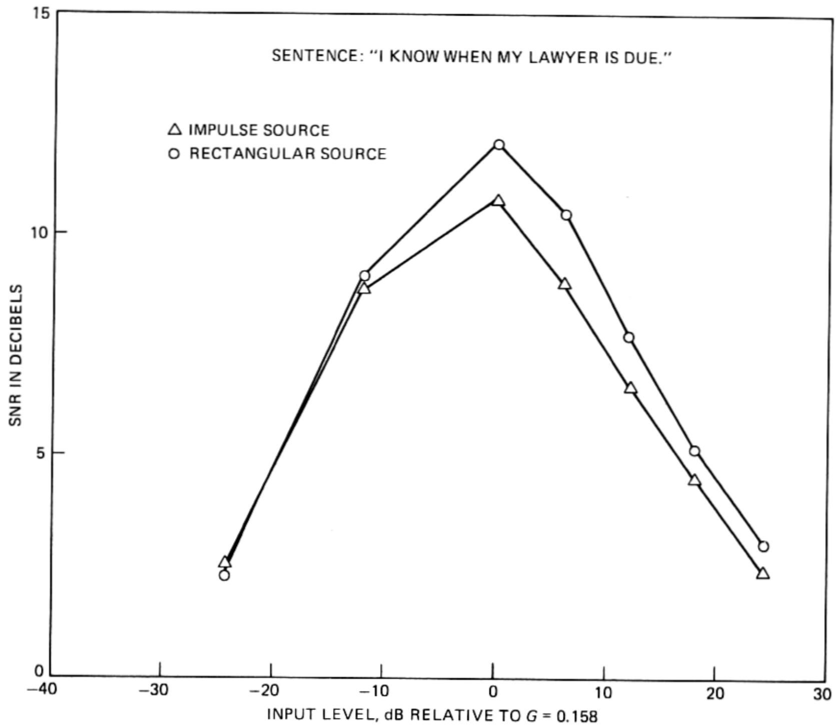
LPC, we anticipate that for certain speakers LPC would be the weaker link in a tandem connection.

As a function of input level, CVSD quality appears to be much lower with weak inputs, which lead to substantial granular quantizing noise, than with strong inputs, for which the main distortion is slope overload. This subjective effect is at variance with SNR indications which show rapidly declining quality as the input level rises into the coder overload range.

The use of broadened LPC excitation pulses lends a more natural quality to the resynthesized speech as well as improving CVSD SNR in the overload region. An allpass filter which also improves SNR for one sentence seems to offer little, if any, enhancement of subjective quality of tandem circuits.

VI. DISCUSSION

Although the conclusions of the previous section must be regarded as tentative, pending formal subjective evaluation of speech processed



(b)

Fig. 12 (*continued*)

in tandem connections, it does appear that efforts to improve the quality of the wideband link would be justified. The CVSD encoder is a 9-year-old design with values of circuit elements chosen to withstand transmission errors occurring at rates as high as 10 percent. If this very demanding requirement is relaxed somewhat and recent advances in delta modulation are incorporated, it may be possible to modify the CVSD to produce higher stand-alone and tandem quality. Alternatively other 16 kb/s wideband coding schemes such as adaptive PCM, adaptive differential PCM or sub-band coding may offer even greater advantages than improved CVSD.[6,11]

REFERENCES

1. R. Steele, "Delta Modulation Systems," London: Peutech Press, 1975 (see Scale system, pp 206).
2. J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech, Berlin: Springer-Verlag, 1976.
3. J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-time digital hardware pitch detector," IEEE Trans. Acoust., Speech, Signal Proc., *ASSP-24* February 1976, pp. 2–8.
4. M. R. Sambur, "An efficient linear prediction vocoder," B.S.T.J., *54,* No. 10 (December 1975), pp 1693–1723.

5. A. H. Gray and J. D. Markel, "Quantization and Bit Allocation in Speech Processing,' IEEE Trans. Acoust., Speech, Signal Proc., *ASSP-24,* December 1976, pp. 459–473.
6. N. S. Jayant, "Digital coding of speech waveforms: PCM, DPCM, and DM quantizers," Proc. IEEE, May 1974, pp. 611–632.
7. R. E. Crochiere and L. R. Rabiner, "Optimum FIR digital filter implementations for decimation, interpolation, and narrowband filtering," IEEE Trans. Acoust., Speech, Signal Proc., *ASSP-23,* October 1975, pp. 444–456.
8. A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Amer., *50,* February 1971, pp. 637–644.
9. M. R. Sambur, A. E. Rosenberg, L. R. Rabiner, and C. A. McGonegal "On removing the buzz in LPC Synthesis," Proc. 1977 IEEE Int. Conf. on Acoust., Speech, Sig. Proc., pp 401–404.
10. L. R. Rabiner and R. E. Crochiere, "On the design of allpass signals with peak amplitude constraints," B.S.T.J., *55,* No. 4 (April 1976), pp. 395–407.
11. 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.