

Design of a Robust Baseband LPC Coder for Speech Transmission Over 9.6 Kbit/s Noisy Channels

VISHU R. VISWANATHAN, MEMBER, IEEE, ALAN L. HIGGINS, MEMBER, IEEE, AND WILLIAM H. RUSSELL, MEMBER, IEEE

Abstract—This paper describes the design of a baseband LPC coder that transmits speech over 9.6 kbit/s (kilobit/second) synchronous channels with random bit errors of up to 1 percent. Presented are the results of our investigation of a number of aspects of the baseband LPC coder with the goal of maximizing the quality of the transmitted speech. Important among these aspects are: bandwidth of the baseband, coding of the baseband residual, high-frequency regeneration, and error protection of important transmission parameters. The paper discusses these and other issues, presents the results of speech-quality tests conducted during the various stages of optimization, and describes the details of the optimized speech coder. This optimized speech coding algorithm has been implemented as a real-time full-duplex system on an array processor. Informal listening tests of the real-time coder have shown that the coder produces good speech quality in the absence of channel bit errors and introduces only a slight degradation in quality for channel bit error rates of up to 1 percent.

I. INTRODUCTION

RECENTLY, there has been a growing interest in developing speech transmission systems that operate at data rates of about 9.6 kbit/s. Several practical reasons exist for this interest, and important among these are: 1) demand on speech quality that is substantially better than is currently available from narrow-band speech coders (data rates below 4.8 kbit/s); 2) unavailability of wide-band channels (data rates above 16 kbit/s) for economical and other reasons; and 3) availability of modems that operate reliably in the range of data rates around 9.6 kbit/s over telephone lines. The types of speech coders that have been investigated for 9.6 kbit/s speech transmission include: voice-excited or baseband coders, subband coders, adaptive predictive coders, and adaptive transform coders. This paper is concerned with baseband coders. A review of the other types of speech coders is provided, for example, in [1].

In a baseband coder, a low-frequency portion (or baseband) of the speech or the residual signal is extracted and transmitted; at the receiver, the baseband signal is used to generate an approximation to the full-band excitation signal, which is then applied to the synthesizer to produce the output speech. The two types of coders, one transmitting the baseband

speech and the other transmitting the baseband residual, may be called, respectively, baseband speech coders and baseband residual coders. Using the linear predictive coding (LPC) method to obtain the underlying spectrum model of the speech signal, we have investigated both types of baseband coders and found that baseband residual coders, in general, produce better overall speech quality than baseband speech coders at the same bit rate [2]. In this paper, we consider only baseband LPC residual coders and, for convenience, we refer to them as BBC.

The objective of our work [3] reported below has been to design a BBC system that 1) transmits good quality speech at a fixed (or synchronous) data rate of 9.6 kbit/s, 2) produces a graceful degradation in speech quality in the presence of random channel bit errors of up to 1 percent, and 3) can be implemented on the CSPI MAP-300 array processor for real-time full-duplex operation. Admittedly, the terms "good quality" and "graceful degradation" provide only a qualitative idea about the nature of the requirements on the coder. In the actual coder design, we optimized experimentally the various aspects of the error-protected BBC system so that it produced the best speech quality at 9.6 kbit/s, while satisfying the following engineering robustness criterion: the speech quality of the error-protected coder at 1 percent channel error must be about the same as or better than the speech quality of the same coder operated without error protection in 0.1 percent channel error. The real-time implementation aspects of the optimized coder are described in another paper [4].

This paper is organized as follows. Section II provides a block diagram description of the BBC system. In Section III, we review briefly previous work on baseband coders. The next five sections deal with the different aspects of the BBC system that we investigated: estimation and coding of the spectral (or LPC) parameters (Section IV); extraction of the baseband from the full-band linear prediction residual (Section V); methods of coding the baseband (Section VI); techniques for regenerating the high band from the transmitted baseband (Section VII); and the speech quality effect of short-term dc bias at each of several places within the coder (Section VIII). While Sections IV, VI, and VII contain some results of speech quality optimization with respect to specific parameters, we present in Section IX the results of parameter tradeoff studies under the 9.6 kbit/s data rate constraint, for the case of error-free transmission. From these results, we chose six BBC systems, which were then compared using a formal subjective speech quality test. The details of the six coders, the test, and the results are all given in Section IX. In Section X, we treat the problem of speech transmission over a noisy channel with

Manuscript received November 2, 1981; revised December 9, 1981. This paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, CO, April 9-11, 1980. This work was supported by the Defense Communications Agency under Contract DCA100-79-C-0003.

V. R. Viswanathan and W. H. Russell are with Bolt Beranek and Newman Inc., Cambridge, MA 02238.

A. L. Higgins was with Bolt Beranek and Newman Inc., Cambridge, MA 02238. He is now with ITT Defense Communications Division, San Diego, CA 92131.

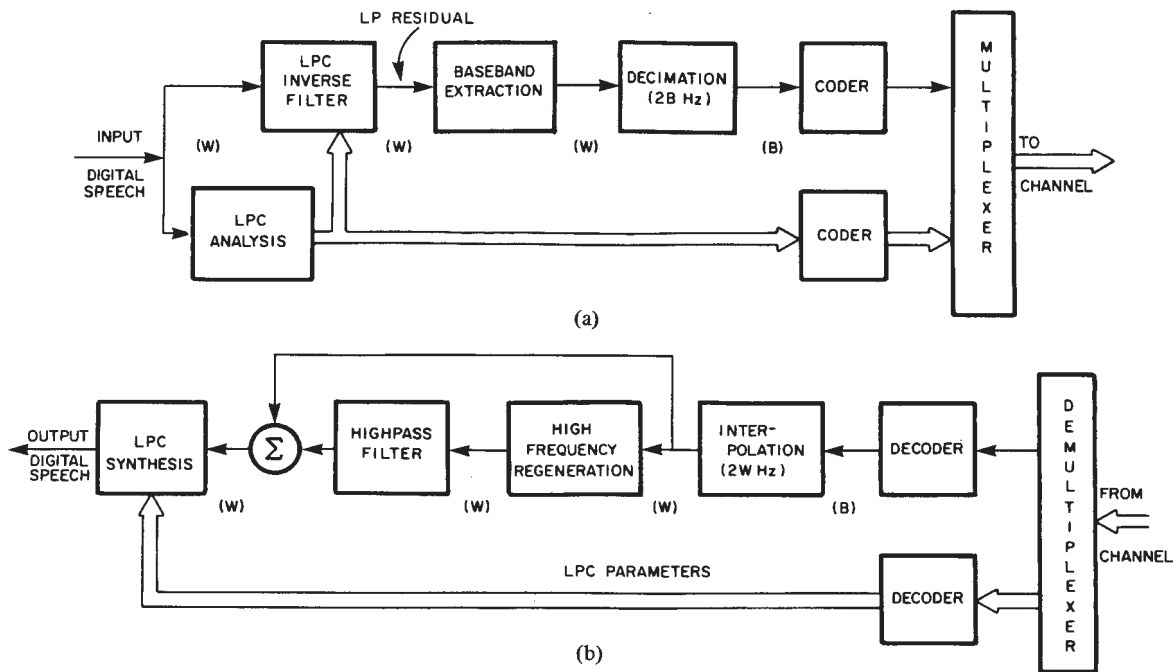


Fig. 1. Block diagram of a baseband residual coder. (a) Transmitter. (b) Receiver.

bit error rates of up to 1 percent and present the results of our work to determine the amount and type of error protection of the transmitted data which satisfies the engineering robustness criterion stated above. Finally, in Section XI, we summarize the details of the optimized 9.6 kbit/s BBC system.

The final design of the 9.6 kbit/s baseband coder devotes about 11.5 percent of the bit rate to the protection of the transmitted coder parameters. This coder design has met the speech quality and robustness goals of our work, and it has been successfully implemented on the CSPI MAP-300 array processor for real-time full-duplex operation. The overall delay between the input and output speech of this real-time speech communication system is about 299 ms. A substantial part of this total delay is due to input and output buffering and the pipelined structure of the transmitter and the receiver, required in the real-time implementation of the coder. This large delay is still within the allowable range for two-way communication using normal conversational protocols [5].

Baseband coders have been investigated by a number of researchers for over two decades. Specific new results and contributions of this work include: development of new high-frequency regeneration methods that yield better speech quality than is possible with existing methods (Section VII); speech quality comparison of several high-frequency regeneration methods, using a formal subjective speech quality test (Section IX), demonstration of a result that coding the baseband signal with an adaptive predictive coder with pitch prediction improves the speech quality considerably in the presence of channel bit errors (Section X); identification of the speech quality effect of short-term dc bias at each of several places within the baseband coder (Section VIII); and investigation of the performance of baseband coders over noisy channels and the design of a robust 9.6 kbit/s baseband coder (Section X).

II. CODER DESCRIPTION

Fig. 1 shows the overall block diagram of a BBC system based on the linear prediction method of spectral analysis. In the figure, we assume that the bandwidth of the speech signal is W Hz and that the bandwidth of the baseband is B Hz ($B < W$). In the work reported here, we have used $W = 3333$ Hz and assumed that the ratio W/B is an integer, denoted by L and called the "number of bands."¹ The terms in parentheses below the signal path in Fig. 1 represent the Nyquist frequency (half the sampling rate) at each of these points. Along the lower branch of the transmitter in Fig. 1(a), LPC parameters are extracted, quantized, and encoded. Along the upper branch, the linear prediction residual is computed by inverse filtering the speech signal with the filter $A(z)$ given by

$$A(z) = 1 + \sum_{k=1}^p a(k)z^{-k} \quad (1)$$

where $a(k)$, $1 \leq k \leq p$, are the predictor parameters and p is the LPC order. A baseband is then extracted from the residual. Baseband extraction usually takes the form of a low-pass filter.² The extracted baseband signal is then decimated (downsampled) to its Nyquist rate of $2B$ Hz, quantized, and encoded. The encoded LPC parameters and baseband signal are multiplexed and transmitted via a communication channel.

¹ For a fixed speech bandwidth, the bandwidth of the baseband can be specified directly or via the number of bands L . We have chosen to retain L as a coder parameter because of its relevance to an important class of high-frequency regeneration methods based on spectral folding (see Section VII).

² If the input of the coder is telephone speech, for example, the lower edge of the extracted baseband may be increased from 0 Hz to, say, 300 Hz, discarding the frequencies below this lower edge. To reduce the sampling rate of the baseband in this case, one may replace the operations of low-pass filtering and Nyquist sampling with a bandpass sampling technique [6], [7].

The receiver of the BBC coder is depicted in Fig. 1(b). The received data are demultiplexed and decoded into the baseband signal (top branch) and LPC parameters (bottom branch). The high-frequency regeneration (HFR) process shown in the figure operates on the baseband to produce a signal with a full-band width W Hz. This process generally involves some nonlinear operation on the baseband (such as waveform rectification), followed by spectral flattening of the distorted signal for the purpose of LPC synthesis. The baseband part of the nonlinearly regenerated fullband signal is, in general, distorted compared to the received baseband. This problem is corrected by a procedure called baseband reintroduction, as shown in Fig. 1(b). The output of the HFR process is high-pass filtered at a cutoff frequency of B Hz and added to the received baseband signal. The resulting signal is then applied to the all-pole LPC synthesizer $H(z) = 1/A(z)$ to produce the reconstructed output speech.

The various operations at the transmitter and at the receiver, mentioned above, are performed over one frame of speech at a time. In our work, we used nonoverlapping frames of 20–30 ms duration.

The data rate and the output speech quality of a BBC system, with a given input speech bandwidth W , are determined by the following items:

- estimation and coding of the LPC parameters
- bandwidth of the baseband B , or equivalently the number of bands, $L = W/B$
- coding of the baseband residual
- method of high-frequency regeneration
- type and amount of error protection of transmission parameters.

Before we consider these items in detail, we give a brief review of the literature on baseband or voice-excited coders.

III. REVIEW OF PREVIOUS WORK

The idea of voice excitation was introduced by Schroeder and David [8]–[10] around 1960 in the context of developing a high-fidelity (analog) channel vocoder, as a way of overcoming the problem of voiced-unvoiced errors and pitch errors that are associated with binary pulse/noise excitation. In [10], a low-frequency band of the speech signal (250–940 Hz) is transmitted as the baseband. The higher frequencies (940–3650 Hz) are modeled by 17 vocoder channels, representing the spectral envelope for those frequencies. A digital version of a voice-excited 16-channel vocoder operating at 9.6 kbits/s has been developed by Gold and Tierney [6].

The use of baseband excitation in LPC coders has been investigated by Un and Magill, who transmitted the baseband residual [11], and by Weinstein, who transmitted the baseband speech [7]. The first of these coders, called RELP (for residual-excited linear prediction), was designed to operate at a data rate of 9.6 kbits/s; the second coder, called VELP (for voice-excited linear prediction), transmitted at 8 kbits/s. Both coders were found to produce speech with a certain “roughness” and “hoarseness” [12]. These effects may be largely due to spectral aliasing caused by the nonlinear distortion performed on the baseband signal to regenerate the high band. Such aliasing can be eliminated by interpolating (over-

sampling) the received baseband to at least twice the Nyquist rate, before the distortion is applied. With this modification, recently suggested by Esteban *et al.* [13], speech quality at 9.6 kbits/s has been reported to be good. A voice-excited LPC coder has also been considered for a low bit-rate transmission at 3.6 kbits/s by Atal *et al.* [14], but the speech quality obtained was not satisfactory.

In all the foregoing investigations, high-frequency regeneration has been accomplished at the receiver by using minor variations of the general method, which is waveform rectification followed by spectral flattening by waveform (peak or center) clipping [10], [14] or double differencing [11] or linear prediction inverse filtering [7]. Recently, a new and computationally simple HFR method, called spectral folding, was proposed by Makhoul and Berouti [15]. Section VII contains the description of the spectral folding method and two new HFR schemes developed in our work.

Coding of the baseband signal has been accomplished using one of the well-known waveform coding techniques: log PCM [7], adaptive PCM [2], [15], adaptive delta modulation [11], adaptive differential PCM [7], subband coding [13], adaptive predictive coding [14], discrete Fourier transform coding [16], or adaptive transform coding [17].

An observation on terminology is in order. Both the RELP and VELP coders mentioned above are voice-excited coders, since both conform to the spirit in which Schroeder and David introduced the idea of voice excitation. The terms “residual-excited” in RELP and “voice-excited” in VELP are thus confusing. This situation has prompted us to use the more obvious and clear terminology of *baseband residual coder* and *baseband speech coder* for RELP and VELP, respectively. In this paper, as mentioned in Section I, we consider only the baseband residual coder.

IV. ESTIMATION AND CODING OF SPECTRAL PARAMETERS

The spectral parameters of the BBC system are the LPC parameters $a(k)$, $1 \leq k \leq p$, that determine both the LPC inverse filter $A(z)$ in (1) used for computing the residual at the transmitter [Fig. 1(a)] and the synthesis filter $H(z) = 1/A(z)$ at the receiver [Fig. 1(b)].

A. Estimation

In a waveform coder that encodes and transmits the full-band residual signal, such as the adaptive predictive coder, the residual signal compensates for any lack of accuracy in estimating the spectral (LPC) parameters. However, in a BBC coder where only a baseband is transmitted, the spectral accuracy at frequencies above the baseband depends exclusively on the accuracy of the LPC parameters in characterizing the spectrum. In this regard, the demands on spectral accuracy in a BBC coder are similar to what one would find in an LPC vocoder using a pulse/noise excitation. In particular, this means that the number of LPC parameters (predictor order) p must be large enough to model the speech spectrum accurately, and the parameters must be quantized so that the model spectrum is minimally distorted. In our work, using the autocorrelation method of linear prediction [18], we

found that for the sampling rate of 6.67 kHz that we used, a predictor order p of 8 was required for transmitting good quality speech. A higher value of p , such as 10, did not produce any noticeable improvement in speech quality. Therefore, we used $p = 8$ in all our investigations of the BBC coder.

While the spectral accuracy mentioned above is of a static nature, there is a dynamic variable that also affects the performance of a BBC system. This variable is the rate of transmission of LPC coefficients in frames per second. Since we use nonoverlapping frames, this transmission rate is, in turn, determined by the frame size. For large frame sizes the spectral parameters attempt to describe a large speech region, which results in a residual with a nonflat short-term spectrum; also, the linear predictor becomes less adaptive to input speech, thus leading to a decline in system performance. A frame size of about 20–30 ms was found to give good results.

B. Coding

The LPC parameters $a(k)$ are to be quantized and encoded. In our work, we used uniform quantization of log area ratios (LAR's), which are obtained by first converting predictor coefficients $a(i)$ to reflection coefficients $K(i)$ and then using the following logarithmic transformation [19]:

$$g(i) = 10 \log_{10} [1 + K(i)]/[1 - K(i)], \quad 1 \leq i \leq p. \quad (2)$$

This procedure has been shown to minimize the maximum error, due to quantization, in the log-power spectrum of the all-pole filter $1/A(z)$ [19]. We employed a total of 33 bits and optimally allocated them among the eight LAR's using a method reported in [19]. The bit allocation we used for $g(1)$ to $g(8)$ is (5, 5, 5, 4, 4, 4, 3, 3).

Since the synthesizer at the receiver will use predictor coefficients obtained from the quantized LAR's, it is important to use the same (quantized) coefficient values in the inverse filter $A(z)$ used in computing the residual at the transmitter.

V. BASEBAND EXTRACTION

The baseband signal is obtained by low-pass filtering the linear prediction residual. The cutoff frequency of the low-pass filter is $B = W/L$. The filtered signal is then decimated to its Nyquist frequency of $2B$ Hz by retaining every L th sample and discarding the others.

Elliptic, Butterworth, or finite impulse response (FIR) filters can be used for this low-pass filtering operation. The first two types of filters produce a frequency-dependent phase delay at the output, while the linear phase FIR filters produce a constant delay that can be easily compensated [20]. To avoid any possible speech quality degradations due to phase distortion, we chose to use a linear phase FIR filter. In all our baseband coder simulations, we employed an FIR filter of order 201, which has a very sharp cutoff characteristic with a transition region of only about 110 Hz width and stop-band attenuation of over 50 dB. This FIR filter introduces a constant phase delay of 100 samples or 15 ms, which we compensated by providing a full frame of delay at the transmitter. At every frame, the full-band residual of the present frame and the baseband residual of the past frame were computed.

VI. CODING OF THE BASEBAND RESIDUAL

The baseband residual can be coded by any of the known waveform coding techniques, as noted in Section III. In our work, we investigated the two techniques: adaptive PCM (APCM) and adaptive predictive coding (APC). Below, we describe the two coding techniques briefly.

A. Adaptive PCM Coder

The APCM coder that we employed has a gain multiplier G followed by an optimum (minimum mean-square error [21]) unit-variance quantizer. G is varied adaptively by setting its value equal to the reciprocal of the rms value of the baseband residual over the current frame. The value of G is quantized before it is used to multiply the baseband residual. We transmitted the quantizer gain G in terms of the energy or mean-square value of the baseband residual; this energy in decibels was quantized using 6 bits. In designing the optimum nonuniform quantizer, we considered three distributions for the quantizer input: Gaussian, Laplacian, and gamma. We chose a Laplacian quantizer because it produced a slightly higher speech quality than either of the other two quantizers.

In our optimization study reported in Sections IX and X, we allowed the use of a noninteger number of bits per sample for baseband residual quantization, e.g., 11 levels per sample. In this case, we combined the quantized levels in a block consisting typically of two to five samples and coded them jointly, using an integer number of bits. For the example of 11 levels per sample, we would code blocks of two samples using 7 bits per block, since a block corresponds to a total of $11 \times 11 = 121$ levels.

B. APC Coder

The APC coder shown in Fig. 2 has a feedback structure with an APCM quantizer placed in the forward path and an adaptive linear predictive filter placed in the feedback path around the quantizer. The adaptive predictor in the feedback path is a pitch predictor, which removes the redundancy due to quasi-periodicity during voiced segments.³ Notice that the quantizer gain G in Fig. 2 is computed from the "second residual," which is obtained by filtering the baseband residual with the pitch-inverse filter $C(z)$. We considered the following two choices for $C(z)$:

$$C(z) = 1 - cz^{-M} \quad (\text{one-tap}) \quad (3)$$

$$C(z) = 1 - c_1 z^{-(M-1)} - c_2 z^{-M} - c_3 z^{-(M+1)} \quad (\text{three-tap}) \quad (4)$$

where c , c_1 , c_2 , and c_3 are pitch predictor coefficients (or taps) and M is the pitch period in number of baseband samples. We computed M as the location of the peak of the autocorrelation function of a 40 ms interval of the baseband residual obtained from the current frame and the trailing part of the preceding frame. For the range of pitch frequencies from 50 to 450 Hz that we considered, M has a range of 5–45 (baseband) samples for $L = 3$ and a range of 4–33 for $L = 4$.

³ Since the baseband residual has, in general, a flat spectral envelope, we did not use a spectral predictor, which is commonly employed in APC coders [22].

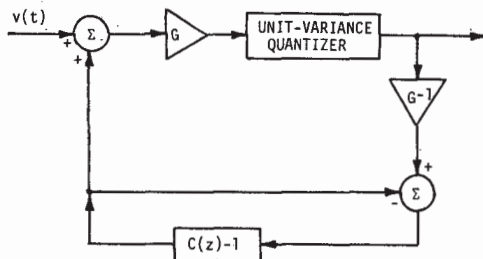


Fig. 2. Adaptive predictive coding of the baseband residual $v(t)$. $C(z)$ is the pitch-inverse filter.

Thus, M can be transmitted without any quantization error in the two cases, using 6 bits and 5 bits, respectively. We used the autocorrelation method of linear prediction for computing the pitch predictor coefficients. For the one-tap case, we quantized c linearly, using 4 bits. For the three-tap case, we quantized linearly $c1$ and $c3$ using 3 bits each, and $c2$ using 4 bits. The energy in decibels of the second residual was quantized using 6 bits and transmitted for computing the gain G at the receiver.

Informal listening tests showed that one-tap and three-tap APC coders produced a significant improvement in speech quality over the APCM coder. Results of an experiment comparing these three cases under a fixed bit rate (9.6 kbits/s) constraint are given in Section IX.

VII. HIGH-FREQUENCY REGENERATION

The HFR problem is stated as follows: given the baseband residual, generate high frequencies so that the full-band excitation signal has a flat spectrum and proper harmonic structure for voiced sounds. Below, we describe briefly the well-known waveform rectification method (minor variations of which have been in use since over two decades ago), a recently introduced spectral folding method [15], and two of the several new methods developed, during the course of our investigation, to improve upon the spectral folding method [3], [23]. The methods described in this section perform the operations of both high-frequency regeneration and baseband reintroduction (see Section II).

A. Rectification

Our implementation of this HFR method consists of the following steps:

- 1) interpolation (upsampling followed by low-pass filtering with a cutoff at B Hz) of the received baseband signal to a sampling rate of $4W$ Hz;
- 2) rectification of the interpolated samples;
- 3) decimation (low-pass filtering with a cutoff at B Hz followed by downsampling) of the rectified signal to the full sampling rate of $2W$ Hz;
- 4) spectral flattening by eighth-order linear prediction analysis and inverse filtering of the decimated signal;
- 5) extraction of the high band by high-pass filtering the spectrally flattened signal (with the cutoff at B Hz); and
- 6) combining the high-band signal with the received baseband signal (at the full sampling rate of $2W$ Hz) to produce the full-band excitation signal.

The higher sampling rate used in Step 1 is necessary to reduce the spectral aliasing from the subsequent nonlinear dis-

tortion (rectification) [13]. Rectification in Step 2 may be performed using a half-wave or a full-wave rectifier, without affecting the output speech quality [15]. The combination procedure in Step 6 involves appropriate scaling and adding of the two flat-spectrum (high-band and baseband) signals so that the resulting signal has a flat overall spectrum rather than a spectrum with a discontinuity at the junction between the two bands [2]. The filtering operations required in Steps 1, 3, and 5 introduce additional frames of delay (see Section V) and make the rectification method computationally rather expensive.

We note that a nonzero short-term (frame-by-frame) dc bias in the interpolated baseband signal can reduce the effectiveness of the rectification method in generating the high frequencies. In the extreme case where the dc bias is so large that all the samples have the same sign, rectification becomes ineffective. Also, the presence of a nonzero dc bias in the decimated signal (after rectification) can reduce the effectiveness of the LPC spectral flattener. Our experimental investigations of the rectification method have produced the following two results:

- 1) removing the short-term dc bias from the signal in the above two cases improves the output speech quality by reducing the amount of perceived roughness; and
- 2) the improved rectification method still produces audible roughness in the form of background noises.

B. Spectral Folding

In this method [15], the process of high-frequency regeneration reduces to inserting $L - 1$ zeros after every baseband sample, where $L = W/B$. This upsampling operation aliases the baseband into the high band. Notice that spectral folding, unlike the rectification method, preserves the baseband. Therefore, the additional steps of high-pass filtering and baseband reintroduction are not required. Furthermore, since the baseband residual is assumed to have a generally flat spectrum, the spectrally folded signal also has a generally flat spectrum. Therefore, the spectral folding method does not require any spectral flattening operation.

Clearly, the spectral folding method is quite simple. However, it produces audible "tonal noises," which increase with the number of bands L and with the pitch of the speaker. A simple modification to the method has been found to largely eliminate a tone in the output speech at multiples of $2B$ Hz. The modification requires the subtraction of the short-term dc in the baseband residual, because the dc is folded into multiples of $2B$ Hz [15], [2]. After spectral folding, the original dc is restored so as not to disturb the average signal level. This modified method, which still produces audible tonal noises at other frequencies, was used in all our speech quality comparison experiments described in later sections.

In our effort to reduce or mask the tonal noises produced by the spectral folding method, we developed several new HFR methods [3], [23], two of which are described below.

C. Spectral Folding with Preflattening

Although the spectrum of the baseband residual is expected to be generally flat, we have observed that occasionally it has a nonflat spectral envelope. Since the high band generated by

spectral folding is made up of aliased versions of the baseband, a nonflat baseband spectrum results in a nonflat high band. This might, in part, be responsible for the tonal noises produced by the spectral folding method. The method of spectral folding with preflattening overcomes the foregoing problem by spectrally flattening the baseband prior to spectral folding. We used a fourth-order LPC inverse filter for the preflattening. The complete method is as follows. The baseband residual is spectrally flattened through LPC inverse filtering, spectrally folded by $L:1$ upsampling, high-pass filtered with a cutoff at B Hz, scaled to compensate for the energy loss due to inverse filtering, and added to the interpolated ($L:1$) version of the received baseband residual. Informal listening tests have shown that this HFR method reduces tonal noises at the expense of slight "pinging sounds."

D. Perturbed Spectral Folding

Simple spectral folding exhibits spectral regularity in the way it reflects the baseband into the higher bands. This regularity might be responsible for some of the tonal noises. The method we call perturbed spectral folding breaks up this spectral regularity. This method is illustrated in Fig. 3. In the lower branch of Fig. 3, the nonzero samples of the spectrally folded (or upsampled) baseband residual are randomly perturbed; perturbation is performed by simply interchanging the sample with an adjacent (zero) sample. The perturbation procedure is as follows. No perturbation is performed if the magnitude of the nonzero sample is larger than a preselected threshold X ; this avoids perturbing samples corresponding to pitch pulses. Nonzero samples with magnitudes less than this threshold are perturbed with a variable probability, with the smaller samples having a higher probability of being perturbed. The perturbed signal is then high-pass filtered and added to the received baseband residual at the full sampling rate of $2W$ Hz (top branch in Fig. 3).

A flowchart of the perturbation scheme is shown in Fig. 4. The values of the thresholds X and C and the parameter D were chosen for best speech quality. The optimized values are: $C = 0.5$ and $D = 0.7$, and X = the long-term rms value of the baseband residual. For the three-band case (i.e., $B = W/3$), this rms value was found to be 35 in our simulation system, which processed 11 bit (including sign) input speech samples.

The perturbed spectral folding method essentially adds a signal-correlated noise to the spectrally folded residual. Perceptually, we have found that this additive noise has the effect of masking the tonal noises. The additive noise also causes slight roughness.

Of the different HFR methods we considered, our informal listening tests indicated that perturbed spectral folding produced the best overall speech quality. This result was also borne out in the formal subjective speech quality test described in Section IX.

VIII. SHORT-TERM dc BIAS

We have found through our experiments that the presence of a nonzero short-term (frame-by-frame) dc bias at each of several points along the signal path in the BBC coder in Fig. 1 could cause speech quality degradations at the coder output.

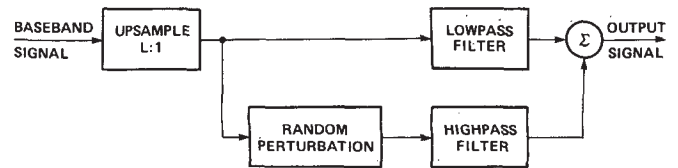


Fig. 3. Block diagram description of the perturbed spectral folding method of high-frequency regeneration. The top branch provides the received baseband at the full sampling rate, while the bottom branch produces the high-frequency region by randomly perturbing the nonzero samples of the spectrally folded input.

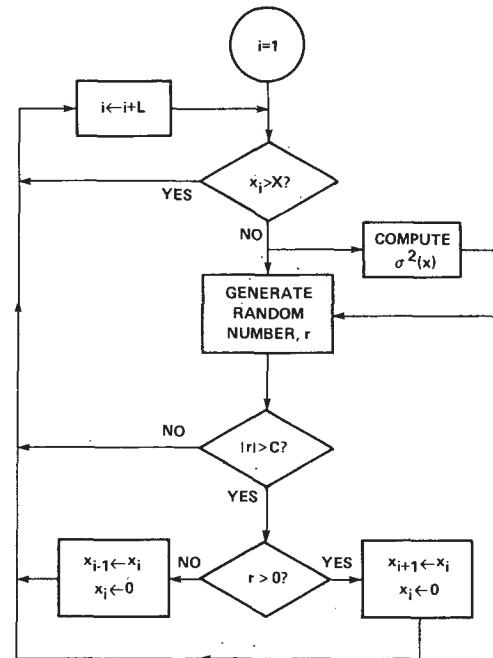


Fig. 4. Flowchart of the scheme for randomly perturbing the nonzero samples of the spectrally folded signal. In our experiments, we generated the random numbers using a zero-mean Gaussian distribution with a variance σ^2 , where $\sigma^2(x)$ is equal to 0 for $|x| > X$, and $D(1 - |x|/X)$ for $|x| < X$. The quantities X , C , and D are the parameters of the HFR scheme, which we optimized experimentally.

We summarize below our main observations.

1) Removing the short-term dc bias from the input speech before LPC analysis [lower branch in Fig. 1(a)] was found to yield reduced background noise in the output speech.

2) If dc is removed before LPC analysis and if the quantizer gain G (see Section VI) is computed as the reciprocal of the standard deviation (rather than the rms value) of the baseband residual, then dc must be removed from the baseband residual before quantizing it; otherwise, continuous overloading of the quantizer (or clipping) can occur in some low-energy regions of speech. If the rms value is used for computing G , then dc removal from the baseband residual is not required.

3) Removing dc from the baseband residual prior to HFR was found to improve the output speech quality, as mentioned in Section VII. For the HFR methods based on spectral folding, high-pass filtering the baseband residual prior to HFR, with the cutoff frequency of the filter below 100 Hz, prevents not only the dc value but also low-frequency components (caused by room acoustics or audio equipment) from being reflected into the higher bands. The speech database used in the formal subjective speech quality test described in Section

IX has such low-frequency components. Without the high-pass filtering mentioned above, the speech processed through the coders contained objectionable roughness and noises. These distortions were eliminated when we used a second-order Butterworth high-pass filter with a cutoff of about 73 Hz. With the use of the high-pass filter, frame-by-frame dc removal is no longer necessary in the receiver.

IX. SPEECH QUALITY EVALUATION

We conducted an extensive investigation of the BBC systems to maximize the output speech quality. This investigation included optimization of the parameters of individual methods for dealing with each of the various aspects of the coder, such as a given HFR scheme, a given baseband coding method, etc. Under the constraint of a fixed data rate of 9.6 kbits/s, tradeoffs were considered between various coder parameters, such as bandwidth of the baseband (or, equivalently, number of bands L), number of levels per sample used for quantizing the baseband residual, frame size (which determines the rate of transmission of LPC and other parameter data), and number of pitch predictor taps used by the APC coder. In this investigation, we used informal listening tests to judge the relative speech quality of competing coders. Also, we considered error-free transmission initially, to get an idea about the range of parameter values to investigate in the subsequent optimization study for noisy channels, which is reported in the next section.

For the tradeoff between number of bands L and number of levels per sample used by the APCM coder, we considered $L = 2-5$ for each of the different HFR methods, and we obtained the best speech quality using $L = 4$ for the rectification HFR method and using $L = 3$ for the HFR schemes based on spectral folding. In another experiment, we compared three BBC systems, which were identical except for the method of baseband coding. These systems employed APCM, one-tap APC, and three-tap APC. The number of quantizer levels per sample for each system was chosen so that its data rate was about 9.6 kbits/s. The average segmental (frame-by-frame) signal-to-quantization-noise (S/Q) ratios for the baseband residual coding in the three cases were found to be 15.6 dB (APCM), 20.7 dB (one-tap APC), and 20.3 dB (three-tap APC). Thus, use of APC produced a substantial increase in the S/Q ratio. Informal listening tests showed that the two APC coders produced a significant improvement in speech quality over the APCM coder and that the one-tap APC system yielded slightly better speech quality than the three-tap system.

In Section VII, we reported that different HFR schemes produced different types of distortions: roughness, tonal noises, and pinging sounds. The informal relative speech quality judgments of these schemes varied over the listeners, the speakers, and the speech material. The relative quality ratings reported in the preceding sections represent merely our best informal estimates. To decide on the HFR scheme to be used in our final system, we felt that it was necessary to perform a more formal evaluation, using a testing procedure that explicitly took into account the variations in judgments over listeners, speakers, and speech material. Below, we de-

TABLE I
DESCRIPTION OF THE SIX 9.6 kbit/s BBC SYSTEMS INCLUDED IN THE SUBJECTIVE SPEECH QUALITY TEST

CODER ID	NO. OF BANDS	NO. OF LEVELS PER SAMPLE	HFR METHOD	BASEBAND CODING
SPF	3	8	SPECTRAL FOLDING	1-TAP APC
FT1	3	8	SPECTRAL FOLDING WITH PREFLATTEING	1-TAP APC
PF1	3	8	PERTURBED SPECTRAL FOLDING	1-TAP APC
FT2	3	10	SPECTRAL FOLDING WITH PREFLATTEING	APCM
WFR	4	22	RECTIFICATION	APCM
PF2	4	22	PERTURBED SPECTRAL FOLDING	APCM

TABLE II
A SET OF SIX SENTENCES USED IN THE SUBJECTIVE SPEECH QUALITY TEST. THE UNDERLINED WORDS ARE EMPHASIZED.

1. Why were you away a year, Roy?
2. Nanny may know my meaning.
3. His vicious father has seizures.
4. Which tea-party did Baker go to?
5. The little blankets lay around on the floor.
6. The trouble with swimming is that you can drown.

scribe the BBC systems selected for this evaluation, the speech quality testing procedure, and the test results.

A. BBC Systems Tested

From the results of the foregoing investigation, we chose six 9.6 kbit/s BBC systems to be evaluated by means of a formal subjective speech quality test. These six systems are defined in Table I, where they are denoted as SPF, FT1, PF1, FT2, WFR, and PF2. As Table I shows, these systems use three or four bands, one of the four HFR methods described in Section VII, and APCM or one-tap APC for coding the baseband residual. The frame size used is 24.0 ms for WFR and PF2, 24.3 ms for FT2, and 24.75 ms for the remaining three systems. Although we tested these coders in the absence of channel bit errors, each coder had a small number of extra bits per frame reserved for error protection of important transmission parameters.

B. Description of the Speech Quality Test

A brief description of the testing procedure is given below. For more details, the reader is referred to [24], [25]. Subjects rated the perceived quality of each system for a set of 36 test sentences. To make the set of 36, we used six sentences, each of which was read by six speakers chosen to represent the full range of speaker variables found in a group of 20 speakers. The sentences used are given in Table II, and the average pitch values of the speakers are given in Table III. The first four test sentences in Table II are phoneme-specific, in the sense that each contains all and only the phonemes of a particular type (glides, nasals, fricatives, and stops), together with vowels. The last two sentences are "general" sentences;

TABLE III
A SET OF SIX SPEAKERS USED IN THE SUBJECTIVE SPEECH
QUALITY TEST

SPEAKER (INITIALS)	MALE/ FEMALE	AVERAGE FUNDAMENTAL (Hz)
DK	M	95
JB	M	118
DD	M	139
AR	F	167
RS	F	209
PF	F	232

which contain several consonant clusters and unstressed syllables. The 216 stimulus sentences (6 systems $\times$ 6 speakers $\times$ 6 sentences) were presented in a specially constructed order designed to counterbalance sequence effects completely. Six subjects were used, three of whom were experienced with the systems under test, and three of whom were naive. Quality judgments were made on an eight-point scale, with 8 representing best quality and 1 representing worst quality.

C. Results of the Speech Quality Test

Mean quality judgments for each system are plotted in Fig. 5. We calculated statistical *t*-scores to determine the significance of the quality difference between any two systems. Briefly, the systems fell into three groups: (PF1 and FT1), (SPF, FT2, and PF2), and (WFR). The differences within groups were insignificant (except that PF1 was just significantly better than FT1, $P < 0.05$), but the differences between groups were very highly significant. When we averaged the quality judgments separately for male and female speakers, we found that the systems SPF, FT2, and PF2 exhibited the most variability between the two mean-quality scores and the system WFR produced the least variability; the remaining two systems FT1 and PF1 produced moderate amounts of variability.

From the results presented above, we have reached the following two conclusions.

- 1) The perturbed spectral folding method is the best HFR method.
- 2) APC coding of the baseband residual noticeably improves the perceived speech quality compared to that obtained using APCM coding.

X. TRANSMISSION OVER A NOISY CHANNEL

One of the goals of this work has been to design a robust 9.6 kbit/s BBC system that tolerates channel bit error rates of up to 1 percent. We used the following engineering criterion as a specific measure of the extent of robustness required of the final BBC system design: the speech quality of the error-protected 9.6 kbit/s coder at 1 percent channel error should be about the same as or better than the speech quality of the same coder when it is operated without error protection in 0.1 percent channel error.⁴ Our study was concerned with op-

⁴ The unprotected, engineering-criterion system will have a bit rate less than 9.6 kbits/s, since it is obtained by discarding the error-protection bits of a 9.6 kbit/s coder.

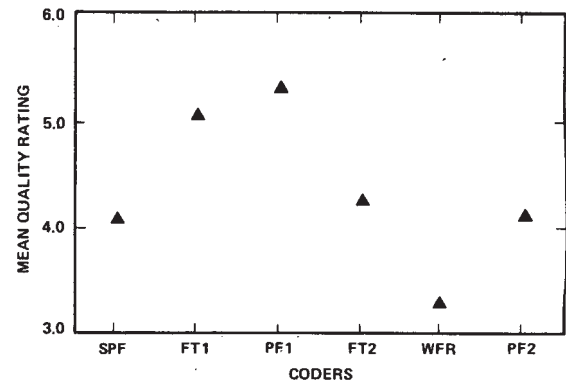


Fig. 5. Mean quality ratings of the six 9.6 kbit/s BBC systems.

timizing 1) the tradeoff between the voice data rate and the error-protection data rate and 2) the distribution of error-protection bits among the individual transmission parameters. We did not protect the coded baseband residual. To protect important parameters of the coder, we used the Hamming (7, 4) code, which protects 4 data bits by adding 3 parity bits; this code detects and corrects all single bit errors in the resulting 7 bit codeword.

Several tests were made to examine how each transmitted parameter, when subjected to 1 percent channel error, would independently affect the output speech. Results of these tests showed that in coders using APCM, channel errors on the baseband residual samples caused the speech to have a continuously rough or "raspy" character. With the use of APC, the roughness in the output speech was reduced considerably. One reason for this improvement provided by APC over APCM is that adequately error-protected and transmitted pitch and pitch tap(s) in the APC system produce the proper pitch periodicity in the output speech; in the APCM system, on the other hand, the channel bit errors on the unprotected residual samples can distort the pitch periodicity. Errors in the LPC coefficients and quantizer gain caused discrete effects such as "pops" and "clicks" in the speech. The block-coding technique, required by the use of noninteger bits per sample (see Section VI-A), was found to produce more noticeable speech quality distortions under channel bit errors than the integer bits per sample case, since a single bit error in the block-coded case resulted in the erroneous decoding of a block of baseband residual samples.

Subsequently, we investigated BBC systems using APC for baseband residual coding and obtained the following results.

- 1) The output speech from coders without any error protection had a "ringing" or reverberant character. Providing error protection for the pitch and pitch predictor tap(s) was found to reduce this effect substantially.
- 2) Considering the two frame rates of 37 and 40 frames/s, we found that the lower frame-rate system produced higher quality speech because of the additional error protection it provided.
- 3) For a similar reason, the one-tap APC case yielded better speech quality than the three-tap case.

Of the various coders we tested, the one-tap APC-coded low-frame-rate system provided the highest quality output speech. For this coder, 30 bits/frame were available for error

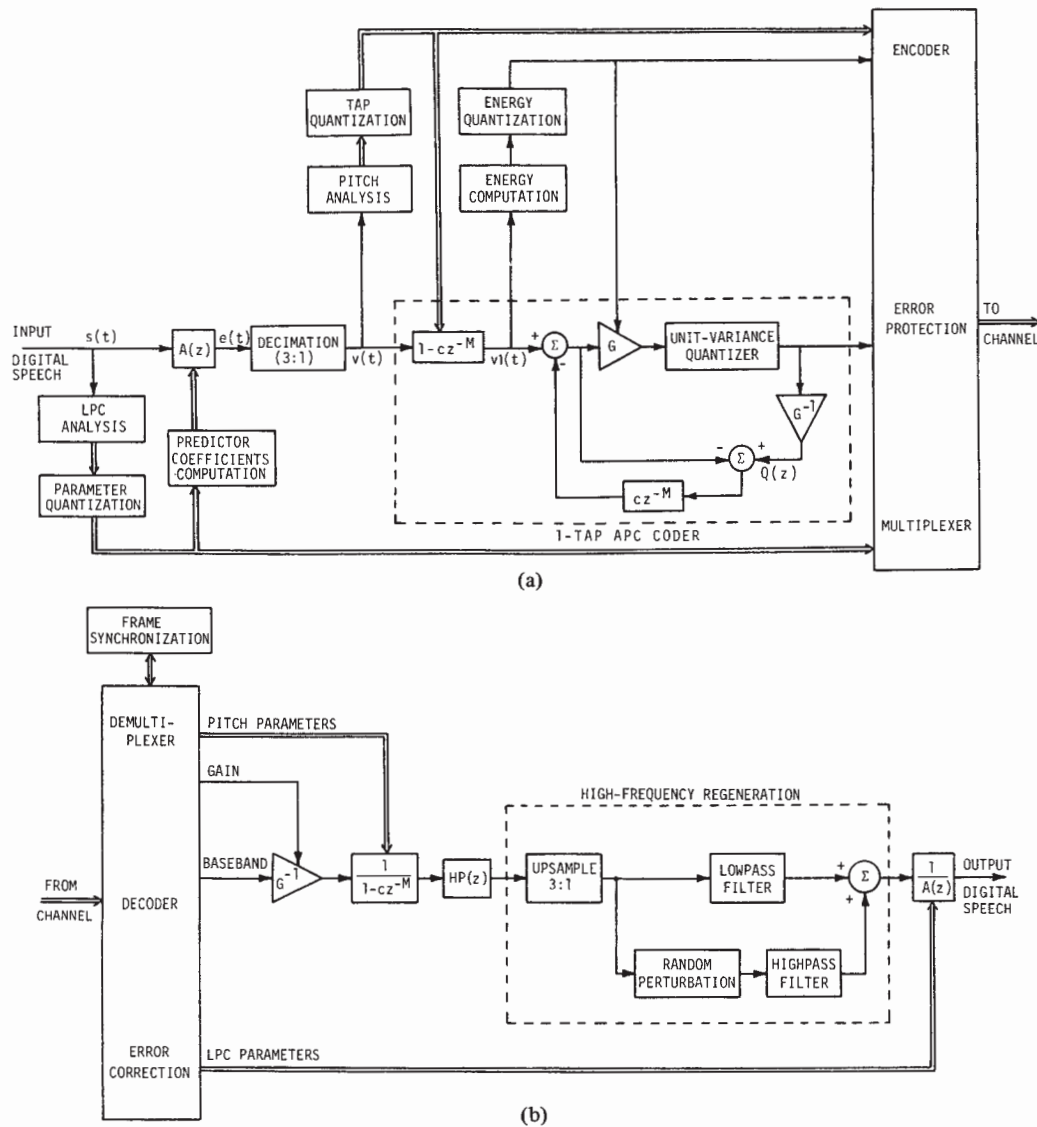


Fig. 6. Block diagram of the optimized 9.6 kbit/s BBC system. (a) Transmitter. (b) Receiver.

protection (i.e., 40 data bits were protected with ten Hamming codewords). The first three LAR coefficients (coded with 5 bits each) and the pitch predictor tap (4 bits) were protected completely. The three most significant bits (MSB's) of the fourth, fifth, and sixth LAR coefficients, one MSB of the seventh and eighth coefficients, and the 5 MSB's of energy (quantizer gain) and pitch were also protected. This error-protection data rate of 30 bits/frame or about 1.1 kbits/s constitutes about 11.5 percent of the channel bit rate of 9.6 kbits/s. This coder was found to satisfy the above-mentioned engineering robustness criterion.

XI. OPTIMIZED 9.6 kbit/s BASEBAND CODER

A. Description of the Coder

A block diagram of the optimized coder is shown in Fig. 6. At the transmitter, the analog input speech is low-pass filtered at 3.33 kHz and sampled at 6.67 kHz. Referring to Fig. 6(a), the sampled speech $s(t)$ is analyzed over 180 speech samples (once every 27 ms). The LPC analysis consists of removing the

short-term dc bias from the speech samples, Hamming windowing, and using the autocorrelation method of linear prediction to compute eight reflection coefficients. The reflection coefficients are quantized using log area ratios (LAR's), with a total of 33 bits. Predictor coefficients are obtained from the quantized LAR's and are used in the inverse filter $A(z)$ to compute the residual $e(t)$. $e(t)$ is decimated 3:1 to obtain the baseband residual $v(t)$ by low-pass filtering $e(t)$ with the stop-band edge at about 1.1 kHz and retaining every third filtered sample. (In practice, one computes every third sample only, to save computation.) Pitch analysis of the baseband consists of computing the autocorrelation function of $v(t)$ for lags 5–45, from an interval of 90 baseband samples (60 from the current frame and 30 from the last frame), determining the pitch value M as the lag corresponding to the peak of this function, and computing the pitch tap c as the ratio of this peak to the value at the zeroth lag. Pitch is transmitted using 6 bits (without quantization), and the pitch tap is linearly quantized using 4 bits. The quantized one-tap filter $(1 - cz^{-M})$ is used to compute from $v(t)$ the second residual

$v_1(t)$. The energy of $v_1(t)$ over the current frame is computed and quantized logarithmically (in decibels) using 6 bits. The gain G within the one-tap APC coder⁵ is set equal to the reciprocal of the quantized energy. The second residual $v_1(t)$ is then coded using the one-tap APC coder, which contains a 3 bit, minimum mean-square error, nonuniform, Laplacian, unit-variance quantizer. The quantized data, APC-coded samples, LAR's, pitch tap, and energy of the second residual, as well as the unquantized pitch, are all binary encoded, error-protected using ten Hamming (7, 4) codewords as indicated in Section X, multiplexed with one synchronization bit, and transmitted over the channel.

At the receiver shown in Fig. 6(b), the received data are demultiplexed, decoded, and error-corrected. The decoded baseband samples are divided by the quantizer gain G and filtered by the pitch-synthesis filter $1/(1 - cz^{-M})$. The output of the pitch filter is high-pass filtered with a second-order Butterworth filter $HP(z)$. High-frequency regeneration is performed next using the perturbed spectral folding method described in Section VII-D. The low-pass and high-pass filters within the HFR process are both FIR filters, with the stopband/passband edge at about 1.1 kHz. The regenerated full-band excitation signal is then applied to the linear prediction all-pole filter $1/A(z)$ to produce the digital speech output. This digital output is passed through a D/A converter and an analog low-pass filter with its cutoff at 3.33 kHz to produce the analog output speech.

B. Real-Time Implementation

The optimized coder has been successfully implemented on the CSPI MAP-300 array processor for real-time full-duplex operation [3], [4]. From the options available on the audio signal interface, we have used 3.2 kHz low-pass filters and a speech sampling rate of 384/58 or about 6.621 kHz. The frame size used for the real-time system is 27.1875 ms (180 samples as in the simulation system). The available number of bits per frame at the 9.6 kbit/s rate is 261, of which 1 bit is unused in the real-time system. To facilitate the real-time implementation, we had to reduce the order of the three FIR filters (one used for baseband extraction at the transmitter and two used in high-frequency regeneration at the receiver) from 201 to 75. The resulting loss in the output speech quality was only slight. The overall delay between the input and output speech in the real-time system is 11 frames or about 299 ms, which includes FIR filtering delays (one frame each at the transmitter and receiver) and delays caused by input and output buffering and by the pipelined structure of the transmitter and receiver real-time implementations [3], [4].

C. Coder Performance

Informal listening tests of the real-time coder have shown that the coder produces good speech quality in the absence of channel bit errors and introduces only a slight degradation in quality for channel bit error rates of up to 1 percent. The

real-time coder was recently tested for speech intelligibility, using the diagnostic rhyme test (DRT) [27].⁶ The DRT scores obtained for high-quality speech input are 89.9 for error-free transmission, 90.8 for 1 percent channel error,⁷ and 83.3 for 5 percent channel error. For error-free transmission, the DRT scores are 89.3 for the coder input in a typical office-noise environment and 82.2 for an airborne command post noise environment.

ACKNOWLEDGMENT

The authors would like to thank J. Makhoul for significantly contributing to the writing of the proposal which led to this project, A. W. F. Huggins for running the formal subjective speech quality test reported in Section IX, and G. Moran of the Defense Communications Agency for suggesting the engineering robustness criterion stated in Section X and for providing the results of the DRT tests of the real-time coder. The optimized coder was implemented in real time by J. Wolf, L. Cosell, and K. Field.

REFERENCES

- [1] J. L. Flanagan, M. R. Schroeder, B. S. Atal, R. E. Crochiere, N. S. Jayant, and J. M. Tribolet, "Speech coding," *IEEE Trans. Commun.*, vol. COM-27, pp. 710-737, Apr. 1979.
- [2] R. Viswanathan, W. Russell, and J. Makhoul, "Voice-excited LPC coders for 9.6 kbps speech transmission," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Washington, DC, Apr. 1979, pp. 568-571.
- [3] R. Viswanathan, J. Wolf, L. Cosell, K. Field, A. Higgins, and W. Russell, "Design and real-time implementation of a baseband LPC coder for speech transmission over 9600 bps noisy channels," Final Rep., Contr. DCA100-79-0003, Bolt Beranek and Newman Inc., BBN Rep. 4327, vol. I: ADA083079, vol. II: ADA083238, Feb. 1980.
- [4] J. Wolf and K. Field, "Real-time speech coder implementation at 9.6 and 16 kb/s," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Atlanta, GA, Apr. 1981, pp. 607-610; also, this issue, pp. 615-620.
- [5] P. T. Brady, "Effects of transmission delay on conversational behavior on echo-free telephone circuits," *Bell Syst. Tech. J.*, vol. 50, pp. 115-134, 1971.
- [6] B. Gold and J. Tierney, "Digitized voice-excited vocoder for telephone-quality inputs," *J. Acoust. Soc. Amer.*, vol. 37, pp. 753-754, April 1965.
- [7] C. J. Weinstein, "A linear prediction vocoder with voice excitation," in *Proc. EASCON 75*, Washington, DC, Sept. 29-Oct. 1, 1975, pp. 30A-30G.
- [8] M. R. Schroeder, "Recent progress in speech coding at Bell Telephone Laboratories," in *Proc. 3rd Int. Cong. Acoust.*, Stuttgart, Germany, 1959.
- [9] M. R. Schroeder and E. E. David, Jr., "A vocoder for transmitting 10 kc/s speech over a 3.5 kc/s channel," *Acustica*, vol. 10, pp. 35-43, 1960.
- [10] E. E. David, M. R. Schroeder, B. F. Logan, and B. F. Prestigiacomo, "New applications of voice-excitation of vocoders," in *Proc. Stockholm Speech Commun. Sem.*, Royal Inst. Technol., Stockholm, Sweden, 1962.
- [11] C. K. Un and D. T. Magill, "The residual-excited linear prediction vocoder with transmission rate below 9.6 kbits/s," *IEEE Trans. Commun.*, vol. COM-23, pp. 1466-1474, Dec. 1975.
- [12] C. J. Weinstein, "Comments on VEP and RELP," unpublished note (ARPA NSC Note 51), 1975.
- [13] D. Esteban, C. Galand, D. Mauduit, and J. Menez, "9.6/7.2 kbps voice excited predictive coder (VEPC)," in *Proc. IEEE Int. Conf.*

⁵ The APC coder shown in Fig. 6(a) uses the noise-feedback configuration [26], which is equivalent to the configuration shown in Fig. 2.

⁶ The tests were performed by Dynastat Inc. for the Defense Communications Engineering Center, U.S. Defense Communications Agency.

⁷ That the DRT score is slightly larger in 1 percent channel error than in 0 percent channel error is not statistically significant.

- Acoust., Speech, Signal Processing*, Tulsa, OK, Apr. 1978, pp. 307–311.
- [14] B. S. Atal, M. R. Schroeder, and V. Stover, "Voice-excited predictive coding system for low bit-rate transmission of speech," presented at the Int. Conf. Commun., San Francisco, CA, June 1975.
 - [15] J. Makhoul and M. Berouti, "High-frequency regeneration in speech coding systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Washington, DC, Apr. 1979, pp. 428–431.
 - [16] G. S. Kang, L. J. Fransen, and E. L. Kline, "Multirate processor (MRP) for digital voice communications," Naval Res. Lab., Washington, DC, NRL Rep. 8295, Mar. 1979.
 - [17] M. Berouti and J. Makhoul, "An adaptive-transform baseband coder," presented at the 97th Meet. Acoust. Soc. Amer., Cambridge, MA, June 1979; Preprint Vol., *Speech Communication Papers*, J. J. Wolf and D. H. Klatt, Eds., pp. 377–380.
 - [18] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.
 - [19] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 309–321, June 1975.
 - [20] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
 - [21] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.
 - [22] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 247–254, June 1979.
 - [23] A. Higgins, R. Viswanathan, and W. Russell, "New high-frequency regeneration (HFR) techniques for voice-excited speech coders," *J. Acoust. Soc. Amer.*, vol. 66, p. S22 (Abstr.), Nov. 1979.
 - [24] R. Viswanathan, J. Makhoul, and A. W. F. Higgins, "Speech compression and evaluation," Final Rep., Contr. MDA903-75-C-0180, Bolt Beranek and Newman Inc., BBN Rep. 3794, ADA055019, Apr. 1978.
 - [25] R. Viswanathan, J. Makhoul, R. Schwartz, and A. W. F. Higgins, "Variable-frame-rate transmission: A review of the methodology and application to narrow-band LPC speech coding," this issue, pp. 674–686.
 - [26] R. Viswanathan, W. Russell, A. Higgins, M. Berouti, and J. Makhoul, "Speech-quality optimization of 16 kb/s adaptive predictive coders," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Denver, CO, Apr. 1980, pp. 520–525.
 - [27] W. D. Voiers, A. D. Sharpley, and C. J. Hehmsoth, "Research on diagnostic evaluation of speech intelligibility," TRACOR Inc., Austin, TX, Tech. Rep. APCR-72-0694, Jan. 1973.



Vishu R. Viswanathan (S'68–M'72) was born in India on November 10, 1943. He received the B.E. degree (with honors) in electrical engineering from Madras University, Madras, India, in 1966, the M.Tech. degree in control systems engineering from the Indian Institute of Technology, Kanpur, India, in 1968, and the M. Phil. and Ph.D. degrees in control and communication from Yale University, New Haven, CT, in 1970 and 1972, respectively.

He was a Junior Research Assistant at the Indian Institute of Technology, Kanpur, during 1967–1968. From 1969 to

1971, he was a Teaching Fellow at Yale University. He was awarded the IBM Fellowship at Yale during 1969–1970 and 1971–1972. For his Ph.D. dissertation, he developed a number of stochastic automata models for application to adaptive and learning control systems. He joined Bolt Beranek and Newman Inc., Cambridge, MA, in 1972, where he has been a Senior Scientist since 1975, conducting research in signal processing, speech analysis, and narrow-band and medium-band speech data compression. He has been Principal Investigator of several projects in these areas.

Dr. Viswanathan is a member of the American Association for the Advancement of Science, Sigma Xi, the IEEE Acoustics, Speech, and Signal Processing, Control Systems, Communications, and Computer Societies, and the IEEE Information Theory Group. He is Vice-Chairman of the IEEE ASSP Society's Technical Committee on Speech Processing. He received the Harding Bliss Award for Excellence in the Department of Engineering and Applied Science at Yale.



Alan L. Higgins (S'76–M'77) was born in San Diego, CA, on November 2, 1952. He received the B.A. and M.S. degrees from the University of California at San Diego in 1974 and 1977, respectively.

In 1977 and 1978, he was a Development Engineer at Scripps Institute of Oceanography, where he developed methods for predicting and measuring near-shore sediment transport on beaches. From November 1978 to July 1981, he was a Scientist at Bolt Beranek and Newman Inc., Cambridge, MA, where he did research and development in the area of speech signal processing and worked on projects in medium-band and narrow-band speech data compression. In August 1981, he joined the ITT Defense Communications Division, San Diego, CA, as a Member of the Technical Staff. At ITT, he is working on projects in isolated word recognition and connected speech recognition. His current interests include digital signal processing and speech communications.

Mr. Higgins is a member of the IEEE Acoustics, Speech, and Signal Processing Society.



William H. Russell (S'65–M'68) was born in Bristol, CT, on June 5, 1943. He received the B.S. degree in electrical engineering from the University of Hartford, Hartford, CT, in 1968 and the M.S. degree in electrical engineering from Northeastern University, Boston, MA, in 1973.

From 1968 to 1973 he was employed at the Raytheon Company, Sudbury, MA, where he was engaged in research on HF propagation and ionospheric disturbances. He has been with Bolt Beranek and Newman Inc., Cambridge, MA, since 1973, working primarily on speech analysis and synthesis, speech data compression techniques, and objective speech quality evaluation.