An All-Digital "Commentary Grade" Subband Coder*

J. D. JOHNSTON AND R. E. CROCHIERE

Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974, USA

Subband coding is a digital encoding method that bandpass filters the signal to be digitized into a set of subbands. Each subband signal is then independently encoded according to perceptual criteria for its frequency band. The digital signals are multiplexed for transmission and demultiplexed and decoded in the receiver. After decoding, the subband signals are summed to give a reproduction of the output signal. The technique permits a greater control over the spectral shape of the digital distortions in the digital encoding system.

This paper presents a design for an all digital two-band subband coder to provide "commentary quality" speech and music over a single slot in a T1 carrier system. It provides a 7-kHz bandwidth at a transmission rate of 56 kb/s with no noticeable distortions or noise under normal "casual listening" environments. It provides a better quality and a more robust performance than a previously discussed analog coder. The design incorporates a number of new digital techniques, including quadrature mirror filtering and high quality analog-to-digital and digital-to-analog conversion circuits.

0 INTRODUCTION

Subband coding is a waveform coding technique in which the input signal is partitioned into two or more frequency bands by bandpass filters [1], [2]. In this paper we will only consider two band designs. Each subband is translated (modulated) to dc (base band), sampled at its Nyquist rate (a sampling rate equal to twice the width of the band), and then digitally encoded. The encoding is usually performed by a technique such as adaptive PCM (APCM) (that is, PCM with an adaptive step size) or adaptive differential PCM (ADPCM) (that is, adaptive PCM encoding of the sample-to-sample difference signal [3]).

This process of dividing the signal into subbands permits each subband to be preferentially encoded according to perceptual criteria for that band. In reconstruction, the subband signals are decoded and bandpass translated (demodulated) back to their original bands. They are then summed to give a replica of the original signal. The process of modulation and demodulation is achieved as a result of the combined operations of sampling and bandpass filtering (sometimes referred to as integer-band sampling [1], [2]).

In this paper we present an all-digital design for the implementation of a subband coder [1], [2] capable of achieving a bandwidth of 7 kHz and a bit rate of 56 kb/s, which is compatible with a standard T1 voice line [4]. The design utilizes ADPCM encoding in subbands, thus providing step-size adaptation and a combination of fixed and dynamic prediction. This work extends the earlier analog design of Johnston and Goodman [5] into a digital form which is more suitable for hardware implementation, incorporates features that were not possible in the analog version, and provides a better overall performance.

The original coder proposed by Johnston and Goodman [5] is illustrated in Fig. 1. The input speech is first divided into two subbands from 0 to 3.65 kHz and from 3.65 to 6.8 kHz, respectively. The subbands are sampled at 8 kHz and 7 kHz for the low and high bands, respectively, and coded with ADPCM coding (with 4 bits per sample) in the low band and ADPCM coding (with 3 bits per sample) in the upper band. The output bits from the coders are multiplexed for transmission over the channel. The inverse set of operations takes place in the receiver.

The all-digital design overcomes a number of difficulties involved in the initial analog approach. By taking advantage of recently proposed techniques of quadrature mirror filters (QMF) [6], the digital approach allows for a precise

* Presented at the 63rd Convention of the Audio Engineering Society, Los Angeles, 1979 May 15–18.

control of the frequency response of the system at the crossover frequency of the two filters (± 0.2-dB ripple versus 0 to −5-dB ripple for the earlier design). In this paper we also propose new designs for the filters in the QMF approach, based on Hanning window designs, which result in the above well-controlled passband behavior of the system as well as a good control over the interband aliasing effects in the system. The QMF approach also eliminates the use of very high Q filters, which presented noise and sensitivity problems in the earlier design. Finally the use of the digital implementation avoids parameter drift and eliminates the need to pairwise tune the receiver with the transmitter.

The "commentary quality" obtained from the coder is suitable for applications such as broadcast services for news, correspondence, sports, and AM radio music transmission. Distortions and noise are not observable under normal "casual listening" environments. The dynamic range of the coder is approximately 60 dB with less than 10% distortion and a bandwidth of approximately 7 kHz.

1 CONSIDERATIONS IN THE DESIGN OF THE ALL-DIGITAL SUBBAND CODER

A properly designed all-digital implementation of the two-band coder avoids most of the difficulties encountered in the earlier analog approach. In this section we consider the various alternatives offered by the digital approach and discuss their merits. In the next section we will present parameters for a final design based on these considerations.

1.1 Implementation of the Subband Filters

One of the most important aspects in the design of the digital coder is the implementation of the filters. Although a variety of alternatives exist for the implementation of these filters, the QMF approach appears to be the most satisfactory. For example, a direct IIR (infinite impulse response [7]) digital filter approach which duplicates the analog approach may be used. However the high Q's required for the IIR filters give rise to large-order limit cycles (nonlinear oscillations) and require large digital word lengths for the representation of the coefficients and data. A direct FIR (finite impulse response [7]) digital filter approach may also be used to duplicate the analog approach. Although this approach avoids the limit cycle problem and permits the use of shorter word lengths, it requires the use of large order (200-tap) transversal filters which are too costly for practical hardware implementation.

By using the QMF bank approach proposed by Esteban and Galand [6], the requirement for sharp cutoff filters is eliminated and the order of the FIR filters can be reduced to 32 taps without affecting the performance of the coder. Fig. 2(a) illustrates the basic concept of the QMF bank and Fig. 2(b) depicts a spectral interpretation of its operation. It is assumed that the input signal to the coder is first sampled and quantized to a PCM format (see Section 1.5). This signal is denoted by $x(n)$, where n is the sample time ($n = 0,1,2,\cdots$). The digital input signal $x(n)$ is then digitally filtered by two filters with impulse responses denoted by $h_1(n)$ for the low-frequency band and $h_2(n)$ for the high-frequency band.

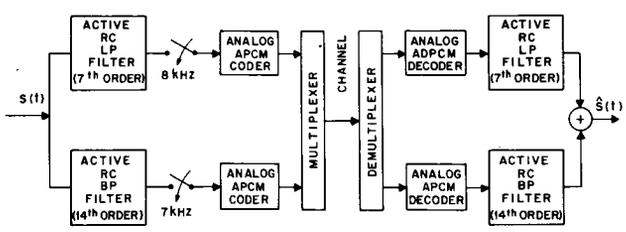


Fig. 1. Analog subband codec design.

According to the design requirements for the QMF [5], these are FIR filters (sometimes called transversal filters), that is, $h_1(n)$ and $h_2(n)$ are zero for $n < 0$ and $n > N - 1$, where N is the number of taps in the filter. Furthermore N is required to be an even number. Both $h_1(n)$ and $h_2(n)$ are obtained from a lowpass prototype design $h(n)$ such that

$$h_1(n) = h(n), \qquad n = 0,1,2,\cdots, N-1 \tag{1a}$$

$$h_2(n) = (-1)^n h(n), \qquad n = 0,1,2,\cdots, N-1 \tag{1b}$$

that is, $h_1(n)$ is identical to $h(n)$ and $h_2(n)$ has the same coefficient values as $h(n)$, except that the signs of odd-numbered coefficients are reversed. Thus the design of one lowpass filter $h(n)$ specifies all the coefficient values in the transmitter and receiver filters.

The lowpass prototype filter $h(n)$ must be a symmetrical filter in order to meet the QMF bank requirements [6], that is,

$$h(n) = h(N-1-n), \qquad n = 0,1,2,\cdots, N/2-1. \tag{2}$$

From Eqs. (1a) and (1b) it is seen that this requirement yields a symmetrical impulse response for $h_1(n)$ (symmetrical about $n = (N-1)/2$) and an antisymmetrical impulse response for $h_2(n)$ (antisymmetrical about $n = (N-1)/2$).

The z transforms of the impulse responses [7] describe the frequency behavior of the filters, and we will let $H(z)$, $H_1(z)$, and $H_2(z)$ denote z transforms of $h(n)$, $h_1(n)$, and $h_2(n)$, respectively:

$$H(z) = \sum_{n=-\infty}^{\infty} h(n) z^{-n}. \tag{3}$$

When $z = e^{j\omega}$, $H(e^{j\omega})$ becomes the Fourier transform (the z transform evaluated on the unit circle) and ω denotes the frequency [in radians] normalized to the sampling fre-

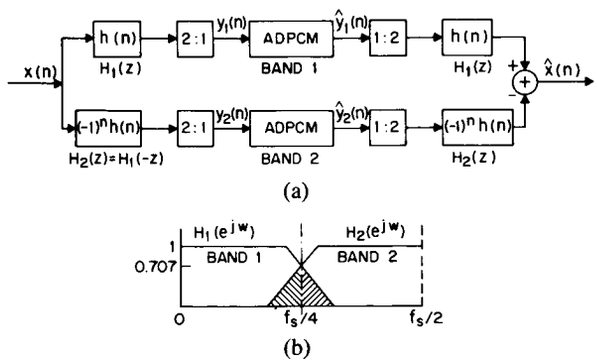


Fig. 2. (a) Quadrature mirror filter bank. (b) Spectral interpretation of its operation.

quency. From Eqs. (2) and (3) it can be shown that

$$H_1(z) = H(z)$$
$$H_2(z) = H(-z) \tag{4a}$$

or equivalently,

$$H_1(e^{j\omega}) = H(e^{j\omega}) \tag{5a}$$
$$H_2(e^{j\omega}) = H(e^{j(\pi-\omega)}). \tag{5b}$$

Thus it is seen that the two filters $H_1(e^{j\omega})$ and $H_2(e^{j\omega})$ are mirror images about $\omega = \pi/2$ [or equivalently the frequency $f_s/4$, where f_s is the sampling frequency as illustrated in Fig. 2(b)].

The outputs of the filters are decimated (subsampled) by a factor of 2 (that is, every other sample is discarded), producing the signals $y_1(n)$ and $y_2(n)$. They are encoded with ADPCM coders. In the receiver the sampling rate is restored (by inserting a zero-valued sample between each pair of decoded samples $\hat{y}_1(n)$ and $\hat{y}_2(n)$. They are then filtered using a similar pair of filters $H_1(z)$ and $H_2(z)$. The reconstructed output $\hat{x}(n)$ is then obtained as the difference of the two subband signals, as seen in Fig. 2(a). Note that with this approach the choice of the subbands is slightly different from that of the earlier design in Fig. 1. In particular, with an input sampling rate of f_s = 14 kHz, the lower band extends from 0 to 3.5 kHz and the upper band from 3.5 to 7 kHz.

It can be shown that the magnitude of the overall frequency response of the system [from $x(n)$ to $\hat{x}(n)$], denoted by $|G(e^{j\omega})|$, has the form [6]:

$$|G(e^{j\omega})|^2 = |H_1(e^{j\omega})|^2 + |H_2(e^{j\omega})|^2. \tag{6}$$

In the frequency range 0 to $f_s/4$, $|H_1(e^{j\omega})| \approx 1$ and $|H_2(e^{j\omega})| \approx 0$, and in the frequency range $f_s/4$ to $f_s/2$, $|H_1(e^{j\omega})| \approx 0$ and $|H_2(e^{j\omega})| \approx 1$ [see Fig. 2(b)]. At the crossover frequency $f_s/4$ (that is, $\omega = \pi/2$) both filters contribute equally such that $|H_1(e^{j\pi/2})|^2 = |H_2(e^{j\pi/2})|^2 = 1/2$. Thus the lowpass prototype filter $H(e^{j\omega})$ should have an attenuation of 3 dB at this frequency. Under these conditions it is seen that

$$|G(e^{j\omega})|^2 \approx 1. \tag{7}$$

Because the filters $H_1(z)$ and $H_2(z)$ are of relatively low order in practice, they have relatively broad transition bands, and a significant amount of aliasing [7] occurs in each subband after decimation in the transmitter. This aliasing energy folds back into the subbands as depicted by the shaded regions in Fig. 2(b). At the receiver, interpolation with filters $H_1(z)$ and $H_2(z)$ leaves significant energy outside of each band. However, because of the special nature of the QMF bank described by Eqs. (1)–(7), the aliasing components which occur due to the decimation process in the transmitter are exactly canceled (to the level of the quantization noise of the coders) by the interpolation process in the receiver. An analysis of the circuit, as given in [6], shows that the output signal $\hat{x}(n)$ is a delayed version of the input signal $x(n)$ with some added quantization noise.

As seen in Eq. (2), the mirror image relationship of the filters $H_1(z)$ and $H_2(z)$ results in the condition that their coefficients are identical, except that their signs alternate [Eqs. (2a) and (2b)]. This property allows both filters to be simultaneously implemented with a single 32-tap filter by accumulating even and odd partial products and computing $y_1(n)$ and $y_2(n)$ as the respective sums and differences of these partial products, as seen in Fig. 3. Another factor of 2 is saved in the implementation by computing only the decimated outputs of the filters, that is, every other sample. A similar efficient implementation for the receiver which takes advantage of the mirror image properties is also possible [6].

1.2 Considerations in the Design of Filters for the QMF Bank

Although Eqs. (1)–(7) and the fact that the filters must be symmetric even-order FIR filters specify a number of constraints on the QMF design, they do not specify an exact filter design. In this section we consider alternatives for this design.

Our first approach to the design of the prototype filter $H(z)$ was to use a 32-tap optimal equiripple design based on the well-known Remez exchange algorithm [8]. (This class of filters is often used in digital signal processing applications.) This resulted in the filter characteristic shown in Fig. 4(a). The filter has a passband ripple of ±0.76 dB and a stopband attenuation of 38 dB. To check the performance of this filter for the coder, the impulse response of the circuit of Fig. 2(a) was measured (by computer simulation) with the ADPCM coders bypassed, that is, $\hat{y}_1(n) = y(n)$ and $\hat{y}_2(n) = y_2(n)$. The total frequency response $G(e^{j\omega})$, shown in Fig. 4(b), has a ripple on the order of ±1.5 dB, which is not satisfactory for music performance.

Another undesirable property of the equiripple filter is that the aliasing components from one subband occur with equal weight in the other subband, independent of their frequency (except at the transition region where they are greater). For example, signal components around 500 Hz in the lower subband produce 6500 Hz aliasing components in the upper subband, which are attenuated by only 38 dB. Since the spectra of most speech and music sounds tend to fall off with increasing frequency, this implies that this aliasing component can represent a significant part of the overall signal energy at 6500 Hz, as seen by the upper subband. Although this aliasing is canceled at the receiver, it can interfere with the efficient coding of the actual signal in the upper subband.

To avoid the problems with the equiripple filter, a prototype filter design based on a 32-tap Hanning window

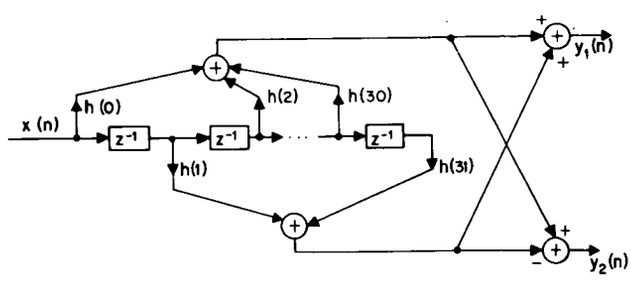


Fig. 3. Efficient implementation of the quadrature mirror filter bank in the transmitter.

design [7] is proposed for QMF designs. This filter has the frequency response characteristic shown in Fig. 5(a). Although the transition region from the passband to the stopband is broader than in the equiripple filter, it can be seen that the stopband attenuation increases with increasing frequency. Thus aliasing between frequency components which are widely separated is attenuated by a larger amount than frequencies that are closely spaced. For example, aliasing between 500 Hz and 6500 Hz is attenuated by about 70 dB, whereas aliasing between 3000 Hz and 4000 Hz is attenuated by only about 20 dB. Thus the Hanning filter provides a better control of the aliasing between subbands.

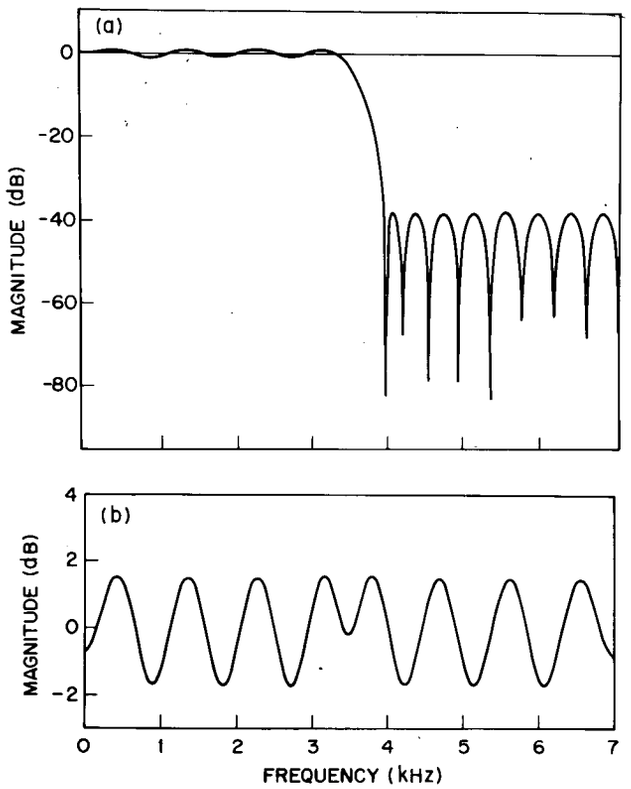


Fig. 4. (a) Frequency response of the 32-tap equiripple filter. (b) Overall frequency response of coder based on this filter design.

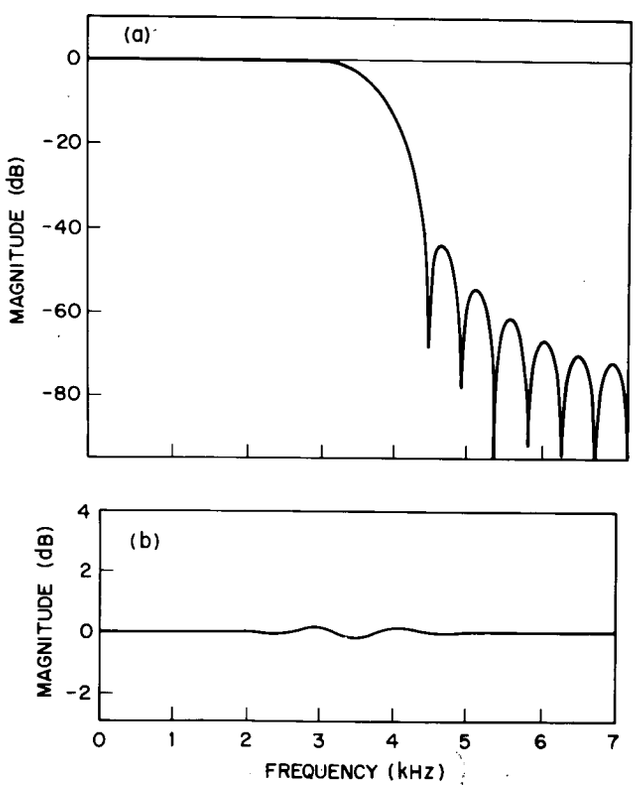


Fig. 5. (a) Frequency response of the 32-tap Hanning filter. (b) Overall frequency response of coder based on this filter design.

A second advantage of the Hanning filter design is seen in Fig. 5(b) which shows the overall frequency response $G(e^{j\omega})$ of the coder with Hanning filters. Because of the nearly flat frequency response of the Hanning filter in the passband, the largest ripple, measured near the transition region, was only on the order of ± 0.2 dB, which is much better than that of the equiripple filter in Fig. 4(b). The Hanning filter design is therefore recommended as a good choice for the design of a QMF subband coder.

Table 1 gives the coefficient values $h(0)-h(15)$ for the first symmetric half of an $N = 32$ tap prototype filter $h(n)$, which was used in our final design. Values for $h(16)-h(31)$ can be obtained from Eq. (2), and values for $h_1(n)$ and $h_2(n)$ can be obtained from Eqs. (1a) and (1b). It should be noted that this is not a Hanning filter design but, in fact, a design that was further tailored to our needs (a Hanning design also works satisfactorily, however). It results in an overall ripple in $|G(e^{j\omega})|$ of $\pm$ 0.025 dB. In the stopband the largest ripple (nearest to the transition band) peaks at -38 dB, and the smallest ripple (farthest from the transition band) peaks at -49 dB. The width of the transition region between the passband and the stopband is approximately 600 Hz (assuming a 14-kHz sampling rate).

1.3 Considerations of Subband Partitioning and Intermodulation Distortion

Dividing the speech band into subbands and coding each subband separately permits each subband to be preferentially encoded according to perceptual criteria for that band. Quantizing noise can be contained within each band and prevented from masking signals outside of the band. As more subbands are used, however, more aliasing components must be canceled between subbands, particularly when using the QMF approach with lower order (16–32-tap) filters. (The QMF approach can be extended to more than two subbands by further partitioning each of the two

Table 1. Coefficients for a 32-Tap ($N = 32$) Prototype Filter for a QMF bank.

$h(0)$	= 0.002245
$h(1)$	= −0.003971
$h(2)$	= −0.001970
$h(3)$	= 0.008182
$h(4)$	= 0.000843
$h(5)$	= −0.014229
$h(6)$	= 0.002069
$h(7)$	= 0.022704
$h(8)$	= −0.007962
$h(9)$	= −0.034964
$h(10)$	= 0.019472
$h(11)$	= 0.054812
$h(12)$	= −0.044524
$h(13)$	= −0.099339
$h(14)$	= 0.132973
$h(15)$	= 0.463674

subbands according to a "tree" structure [6]). If a coder in a particular subband is temporarily overloading, some of this aliasing is not canceled and appears as "intermodulation distortion"; thus a careful trade must be made between the choice of the number of subbands, the order of the FIR filters, the number of bits used for quantization, and the resulting quantization and intermodulation distortion. By increasing the number of subbands a better performance due to the subband partitioning can be obtained provided that the length of the FIR filters is simultaneously increased (thus increasing the complexity and the amount of processing required of the hardware). The two-band approach presented here represents one compromise between hardware complexity and performance.

1.4 Design of the ADPCM Coders

The ADPCM coders use a single predictor loop and the robust step-size adaptation algorithm of Goodman and Wilkinson [9]. It is a modification of the algorithm of Jayant, Flanagan, and Cummisky [3]. Fig. 6 illustrates a block diagram of the coder for one of the subbands. The input subband signal to be coded is $y_i(n)$. Prior to encoding, a predicted version of $y_i(n)$ is determined in a feedback path (the differential parts of ADPCM) around the quantizer, and this value is subtracted from $y_i(n)$. The difference signal $e_i(n)$ is then quantized with a PCM quantizer with an adaptive step size $\Delta_i(n)$. Since this difference signal is generally smaller in amplitude than the signal $y_i(n)$, a smaller step size can be used for encoding, thus reducing the total quantization noise introduced by the quantizer [3]. The output of the quantizer is the bit stream $b_i(n)$ which is transmitted to the receiver. The quantized signal is also decoded to give the signal $\hat{e}_i(n)$, which is the quantized version of $e_i(n)$. This signal is the input signal to the predictor feedback loop. The signal $\hat{y}_i(n)$ is the decoded version of the signal $y_i(n)$, and it is used to form the predicted signal for the next time sample. The signal $\hat{y}_i(n)$ is delayed for one time sample and then multiplied by the coefficient α_i to form the new predicted signal. The value of the predictor parameter α_i will be discussed in more detail in Section 2.2.

In the receiver a modified version of this circuit is used for ADPCM decoding [3]. The received bit stream $b_i(n)$ is used to control the step-size adaptation, and it is decoded to produce the quantized difference signal $\hat{e}_i(n)$. A similar feedback circuit is used to obtain the decoded signal $\hat{y}_i(n)$, which is the output for the ADPCM decoder. Thus the decoder circuit is similar to the coder circuit, except that the quantizer becomes a decoder and the signals $y_i(n)$ and $e_i(n)$ are not present [3].

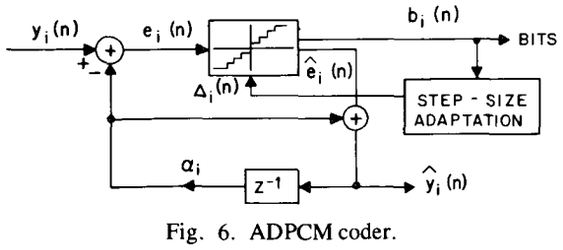


Fig. 6. ADPCM coder.

The robust step-size adaptation is based on the relation

$$\Delta(n) = [\Delta(n-1)]^{\beta} \cdot M(b(n-1)) \tag{8a}$$

where $\Delta(n)$ is the step size (that is, the spacing between quantizer levels) used for encoding (decoding) at the nth time sample and $\Delta(n-1)$ is the step size that was used for the $(n-1)$th time sample. The value of $\Delta(n-1)$ is raised to a power β, where $\beta \leq 1$ is a coder parameter (to be discussed in more detail later). It is then multiplied by a (positive) scale factor $M(\cdot)$ which is a function of the previous code word $b(n-1)$ to give the step-size estimate $\Delta(n)$. The choice of specific scale factor parameters $M(\cdot)$ as a function of the code word will be discussed in greater detail in Section 2.3. In addition the step size $\Delta(n)$ is restricted to the range

$$\Delta_{\min} \leq \Delta(n) \leq \Delta_{\max} \tag{8b}$$

where $\Delta_{\min}$ and $\Delta_{\max}$ are the minimum and maximum allowed values of $\Delta(n)$. The ratio $\Delta_{\max}/\Delta_{\min}$ is typically on the order of 1024, and thus provides a dynamic range on the order of 60 dB for the coders while maintaining $\Delta(n)$ within the range of the digital arithmetic.

In general if the previous codeword $b(n-1)$ indicates that an upper (positive or negative) quantizer level was used in encoding, a value of $M(\cdot) > 1$ is used to increase the size of the new step size $\Delta(n)$. If $b(n-1)$ indicates that a lower (positive or negative) amplitude level was used by the quantizer, a value of $M(\cdot) < 1$ is used to reduce the estimation of the new step size $\Delta(n)$. Thus the step-size adaptation algorithm is constantly attempting to adjust the step size $\Delta(n)$ such that it tracks the rms level of the signal and scales the quantizer characteristic to span the amplitude range of the signal.

The proportion of the amplitude range that is spanned by the quantizer at a particular time [that is, for a particular step size $\Delta(n)$] determines its "loading" [10]. If the range of the quantizer is too small relative to the signal range, the quantizer will overload and clip the signal. If it is too large, the quantizer step size will be too large, and this will result in an excessive quantization error or noise (often referred to as granular noise). Thus the proper "loading" of the quantizer is an important factor in maintaining a good reproduction of the signal. The loading is controlled by the choice of the parameters β and $M(\cdot)$. By taking the logarithm, Eq. (8a) can be written in the form

$$d(n) = \beta d(n-1) + m(b(n-1)) \tag{9a}$$

where

$$d(n) = \log \Delta(n) \tag{9b}$$

$$m(\cdot) = \log M(\cdot). \tag{9c}$$

The step-size adaptation is then implemented by the circuit shown in Fig. 7. The first table lookup converts values of $b(n-1)$ to $m(b(n-1))$ values, and the second table lookup implements the exponential conversion from $d(n)$ to $\Delta(n)$. Thus it is seen that the adaptation circuit consists of two table lookups and a first-order recursive digital filter which can easily be implemented in digital hard-

ware. Choices for the values of $M(\cdot)$ will be discussed in Section 3.

An extra dc input $(1 - \beta)d_c$ is also applied to the circuit in Fig. 7, and it is used to control the loading of the quantizer. Thus Eq. (9a), in practice, is modified to the form

$$d(n) = \beta d(n-1) + m(b(n-1) + (1-\beta)d_c \qquad (10)$$

where, in practice, d_c is chosen to have a value of approximately

$$d_c \approx \log(\Delta_{max}/10). \qquad (11)$$

The reasons for this modification will be discussed below.

In the remainder of this section we will discuss some theoretical aspects of the quantizer loading. Further information can be obtained from [3], [9], [10].

The adaptation leakage factor β, chosen to be $\beta = 127/128$, forces the realignment of the step sizes between transmitter and receiver after channel errors occur. Realignment will also occur when the step size reaches its maximum or minimum value according to Eq. (8b), even if β were chosen to be 1. However, for music signals several minutes may pass before this maximum or minimum is achieved. Since the cancellation of aliasing in the QMF bank depends strongly on the exact tracking of the step sizes in each subband, it is therefore preferable to use a value of $\beta < 1$ to dissipate any effects of channel errors more rapidly.

Another effect of the leakage due to β is that the log of the step size, that is, $d(n)$ in Fig. 7, tends to decay to zero in the absence of the inputs $m(b(n-1))$ and $(1-\beta)d_c$. By adding the term $(1-\beta)d_c$ the step size toward which $d(n)$ decays can be set to any arbitrary level. This can be used to help compensate for the effect of β when $\beta < 1$ on the loading of the quantizer.

The quantizer loading L is defined as

$$L = \sigma/\hat{\sigma} \qquad (12)$$

where σ is a statistical estimate of the rms input signal intensity and $\hat{\sigma}$ is the estimate of σ by the adaptation algorithm (which is proportional to the step size Δ) [10]. Thus L is a theoretical measure of how well the adaptation algorithm tracks the actual rms signal level ($L = 1$ is ideal). At some optimum signal intensity, denoted by σ_c, this loading is equal to 1, and the performance of the coder is maximized. For signal intensities above σ_c, L is greater than 1 (for $\beta < 1$), and the quantizer begins to overload. Similarly for $\sigma < \sigma_c$, L is less than 1, and the coder is dominated by granular noise. In practice σ_c is set to the normal operating level of the coder [10]. This dependence of L on σ can be expressed in the following form (derived below):

$$L = \left[\frac{\sigma}{\sigma_c}\right]^{(1-\beta)/(1-\beta+\gamma)} \qquad (13)$$

where γ is a constant related to the statistics of the signal and the multipliers $m(\cdot)$. Typical values of γ are on the order of 0.1.

The derivation of Eq. (13) can be obtained by examining the "steady-state" condition of Eq. (10) such that, "statistically," $\bar{d} = d(n) = d(n-1)$. The "statistical" value of $m(b(n)) = \bar{m}$ can then be derived as

$$\bar{m} = (1-\beta)(\bar{d} - d_c). \qquad (14)$$

As shown by Mitra [10], $\bar{m}$ is related to the loading factor L according to

$$\bar{m} = \gamma \log L. \qquad (15)$$

Under the above "steady-state" condition the estimate $\hat{\sigma}$ is

$$\hat{\sigma} = K \exp(\bar{d}) \qquad (16)$$

where K is a constant of proportionality. Substituting Eq. (16) into Eq. (12) and solving for $\bar{d}$ then gives

$$\bar{d} = \log\sigma - \log L - \log K. \qquad (17)$$

At the optimum input signal intensity level $\sigma = \sigma_c$, $L = 1$, $\bar{d} = d_c$, and Eq. (17) becomes

$$d_c = \log\sigma_c - \log K. \qquad (18)$$

Substituting Eqs. (17) and (18) into Eq. (14) and canceling terms gives

$$\bar{m} = (1-\beta)(\log\sigma - \log L - \log\sigma_c) \qquad (19)$$

and finally, equating Eqs. (15) and (19) and solving for L gives the desired expression for the quantizer loading in Eq. (13).

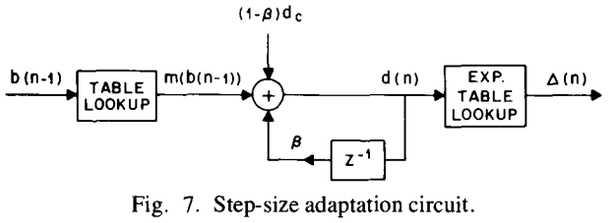


Fig. 7. Step-size adaptation circuit.

1.5 A/D and D/A Conversion

Another consideration in the design of the all-digital coder is that of analog-to-digital and digital-to-analog conversion preceding and following the coder. Conceptually it is possible to simply low-pass filter the input analog signal to 6.7 kHz and quantize it with a PCM quantizer. At the output the equivalent digital-to-analog conversion can be performed followed by a similar analog low-pass filter. The problem with this straightforward approach is that it requires sharp-cutoff large-order (20th-order elliptic) analog filters. These filters would be subject to the same accuracy, noise, and group-delay problems encountered in the analog coder discussed in the Introduction and would thus defeat the entire purpose of the all-digital coder design.

To avoid the above problems, the analog-to-digital and digital-to-analog conversion is performed in two stages as shown in Fig. 8. For digital-to-analog conversion the analog input signal $s(t)$ is prefiltered with a 5th-order LC elliptic filter with a frequency response as shown in Fig. 9(a). The edge of the passband is at 8.5 kHz. Since most of

the group-delay distortions, amplitude sensitivity, and ringing of the analog filter occur near this frequency, these distortions are outside the relevant signal band. The stop-band of the analog filter begins at 15.5 kHz, allowing a total transition region of 7kHz. Because of this broad transition region it is a low-Q filter. This, coupled with the fact that it is a passive LC filter, results in a design that does not suffer from noise or sensitivity problems. The dashed line in Fig. 9(a) shows the frequency response of the aliasing component after PCM conversion at 28 kHz. Since this aliasing is mostly in the 7–14-kHz region, it is later removed by the digital filter.

The digital filter is a 96-tap sharp-cutoff FIR filter with the frequency response shown in Fig. 9(b). It has a pass-band ripple of ±0.76 dB, a transition region from 6.7 to 7 kHz, and a 55-dB stop-band attenuation. Since the output of the filter is decimated by a factor of 2, only every other sample of the output needs to be computed. Also, since this filter is an all-digital FIR filter, it can be implemented with any degree of accuracy desired and therefore does not suffer from any of the drift, noise, or group-delay problems encountered in analog filters.

The analog-to-digital conversion in Fig. 8(a) is achieved by a two-stage process according to Fig. 10. The input

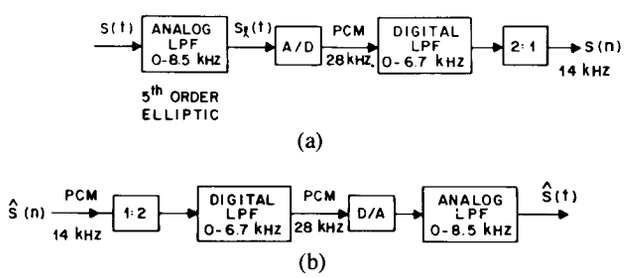


Fig. 8. (a) A/D conversion circuit. (b) D/A conversion circuit.

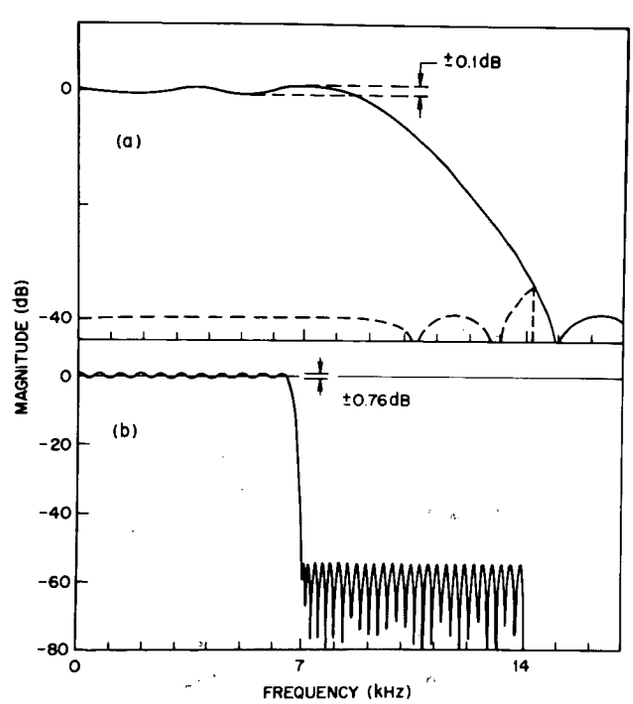


Fig. 9. (a) 5th-order analog LC elliptic filter. (b) 96th order digital filter for A/D and D/A conversion circuits.

signal $s_1(t)$ is stored in a sample and hold and presented to a 12-bit analog-to-digital converter through a gain, $G = 1$. The output of the analog-to-digital converter is stored in the first 12 bits of a 16-bit shift register with the remaining 4 bits zeroed out. If the first two most significant bits (MSB) are zero, the gain selector is switched to a gain $G = 4$ and the analog-to-digital conversion is repeated. To compensate for this gain, the bits in the 16-bit shift register are shifted 2 bits to the right. If the first 4 bits are found to be zero, a gain of $G = 16$ and a shift of 4 bits is used. In this way the analog-to-digital conversion has a 12-bit accuracy and a 16-bit dynamic range.

An extra benefit of the above design is that the digital filter removes one-half of the quantization noise introduced by the analog-to-digital conversion, thus enhancing the signal-to-noise ratio by an additional 3 dB. The output of the analog-to-digital conversion is therefore a "clean" PCM signal with effectively 12.5 bits of accuracy and 16 bits of dynamic range. A similar high-quality digital-to-analog conversion is achieved with the reverse set of operations illustrated in Fig. 8(b).

1.6 Overall System Considerations

The commentary quality coder is designed for a transmission rate of 56 kb/s to be compatible with a 56-kb/s Dataphone digital service (DDS) port. This bit rate is achieved by the use of 4-bit ADPCM encoding and a 7-kHz sampling rate in each subband.

If the coder is used at the 64-kb/s T1 interface, a 5-bit ADPCM coding may be used in the lower subband for an even better overall quality. This results in a 63-kb/s coder design.

2 CODER PERFORMANCE FOR SPEECH AND MUSIC

In the previous section the details of the design of the all-digital coder have been explained. However, the values of the parameters which control the characteristics of prediction and step-size adaptation in the coder were left unspecified. These parameters must be carefully tailored to the characteristics of the signals that are to be coded. In this section we consider the issues involved in properly selecting these parameters.

It is generally found that the most critical signals to encode are those of music. A coder designed for good performance for music sounds will generally have a satisfactory performance for speech.

The statistics of a number of musical sounds are first

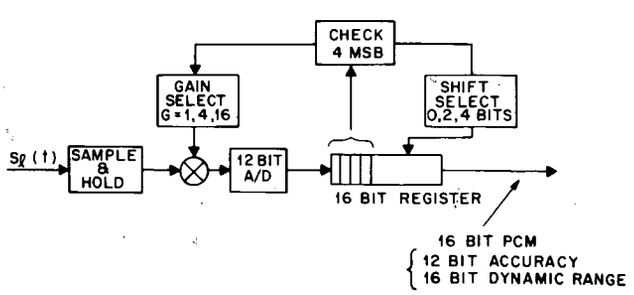


Fig. 10. Two-step A/D converter circuit.

examined. The coder parameters are then chosen to provide a "commentary grade" performance over a wide range of instrumental and vocal music without sacrificing any group of musical sounds. The performance of the coder is then analyzed with these parameters.

2.1 Characterization of Music and Vocal Sounds

There are two important items that relate to the coder when considering the characterization of music signals. They are (1) the frequency spectra of the signal (long term and short term) and (2) the amplitude behavior of the signal.

Music, unlike speech, is characterized by many different individual spectra [11]. Instead of a small number of resonances (formants) as in speech, music sounds may have highly complicated spectra, over a short term, which are constructed from many fundamentals and their harmonically (and in some cases nonharmonically) related tones. These fundamentals may be as low as 16 Hz (pipe organ 32′ stops) or as high as 10–12 kHz (triangles). Thus in music one seeks to find a characterization of sound that allows a predictor that works well on all sounds without sacrificing one or more particular instruments.

In Fig. 11 we present typical waveforms and their spectra for a number of musical sounds: vocal, full orchestra, guitar (soft), guitar (loud, popular), flute, and trumpet. (Also, see Table 2 for the source of these sounds).

The vocal and full orchestra sounds are characterized by very complicated time waveforms and spectra. Their spectra drop by about 40 dB across the total (0–7-kHz) band. Across any narrow frequency range, however, the dynamic range is only on the order of 20 dB for the full orchestra and 30 dB for vocal sounds. The (soft) guitar has a rapid rolloff in energy across the spectrum (about 60 dB) with a localized spectral dynamic range of about 30 dB. The (loud) guitar has more high-frequency energy than any of the other signals and a local spectral dynamic range of nearly 40 dB. The flute has a highly resonant spectrum with only 2 or 4 major harmonics of extremely high Q (over 50-dB peaks). The trumpet also has high-Q (40-dB) resonances. It has more high-frequency harmonics, however, and a substantially different time waveform.

The attack characteristics of the waveforms, as well as the spectra of each instrument, contribute significantly to the perceptual information. In our examples the vocal, orchestral, and flute sounds generally exhibit an attack that is smooth and not overly rapid. The guitar, by nature, exhibits a very rapid attack which the ADPCM coder must follow in order to preserve the character of the instrument. Trumpets (and other brass instruments) exhibit an attack that can vary according to the wishes of the performer, from slow to fairly rapid.

This group of sounds and instruments, while by no means exhaustive, has been selected to present an array of some of the more difficult conditions.

2.2 Selection of Predictor Coefficients

The choice of the predictor coefficients α_1 (lower band) and α_2 (upper band) controls the values of the predicted signal that are subtracted from the input signal $y_i(n)$ ($i = 1$, 2, see Fig. 6). If the values of these coefficients are equal to the sample-to-sample correlations which are present in the signals to be coded, then the signal-to-noise ratios S/N are maximized [3]. In coding speech signals, maximizing S/N usually increases the perceived quality; thus the prediction coefficient is set to the average sample-to-sample correlation value for speech. Coders for music signals, however, experience a much greater range of signals, often with greatly different values of sample-to-sample correlations.

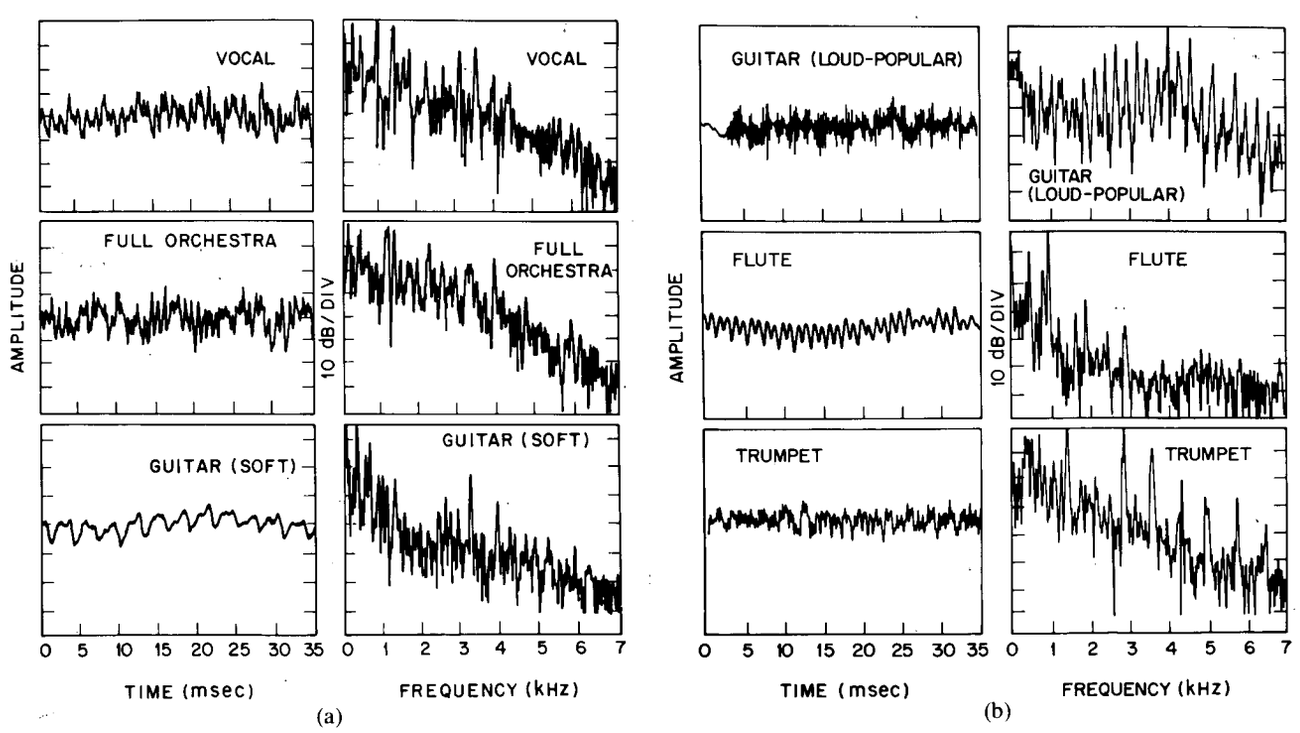


Fig. 11. Typical waveforms and spectra for different musical sounds.

This variation in sample-to-sample correlation must be considered and the objective of maximizing S/N reconsidered.

To determine appropriate choices for α_1 and α_2 the sample-to-sample correlation was measured for the upper and lower subband signals for a number of different musical sounds. Table 2 shows the sample-to-sample correlations observed for typical passages of the sounds in the upper and lower subbands. The fact that the correlations for the upper band are negative is a consequence of the frequency inversion of that band due to the modulation and decimation in the quadrature mirror filter.

It can be observed that the correlations in the upper band are fairly consistent, with an average that runs from -0.69 to -0.85. This consistency allows the use of a predictor coefficient $\alpha_2 = -0.82$, which gives good results for most sounds.

The lower band, however, shows no consistency in the sample-to-sample correlation and it has a range of -0.24 (trumpet) to $+0.98$ (guitar). Because of this large diversity, a fixed predictor value cannot be selected which is best for all sounds. The use of the average value (about 0.5) does not lead to good performance, especially for signals that have lower correlations. In informal listening it was observed that the use of high predictor values ($\alpha_1 = 0.6$) on signals with relatively low sample-to-sample correlations caused more objectionable degradations than the use of low predictor values ($\alpha_1 = 0.2$) on signals with high sample-to-sample correlation. An examination of the distortion introduced by the use of a high predictor value on a relatively uncorrelated signal showed that the coder was often driven into slope overload. This caused intermodulation distortion which appeared as lines in the error spectrum instead of the typically flat noise spectrum. These nonharmonic tones have an undesirable musical content, whereas the flat quantization noise is simply perceived as a low-level noise with no musical content. A reasonable compromise for the fixed predictor value for the lower subband, based on informal listening, was found to be $\alpha_1 = 0.16$.

Table 2. Range of Sample-to-Sample Correlations* for music sounds.†

Sound	Subband 1 Min	Subband 1 Max	Subband 2 Max	Subband 2 Min
Vocal	0.54	0.93	−0.40	−0.90
Full orchestra	0.42	0.77	−0.77	−0.89
Guitar (soft)	0.74	0.98	−0.61	−0.89
Guitar (loud, popular)	0.00	0.71	−0.59	−0.93
Flute	0.67	0.95	−0.41	−0.81
Trumpet	−0.24	0.51	−0.61	−0.90

* Correlation is measured over 256 sample blocks according to:

$$c = \frac{\sum_{n=1}^{255} y(n)y(n-1)}{\sum_{n=1}^{255} y^2(n)}.$$

† The sources of the music sounds are:

Vocal — Starland Vocal Band, "American Tune," Windsong BHL 1-13515.

Full orchestra and trumpet — Eugene Ormandy, Philadelphia Orchestra, March Album, "Trumpet Voluntary" by Jeremiah Clarke, Masterworks MG32314.

Flute — Oslo Philharmonic/Odd Gruner-Hegge Conductor, "Peer Gynt Suite—Morning," by Edvard Grieg, RCA Victrola VICS 1067S.

Guitar (soft and loud, popular) — Paul Simon, "There Goes Rhymin' Simon," "Take Me to the Mardi Gras," CBS/Columbia KC32280.

2.3 Selection of Step-Size Adaptation Parameters

The dynamic amplitude behavior of each subband was examined to obtain an indication of the appropriate values for $M(\cdot)$ in the step-size adaptation of the ADPCM coders. Table 3 shows the values of $M(\cdot)$ used in initial experiments to observe the behavior of the coders. An examination of the error signals, and informal listening observations, indicated that the upper band was adapting satisfactorily. The lower band, however, showed signs of overloading in the attack portions of the signal. The highest multiplier was therefore modified, as shown in Table 3, to prevent this distortion during the attack. This resulted in a coder performance that had a 0.5-dB lower S/N but a better perceived performance due to the elimination of the overloading. (A careful balance between the selection of predictor coefficients and adaptation coefficients is necessary to completely control slope overload in the attack regions.)

In the 63-kb/s coder design the selection of $M(\cdot)$ values in the lower band was modified to give a coder with a 5-dB better S/N performance than the 56-kb/s coder. This selection, also shown in Table 3, shows an increased overhead for attack and virtually eliminates any problems of overloading.

2.4 S/N Performance of the Coder

Fig. 12 illustrates plots of the rms signal energy and S/N as a function of time for the various concatonated musical sounds. It is seen that the signal intensity between the different musical sounds can vary by as much as 18 dB. The 56-kb/s coder maintained a relatively constant S/N of about 19.2 dB across all of these sounds. The 63-kb/s coder shows a similar S/N response, except that it is increased by

Table 3. $M(i)$ Values for step-size adaptation.

Quantizer Step i	Upper Subband	Lower Subband Initial Values (4 bits)	Lower Subband Modified Values (4 bits)	Lower Subband Modified Values (5 bits)
1	0.9	0.9	0.9	0.9
2	0.9	0.9	0.9	0.9
3	0.9	0.9	0.9	0.9
4	0.9	0.9	0.9	0.9
5	1.2	1.2	1.2	0.9
6	1.6	1.6	1.6	0.9
7	2.0	2.0	2.0	1
8	2.4	2.4	3.0	1
9				1.2
10				1.2
11				1.6
12				1.6
13				2
14				2
15				3
16				3.75

5 dB.

In Fig. 13 the rms signal intensities and S/N are plotted separately for the upper and lower subbands. It is seen that the lower subband generally has a larger signal intensity than the upper subband and that the proportion of energy between these two bands varies with different musical sounds. The S/N for the lower subband fluctuates around a value of 19.2 dB (for the 56-kb/s coder). In the upper subband the S/N fluctuates around a value of 21.9 dB with the exception of the flute, which has virtually no signal energy in the upper band.

Fig. 14 illustrates a typical example of the spectrum of the quantization noise in the coder for the (soft) guitar sound, and it is also representative of the noise spectra for other musical sounds. As seen in the figure, the noise is about 20 dB greater in the lower band, which conforms with the contour of the spectrum of the original sound (see Fig. 11). Thus the larger signal energy in the lower band effectively masks the larger noise energy in this band.

Another phenomenon that was observed in the performance of the coder was the effect of very-low-frequency noise (such as turntable rumble) on the step-size adaptation in the lower band. This rumble, for example, can be seen as the very-low-frequency components in the time waveforms of the flute and the (soft) guitar in Fig. 11. Because of the insensitivity of the ear (and most sound-reproduction equipment) at this low frequency, this rumble is not highly audible. It has the effect, however, of pushing up the step size in the lower band and thus increasing the quantizing noise level in that band. This phenomenon can be eliminated in the coder by appropriately high-pass filtering the input signal (at about 20 Hz) prior to coding.

3 CONCLUSIONS

As mentioned in the beginning of the paper, the intention was to design a coder that has a sufficient quality for AM radio use. We feel that this objective can be achieved with the 56-kb/s coder and that the 63-kb/s coder represents a quality which exceeds that of most AM radio signals and nearly all AM receivers. The bit rate of the coder is the same as that of an ordinary μ-law coder which allows the use of this coder in a normal slot of a T1 system. The 56-kb/s rate is acceptable to the DDS network and can be used at any place where a DDS port is available. It would be possible for stations to replace existing leased lines, commonly used for remote music and voice, with a DDS port, thus eliminating the analog transmission part of the system. This would provide a greater measure of robustness along with an improved dynamic range and frequency response.

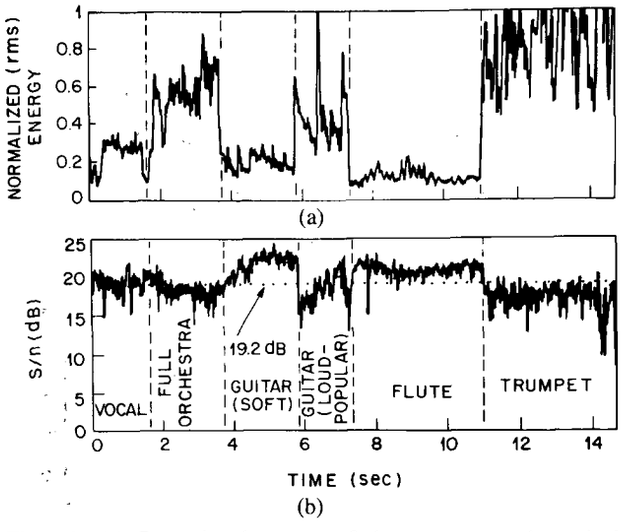


Fig. 12. (a) Rms signal energy of the concatonated musical sounds. (b) S/N performance of the 56-kb/s coder as function of time.

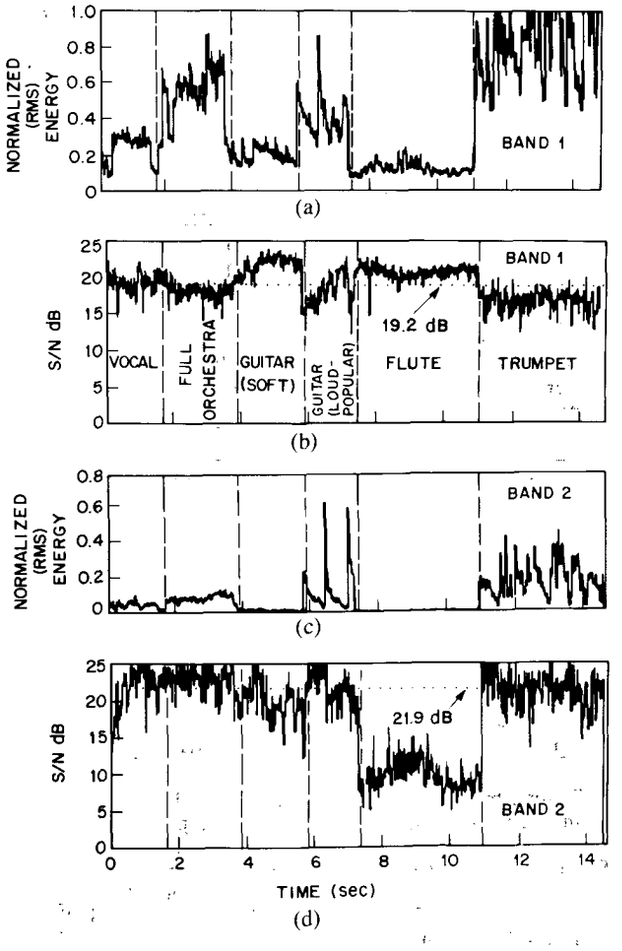


Fig. 13. (a) Rms energy. (b) S/N in the lower subband. (c) Rms energy. (d) S/N in the upper subband.

4 ACKNOWLEDGMENT

The authors would like to thank F. R. Moore for his comments and discussion in the preparation of this manuscript.

5 REFERENCES

[1] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Sub-Bands," *Bell Sys. Tech. J.*, vol. 55, pp. 1069–1085 (1976 Oct.).

[2] R. E. Crochiere, "On the Design of Sub-Band Coders for Low-Bit-Rate Speech Communication," *Bell Sys. Tech. J.*, vol. 56, pp. 747–770 (1977 May-June).

[3] N. S. Jayant, J. L. Flanagan, and P. Cummisky, "Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantizers," *Proc. IEEE*, vol. 62, pp. 611–632 (1974 May).

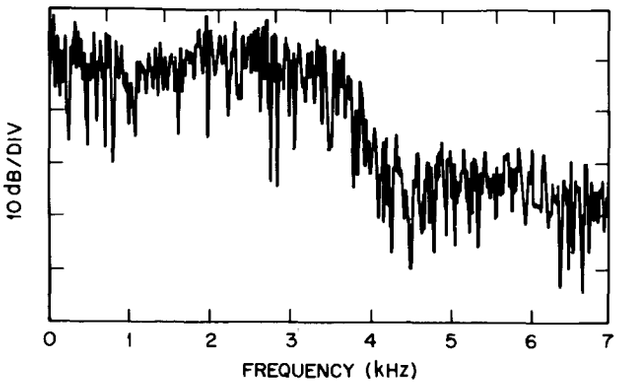


Fig. 14. Typical spectrum of the quantization noise in the 56-kb/s coder.

[4] Members of Technical Staff of Bell Telephone Laboratories, *Transmission Systems for Communications* (Bell Telephone Laboratories, Inc., Murray Hill, NJ, 1970).

[5] J. D. Johnston and D. J. Goodman, "Digital Transmission of Commentary—Grade (7 kHz) Audio at 56 or 65 kb/s," *Proc. 1979 IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, (Washington, DC, April 2–4), pp. 442–444.

[6] D. Esteban and C. Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes," *Proc. 1977 IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.* (Hartford, CT, May), pp. 191–195.

[7] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

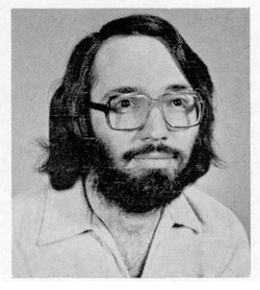
[8] J. H. McClellan, T. W. Parks, and L. R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 506–526 (1973 Dec.).

[9] D. J. Goodman and R. M. Wilkinson, "A Robust Adaptive Quantizer," *IEEE Trans. Commun.*, vol. COM-23, pp. 1362–1365 (1975 Nov.).

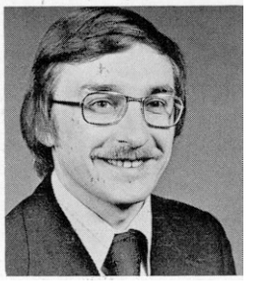
[10] D. Mitra, "An Almost Linear Relationship Between the Step Size Behavior and the Input Signal Intensity in Robust Adaptive Quantization," *IEEE Trans. Commun.*, vol. COM-27, pp. 623–629 (1979 Mar.).

[11] A. H. Benade, *Fundamentals of Musical Acoustics* (Oxford Press, New York, 1976).

THE AUTHORS



J. D. Johnston



R. E. Crochiere

James D. Johnston was born in 1953, in Niles, Ohio. He received his B.S.E.E. and M.S.E.E. degrees from Carnegie-Mellon University in 1975 and 1976, respectively.

Since then, he has been employed by Bell Telephone Laboratories, and is currently an associate member of technical staff in the Acoustics Research Department in Murray Hill, NJ. His research interests include wide and narrow bandwidth waveform coding techniques, fast small-scale digital processors, and analog-to-digital and digital-to-analog techniques.

He is a member of the Audio Engineering Society, and has published papers in IEEE Communications Transactions, ICC and NTC Records, and Electronics Letters. He has an active interest in theater and musical sound reinforcement.

•

Ronald E. Crochiere was born in Wausau, Wisconsin, in 1945. He received the B.S. degree in 1967 from the Milwaukee School of Engineering, Milwaukee, Wisconsin and the M.S. and Ph.D. degrees in 1968 and 1974, respectively, from the Massachusetts Institute of Technology, Cambridge, Massachusetts, all in electrical engineering. From 1968 to 1970 he was employed with Raytheon Co. In 1970 he returned to M.I.T. and joined the Research Laboratory of Electronics where he was engaged in graduate studies for the Ph.D. degree.

In 1974, Dr. Crochiere joined the Acoustics Research Department of Bell Laboratories where he has been involved in research activities in concepts of decimation and interpolation, subband and transform coding of speech, and the measurement of digital speech quality. In 1976 he received the IEEE ASSP paper award for his paper on decimation and interpolation of digital signals.

Dr. Crochiere is an active member of the ASSP ADCOM committee and the ASSP Technical Committee on Digital Signal Processing. He has served for two years as a technical editor on digital signal processing for the ASSP transactions and is presently secretary-treasurer for the ASSP society ADCOM.