THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 5, May–June 1977
Printed in U.S.A.

On the Design of Sub-band Coders for Low-Bit-Rate Speech Communication

By R. E. CROCHIERE

(Manuscript received November 5, 1977)

This paper presents a detailed discussion of issues involved in the design of sub-band coders for low-bit-rate speech communications. Specifically, bit rates in the range of 7.2 to 16 kb/s are emphasized. Design guidelines, based on results of extensive computer simulations and subjective comparisons, are presented for selection of sub-band coder parameters. Practical considerations for selecting sub-bands under integer-band sampling and multiplexing constraints are also discussed, and a method for synchronous multiplexing of the sub-band data, without buffering, is proposed. Several examples of sub-band coders for transmission rates of 7.2, 9.6, and 16 kb/s are presented, and the quality of these coders is compared against that of ADPCM and ADM coders.

I. INTRODUCTION

In recent work by Crochiere, Webber and Flanagan,[1] an approach to speech encoding has been proposed which is based on the partitioning of the speech band into sub-bands and encoding the sub-bands individually. The technique offers attractive possibilities for coding speech economically at bit rates in the range of 7.2 to 16 kb/s. At 16 kb/s good quality encoding, comparable to that of 26.5 kb/s adaptive differential (fixed predictor) PCM (ADPCM) encoding, is possible. Potential applications exist in areas of narrow-band communications, mobile radio, and voice storage applications.

When the bit rate is extended down into the upper data rate range of 9.6 and 7.2 kb/s, moderate quality encoding can be achieved comparable to that of 19 and 18 kb/s adaptive delta modulation (ADM), respectively. Interesting potential applications exist for voice coordination on digital data lines and for secure voice communications by digital encryption and transmission over conventional data lines.

In the design of sub-band coders, a variety of issues and "trade-offs" must be dealt with. The number of sub-bands, the partitioning of sub-

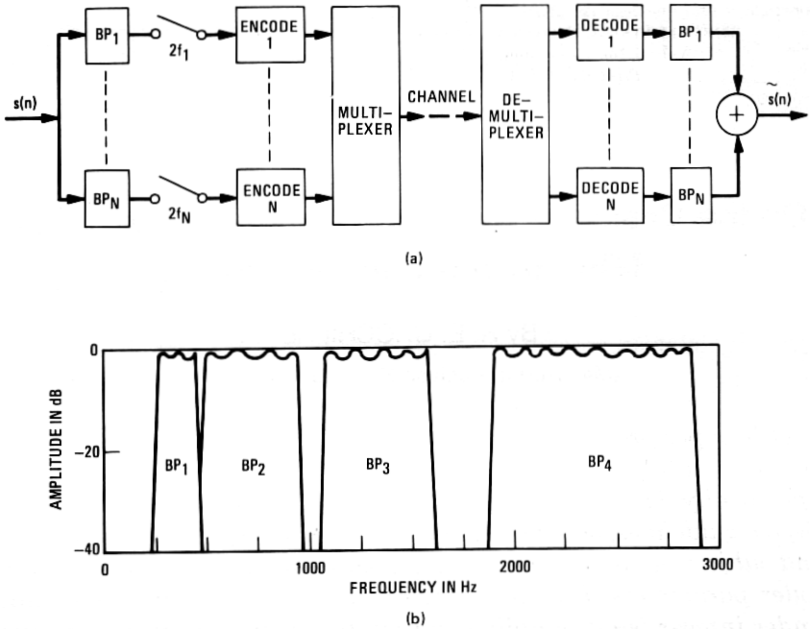

Fig. 1—(a) Implementation of a sub-coder based on integer-band sampling. (b) Frequency-domain illustration of the sub-band partitioning of the speech band.

bands (and gaps between bands), coder parameters, parceling of bits among sub-bands, and compromises between bits/sample and bandwidth are all variables that must be considered. In addition, a number of constraints are introduced by practical considerations of multiplexing the digitized sub-band signals and by considerations of efficient hardware implementation. In this paper, we attempt to clarify these issues and present useful criteria and guidelines for designing sub-band coders. In many respects, the only truly meaningful criterion for selecting parameters of the sub-band coder is a perceptual one. Therefore, design criteria have been supported, as much as possible, by results of extensive computer simulations and listener preference tests.

II. A REVIEW OF SUB-BAND CODERS

In the sub-band coder, the speech band is partitioned into sub-bands by bandpass filters. Each sub-band is low-pass translated, sampled at its Nyquist rate, and digitally encoded. By this process of dividing the speech band into sub-bands, each sub-band can be preferentially encoded according to perceptual criteria for that band. On reconstruction, sub-band signals are decoded and bandpass translated back to their original bands. They are then summed to give a replica of the original speech signal.

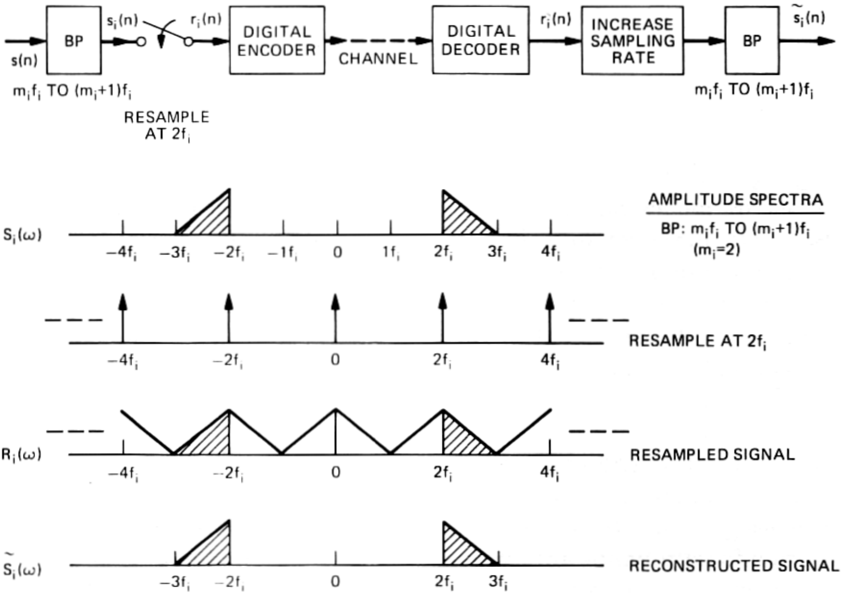

Fig. 2—Integer-band sampling technique and a frequency-domain interpretation.

A variety of techniques exists for performing the low-pass and bandpass translations. However, one approach is particularly attractive for hardware implementation since it eliminates the need for modulators. It is based on the integer-band sampling method proposed in Ref. 1 and will be the method primarily considered in this paper.

The integer-band sampling implementation of the sub-band coder is illustrated in Figs. 1 and 2. The speech band is partitioned into N sub-bands by bandpass filters BP_1 to BP_N. It will be assumed in this paper that the filters are discrete-time (e.g., digital or CCD) filters. Typically four or five bands are used and, at lower bit rates, small gaps are permitted between bands to conserve bandwidth and bit rate, as illustrated in Fig. 1b.

The output of each filter in the transmitter is resampled at a rate of $2f_i$, where f_i is the width of the sub-band and i refers to the ith sub-band. The sampled sub-band signals are digitally encoded and time multiplexed for transmission over the digital channel. At the receiver the digital signals are demultiplexed and decoded. The sub-band signals are reconstructed by filtering the outputs of the decoders with another set of bandpass filters, identical to BP_1 to BP_N, that act as interpolating filters. Prior to this filtering, the sampling rates of the decoder outputs are increased to the original sampling rate of $s(n)$ by filling in with zero-valued samples. The outputs of these filters are then summed to give a reconstructed replica $\hat{s}(n)$ of the original speech signal $s(n)$.

The integer-band sampling scheme imposes certain constraints on the choice of sub-bands, as illustrated in Fig. 2. Sub-bands are required to be between $m_i f_i$ and $(m_i + 1)f_i$, where m_i is an integer. This constraint is necessary to avoid aliasing in the sampling process.

Encoding in sub-bands offers several advantages over full-band coding.[1] Quantization noise can be contained in bands to prevent masking of one frequency range by quantizing noise in another frequency range. Separate quantizer step-sizes are used in each band. Therefore, bands with lower signal energy will have lower quantizer step-sizes and contribute less quantization noise. Finally, the partitioning of the speech band into sub-bands enables the parceling of bits in bands according to perceptual criteria. In lower bands where pitch and formant structure must be accurately preserved, a larger number of bits/sample can be used for encoding, whereas in upper bands where fricatives and noise-like sounds occur in speech, fewer bits/sample can be used.

In the following sections, we focus on the various issues involved in the design of sub-band coders. Section III addresses issues of coder selection for sub-bands and the choice of their parameters. "Trade-offs" involved in the allocation of bits among bands are also discussed. Section IV deals with problems of sub-band partitioning of the speech spectrum under the constraints of integer-band sampling requirements and multiplexing requirements. Section V involves issues in the design of filters for the sub-bands. Finally, Section VI presents further results on comparisons of sub-band coder performance with other waveform coding methods.

III. SELECTION OF CODERS AND CODER PARAMETERS FOR SUB-BANDS

Because encoders are individually tailored to each sub-band, a spectrum of coders and parameters must be considered. For the lower-frequency sub-bands, typically 3 or 4 bits/sample encoders are used, and for upper bands 2 or less bits/sample are used. Since the characteristics of the sub-band signals are considerably different from those of full-band speech, encoding techniques developed for encoding of full-band speech signals do not necessarily lead to good results for encoding of sub-band signals. In this section, we therefore address issues in the design of encoders for sub-band signals and in the parceling of bits among bands.

The choice of encoder parameters is determined in part by the static or long-term spectral characteristics of the speech waveform. Figure 3a illustrates typical long-term speech spectra (averaged over a sentence) based on measurements made by Beranek[2] and Dunn and White.[3] The same spectra are plotted in Fig. 3b with a warped frequency scale based on a constant (5 percent/division) contribution to the articulation index[2] in order to illustrate the relative perceptual importance of the various

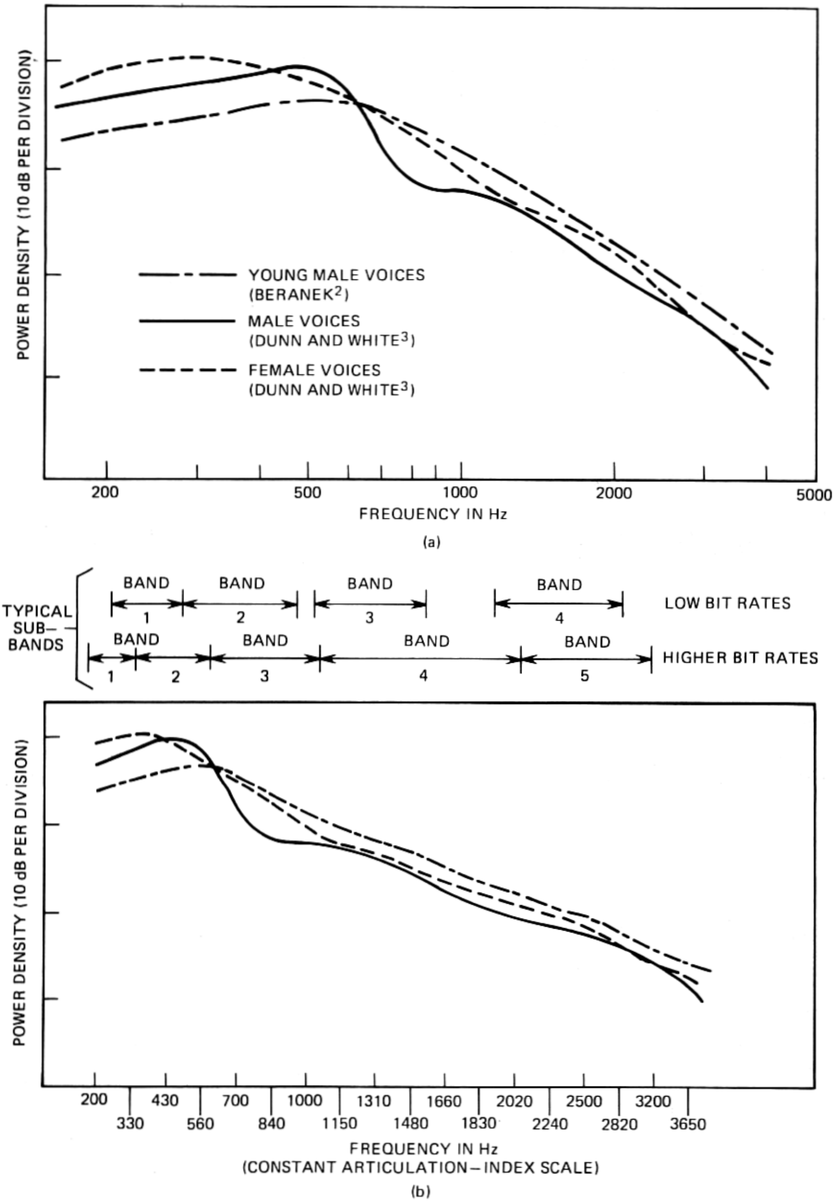

Fig. 3—Long-term spectrum of speech based on measurements by Beranek[2] and Dunn and White.[3] (a) Logarithmic frequency scale. (b) Frequency scale based on a constant contribution to the articulation index.

frequencies. Two possibilities for sub-band selection for low and high bit rates (to be discussed later) are illustrated above Fig. 3b. It is seen that across the entire speech spectrum there is a characteristic drop in

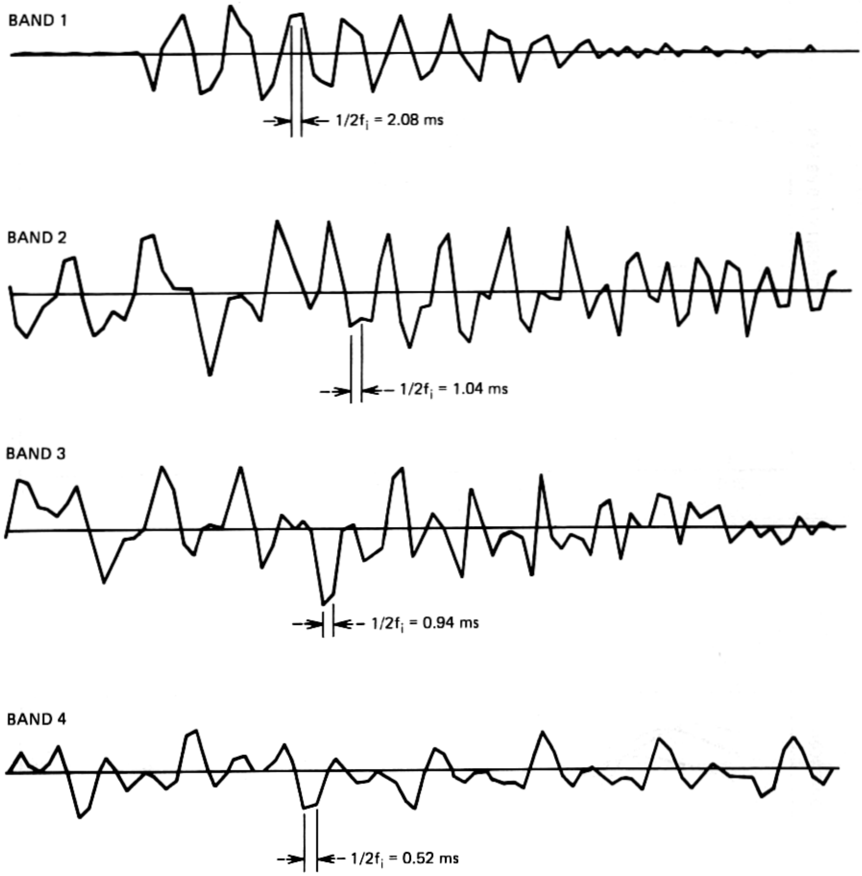

Fig. 4—Typical waveforms of uncoded sub-band signals for bands 1 to 4. Eighty samples are plotted on each line.

power density with increasing frequency. Across any one band, however, the drop in power density is relatively small. Since sub-bands are, in effect, low-pass translated and sampled at their Nyquist rate, they appear essentially as flat spectrum signals at the low sub-band sampling rates and have essentially no sample-to-sample correlation. Figure 4 shows examples of sub-band signals for bands 1 to 4. Because of their low sample-to-sample correlation, encoding is best performed by adaptive PCM (APCM). Encoding based on differential or fixed prediction, commonly used for full-band encoding, does not lead to good results for encoding of sub-band signals.

The step-size adaption strategy used in simulations for the APCM coders is based on the one-word step-size memory approach proposed by Jayant, Flanagan, and Cummiskey.[4,5] The coder input signal, denoted

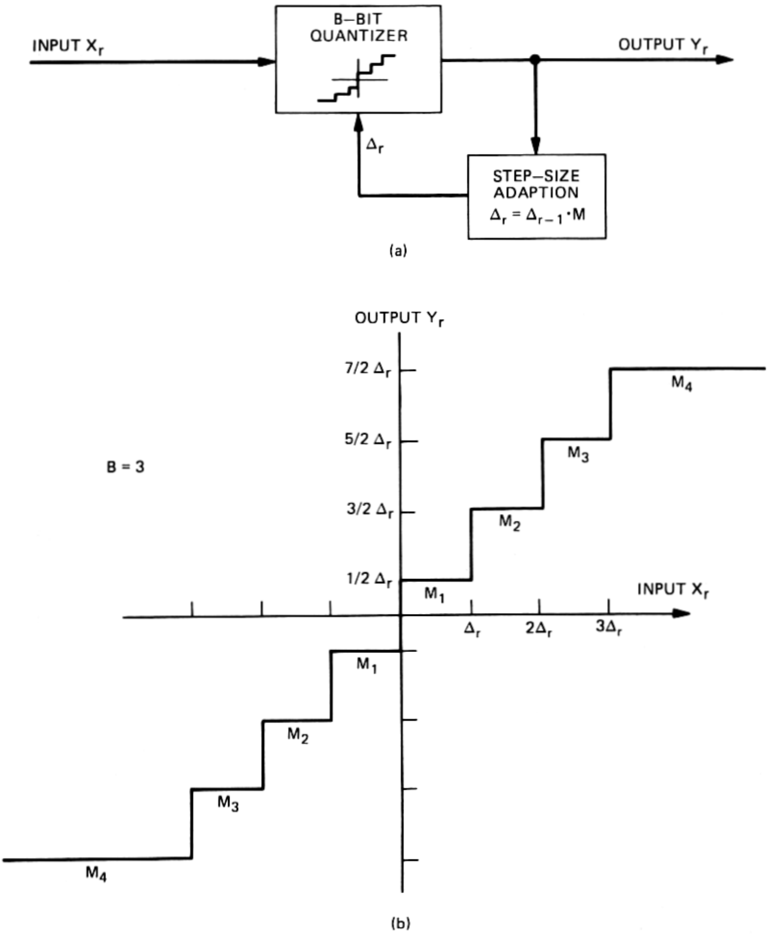

Fig. 5—Step-size adaption algorithm and quantizer characteristics of the APCM coders.

as x_r for the rth sample, is quantized to one of 2^B levels according to the quantizer characteristics shown in Fig. 5, where B is the number of bits in the coder. The step-size adaption circuit examines the quantizer output bits for the $(r-1)$th sample and computes the quantizer step-size, Δ_r, for the rth sample according to the relation

$$\Delta_r = \Delta_{r-1} M(L_{r-1}), \tag{1a}$$

where

$$\Delta_{\mathrm{MIN}} \leqq \Delta_r \leqq \Delta_{\mathrm{MAX}}, \tag{1b}$$

and where Δ_{r-1} is the step size used for the $(r-1)$th sample. $M(L_{r-1})$

Table I — APCM coder parameters

B =	4	3	2	$1\frac{1}{2}$	$1\frac{1}{3}$	$1\frac{1}{4}$
M_1	0.9	0.85	0.85	0.92	0.92	0.92
M_2	0.9	1.0	1.9	1.4	1.4	1.4
M_3	0.9	1.0				
M_4	0.9	1.5				
M_5	1.2					
M_6	1.6					
M_7	2.0					
M_8	2.4					
Typical s/n (dB)	18	11.5	7	4	3.3	2.5

is a multiplication factor whose value depends on the quantizer magnitude level L_{r-1} at time $r - 1$. It can take on one of 2^{B-1} values, $M_1, M_2, \cdots M_{2^{B-1}}$. If the lower-magnitude quantizer levels are used at time $r - 1$, a value of $M(L_{r-1}) = M_i$ less than one is used to reduce the next step size. If upper-magnitude levels are encountered, a value of M_i greater than one is chosen. In this way, the coder continuously adapts its step size in an attempt to track the short-time variance of the input signal. For practical reasons, the step size, Δ_r, is constrained to be between some minimum and maximum value Δ_{MIN} and Δ_{MAX}, respectively.

Typical values of M_i for 2-, 3-, and 4-bit APCM coders are given in Table I. These values were determined experimentally and were found to agree reasonably well with values reported by Jayant[4] for encoding of full-band speech. As observed by Jayant, small changes in these values do not strongly affect the performance of the coders. Typical signal-to-quantizing noise ratios (s/n) found for encoding sub-band signals are also reported in Table I.

An interesting modification to the above algorithm, proposed by Goodman,[6] allows for encoding at an average bit rate of $1 + 1/K$ bits/sample, where K is an integer. In this approach, the sign of the signal x_r is encoded for each sample, r, and the magnitude of the signal is encoded with one bit every K samples. The step-size adaption is essentially that of (1) with M_i and the quantizer magnitude level repeated for $K - 1$ samples at the decoder. For example, if $K = 2$, a sign and a magnitude bit are transmitted on odd numbered samples. On even numbered samples, only the sign bit is transmitted and the magnitude bit is assumed to be that of the previous sample. The sign bit transmits essentially the "zero crossing" or phase information and the magnitude bit conveys the amplitude information in the waveform at a reduced rate.

The $1 + 1/K$ bit coder is found to be useful for encoding the uppermost bands when overall bit rates must be kept low. The upper bands contain primarily the fricative and noise-like sounds in the speech and can therefore be quantized more coarsely than lower bands without a perceived loss in quality. Typical adaption parameters found to be useful for $1 + 1/K$ bit coders are also given in Table I.

The quantities Δ_{MAX} and Δ_{MIN} in the above algorithms represent practical constraints in the adaption logic. Their ratio determines the dynamic range that the coder can handle and their absolute values determine the center of this dynamic range. In simulations, a ratio $\Delta_{MAX}/\Delta_{MIN} = 128$ was consistently used, resulting in a useful dynamic range of about 40 dB for the coders. The actual values of Δ_{MIN} and Δ_{MAX} must be different for each sub-band, however, to match properly the dynamic range characteristics of the sub-band coder to that of the long-term speech spectrum. This is easily seen in Fig. 3. Since upper sub-bands have lower power densities than lower sub-bands, they should have smaller values of Δ_{MAX} and Δ_{MIN} in their coders. A useful criterion for choosing relative values of $\Delta_{MIN}(\Delta_{MAX} = 128\Delta_{MIN})$ can be derived by assuming that the power-density spectrum in sub-band i is approximately flat across the band and has a value S_i. The long-term variance, $\sigma_i{}^2$, of the sub-band signal is then proportional to $S_i f_i$.

To match the center of the dynamic range of the coders in each band, Δ_{MIN} should be selected to be proportional to the square root of the long-term variance of the signal in that band. Therefore, the ratio of $\Delta_{MIN}(\text{band } i)$ in band i to $\Delta_{MIN}(\text{band } j)$ in band j can be determined as

$$\frac{\Delta_{MIN}(\text{band } i)}{\Delta_{MIN}(\text{band } j)} \cong \frac{\sigma_i}{\sigma_j} \cong \sqrt{\frac{S_i f_i}{S_j f_j}} \tag{2}$$

or if values are expressed in dB, (2) becomes

$$\left.\frac{\Delta_{MIN}(\text{band } i)}{\Delta_{MIN}(\text{band } j)}\right|_{\text{dB}} \cong S_i|_{\text{dB}} - S_j|_{\text{dB}} + 20 \log \sqrt{\frac{f_i}{f_j}}. \tag{3}$$

Equation (3) states that the ratio of minimum step size (in dB) of band i to band j is equal to the difference in power densities (in dB) between band i and band j plus a correction factor to account for the differences in bandwidths. Values of S_i and S_j can be obtained from Fig. 3. Although eq. (3) is only approximate, it serves as a useful criterion for choosing relative values of Δ_{MIN} for coders. Good agreement was found with experimentally derived values.

A final consideration in the selection of coders for sub-bands relates to the questions of how many bits/sample should be allocated to each sub-band under constraints of fixed total transmission rate and how should the sub-band bandwidths and gaps between bands be traded against bits/sample for the coders. The answer to both questions is highly dependent on perceptual criteria and is greatly influenced by the overall allowed transmission rate. Therefore, we do not propose to answer these questions in detail but simply provide some insight.

A useful measure for assisting in the parceling of bits among sub-bands is the signal-to-quantizing noise ratio (s/n) as a function of frequency. Figure 6 shows typical s/n values as a function of frequency that are found to give preferred signal quality at bit rates of 16, 9.6, and 7.2 kb/s, respectively. At 16 kb/s it is found that good quality coding can be achieved with an allocation of 4 bits/sample ($\approx$18 dB s/n) in the lower sub-bands, 3 bits/sample ($\approx$11.5 dB) in the middle sub-bands, and 2 bits/sample ($\approx$7 dB) in the upper sub-bands. Contiguous sub-bands are used. One possible choice of sub-bands is shown above Fig. 6 and will be discussed in greater detail in the next section.

In the other extreme, moderate quality coding at transmission rates of 7.2 kb/s can be achieved by trimming the lowest band to 3 bits/sample, the second band to 2 bits/sample, and the upper bands to $1\frac{1}{3}$ or $1\frac{1}{4}$ bits/sample. In addition, to conserve bandwidth, gaps may be allowed between sub-bands as shown in the band arrangement above Fig. 6. While these gaps introduce a slightly reverberant quality to the coder, the reverberation is generally preferred at this transmission rate to a further reduction in bits/sample and a corresponding increase in noise in the coders, which would be necessary if gaps were not present.

At the intermediate transmission rate of 9.6 kb/s, a distribution of 3, 2, and $1\frac{1}{2}$ bits/sample is possible across the frequency ranges, as shown in Fig. 6 by the solid line. A second alternative, which is also judged close in quality, is given by the dotted line. In this case, 3 bits/sample is used only in the lowest band and 2 bits/sample is used for encoding all upper bands. In both cases, gaps are allowed between bands, as shown above the figure. In listener preference comparisons, 63 percent of the listeners preferred the quality of the first bit/sample distribution (solid line) and 37 percent preferred the quality of the second distribution (dotted line). A third approach was also tried at 9.6 kb/s, which involved 3 bits in the lowest band, 2 bits in the second band, and $1\frac{1}{2}$ bits in the two upper bands, with no gaps appearing between bands. In this way, the reverberant quality of the coder was traded for slightly lower overall s/n. This approach was preferred by only 13 percent of the listeners over that of the first distribution (solid line) and by only 37 percent of the listeners over that of the second distribution (dotted line). Therefore, at 9.6 kb/s, a slight reverberant quality in the coder is preferred by listeners over the lower s/n obtained if no gaps between sub-bands are used.

As observed in the above discussion, many "trade-offs" are possible and the only meaningful criterion for comparing them is a perceptual one. Often it is a matter of trading one type of distortion for another with the hope of finding a compromise that is most acceptable to the majority of listeners.

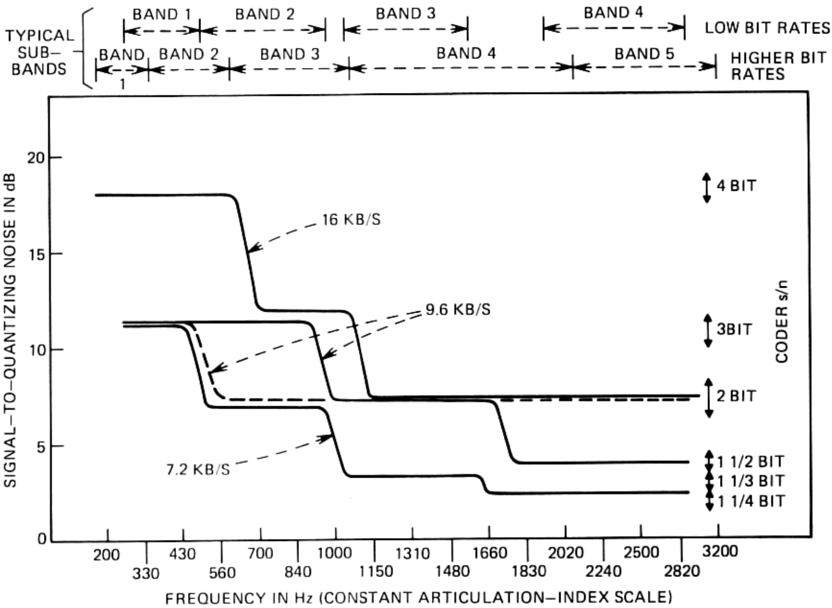

Fig. 6—Signal-to-quantizing noise ratio (s/n) as a function of frequency for bit allocations for 16-μ, 9.6-μ, and 7.2-kb/s coders.

IV. PARTITIONING OF THE SPEECH BAND INTO SUB-BANDS AND MULTIPLEXING OF DATA

The selection of sub-bands involves a variety of considerations. Of preliminary interest is the number of bands. Next, bandwidths and locations of sub-bands must be chosen. This choice is strongly influenced by constraints imposed by the integer-band sampling technique and multiplexing requirements. In this section, we discuss these issues and present candidates for sub-band coders at various bit rates.

Through simulations, a good compromise in the number of bands necessary for sub-band coding was generally found to be about four or five bands. When less than four bands are used, bandwidths become too wide and do not allow for full utilization of the advantages of sub-band encoding. Designs with more than four or five bands tend to consume bandwidth in transition bands of filters in addition to requiring more hardware for practical implementation.

The partitioning of the speech band into sub-bands presents a more difficult problem. A useful preliminary guideline for choosing sub-bands, suggested in Ref. 1, is to partition the speech band into sub-bands that represent approximately equal contributions to the articulation index (AI) under noiseless conditions. In this way each sub-band contains a significant portion of the important frequencies of the speech band. Lower sub-bands should have narrower bandwidths and bandwidths

Table II — Choice of bands for integer-band sampling and 9.6-kHz sampling rate

Decimation Ratio	f_i	$2f_i$	$3f_i$	$4f_i$
1	4800	9600	14400	19200
2	2400	4800	7200	9600
3	1600	3200	4800	6400
4	1200	2400	3600	4800
5	960	1920	2880	3840
6	800	1600	2400	3200
7	686	1371	2057	2743
8	600	1200	1800	2400
9	533	1067	1600	2133
10	480	960	1440	1920
11	436	873	1309	1745
12	400	800	1200	1600
13	369	738	1108	1477
14	343	686	1029	1371
15	320	640	960	1280
16	300	600	900	1200
17	282	565	847	1129
18	267	533	800	1067
19	253	505	758	1011
20	240	480	720	960
21	229	457	686	914
22	218	436	655	873
23	209	417	626	835
24	200	400	600	800
25	192	384	576	768

should become progressively wider with increasing frequency. Gaps between sub-bands can also be determined by this criterion. The allocation of bits in sub-bands, however, is made according to subjective quality considerations, as discussed in the previous section.

The integer-band sampling scheme imposes the constraint that the ratio of upper to lower band edges of sub-bands be $(m_i + 1)/m_i$, where m_i is an integer that may be different for different bands (see Fig. 2). For hardware considerations, it is required that the sampling rates for sub-bands be derivable from a common clock. Furthermore, for digital or CCD hardware implementations, it is desirable to relate these sampling rates to the sampling rate of the bandpass filters by ratios that are integers. Finally, the requirements for multiplexing digitally encoded sub-band signals dictate that the transmission bit rates of each sub-band be a rational fraction of the total bit rate so that the data can be framed and synchronized. Also, a small fraction of this total bit rate must be reserved for synchronizing and framing information.

This multitude of constraints greatly restricts the choices for sub-bands. To assist in the selection of sub-bands, it is helpful to construct tables such as Table II. It is assumed in Table II that the sampling rate of the bandpass filters is 9.6 kHz. Column 1 indicates the integer decimation (reduction) ratios that relate sub-band sampling rates to 9.6 kHz. Column 2 gives bandwidths, f_i, and column 3 gives $2f_i$ sampling rates for the possible sub-bands. Columns 2 through 4 specify choices for band edges $m_i f_i$ ($m_i = 1,2,3, \cdots$). Therefore, all choices for sub-bands are discernible from the tables once the sampling rate for the filters is chosen. Considerations in selecting sub-bands on the basis of articulation index, the distribution of bits/sample across bands, and the total transmission rate quickly reduce the choices of sub-bands further to only a few possibilities. The final choice is still not complete, however, without an analysis of multiplexing requirements. Practically, the transmission rate of each sub-band must be a rational fraction of the total bit rate so that the sub-band data can be multiplexed into a repetitive framed sequence. The lowest common denominator of these rational fractions, including the fraction of transmission rate reserved for synchronization, determines the smallest frame size.

To illustrate these points more clearly, it is helpful to analyze several examples of coders. Table III shows one choice of sub-bands that can be used for 9.6 and 7.2 kb/s four-band coders. The selection of sub-bands is obtained from Table II and corresponds to the low-bit-rate sub-band arrangement illustrated in Figs. 1(b), 3(b), and 6. As seen in Fig. 3(b) or Fig. 6, the bands all have approximately equal width on the warped frequency (constant AI) scale. The lowest sub-band is slightly narrower due to constraints imposed by integer-band sampling. A 107-Hz gap appears between sub-bands 2 and 3 and a 320-Hz gap appears between sub-bands 3 and 4, giving the coders a slightly reverberant quality.

Coder examples A and B represent 9.6 kb/s coders with bit parceling among sub-bands according to distributions shown in Fig. 6 by solid and dotted lines for 9.6 kb/s. Example C is a 7.2 kb/s coder with the bit allocation in Fig. 6. Also included in Table III are sampling rate reduction (decimation) ratios and sampling rates for sub-bands. Relative values of minimum coder step-size (expressed in dB) that match the long-term speech spectrum, as discussed in Section III, eq. (3), are given in column 5. Finally, typical s/n values observed for the examples are given at the bottom of the table. They were measured by comparing simulations with and without coders and represent distortions only contributed by coders and not due to band gaps or filtering.

A fourth coder, example D, was designed for 16 kb/s. The design is based on a filter sampling rate of 10.67 kHz ($\frac{2}{3} \times 16$), which gives the choice of sub-bands shown in Table IV. This led to a slightly better selection of sub-bands for the 16 kb/s coder and resulted in the five-band coder design given in Table V. The sub-band selection corresponds to that shown above Figs. 3(b) and 6. Lower sub-bands overlap slightly to allow for transition bands of filters so that no gaps appear in this frequency range.

Table III — Sub-band coder designs for 9.6 and 7.2 kb/s

Band	Decimate From 9.6 kHz	Band Edges (Hz)	Sub-band Sampling Rates (Hz)	Δ_{MIN} Ratios (dB)	Example A 9.6 kb/s coder		Example B 9.6 kb/s coder		Example C 7.2 kb/s coder	
					Bits/samp.	kb/s	Bits/samp.	kb/s	Bits/samp.	kb/s
1	20	240–480	480	0 (Ref.)	3	1.44	3	1.44	3	1.44
2	10	480–960	960	−3	3	2.88	2	1.92	2	1.92
3	9	1067–1600	1067	−8.5	2	2.13	2	2.13	$1\frac{1}{3}$	1.42
4	5	1920–2880	1920	−14	$1\frac{1}{2}$	2.88	2	3.84	$1\frac{1}{4}$	2.40
Sync						0.27		0.27		0.02
			Total Bit Rate (kb/s)			9.60		9.60		7.20
			Typical s/n (dB)			10.8		10		8.9

Table IV — Choice of bands for integer-band sampling and 10.67-kHz sampling rate.

Decimation Rate	f_i	$2f_i$	$3f_i$	$4f_i$
1	5333	10667	16000	21333
2	2667	5333	8000	10667
3	1778	3556	5333	7111
4	1333	2667	4000	5333
5	1067	2133	3200	4267
6	889	1778	2667	3556
7	762	1524	2286	3048
8	667	1333	2000	2667
9	593	1185	1778	2370
10	533	1067	1600	2133
11	485	970	1455	1939
12	444	889	1333	1778
13	410	821	1231	1641
14	381	762	1143	1524
15	356	711	2133	1422
16	333	667	1000	1333
17	314	627	941	1255
18	296	593	889	1185
19	281	561	842	1123
20	267	533	800	1067
21	254	508	762	1016
22	242	485	727	970
23	232	464	696	928
24	222	444	667	889
25	213	427	640	853
26	205	410	615	821
27	198	395	593	790
28	190	381	571	762
29	184	368	552	736
30	178	356	533	711

Table V — Sub-band coder design for 16 kb/s

Band	Decimate From 10.67 KHz	Band Edges (Hz)	Sub-band Sampling Rates (Hz)	Δ_{MIN} Ratios (dB)	Example D 16-kb/s Coder	
					Bits	kb/s
1	30	178–356	356	−2	4	1.42
2	18	296–593	593	0 (Ref.)	4	2.37
3	10	533–1067	1067	−6	3	3.20
4	5	1067–2133	2133	−11.5	2	4.27
5	5	2133–3200	2133	−18	2	4.27
Sync						0.47
				Total Bit Rate (kb/s)		16.00
				Typical s/n (dB)		13.6

The analysis of the multiplexing requirements for coder examples A through D is summarized in Table VI. The required frame length for multiplexing is 180 bits for the 9.6-kb/s coders, 405 bits for the 7.2-kb/s coder and 135 bits for the 16-kb/s coder. The frame length corresponds

Table VI — Multiplexing and framing information for sub-band coder examples

Band	Fraction of Total Bit Rate	Samples/Frame
Example A (9.6 kb/s)	Frame Length = 180 Bits	
1	27/180	9
2	54/180	18
3	40/180	20
4	54/180	36
Sync	5/180	—
Example B (9.6 kb/s)	Frame Length = 180	
1	27/180	9
2	36/180	18
3	40/180	20
4	72/180	36
Sync	5/180	—
Example C (7.2 kb/s)	Frame Length = 405 Bits*	
1	81/405	27
2	108/405	54
3	80/405	60
4	135/405	108
Sync	1/405	—
Example D (16 kb/s)	Frame Length = 135 Bits	
1	12/135	3
2	20/135	5
3	27/135	9
4	36/135	18
5	36/135	18
Sync	4/135	—

* See text.

to the number of bits that must be stored or transmitted before the multiplexing pattern repeats itself. It is determined by the lowest common denominator of the fractions of total bit rate contributed by sub-bands and by the synchronization channel in column 2. If the frame length is too large, a different sub-band arrangement or bit allocation must be chosen. For example, in the 7.2-kb/s coder, only 1 bit in a frame of 405 bits is reserved for synchronization. If, the third sub-band is quantized with $1\frac{1}{4}$ bits/sample, a frame length of 135 bits is possible with 2 bits reserved for synchronization. This is achieved, of course, at a cost of a slightly reduced coder quality. Column 3 in Table VI gives the number of sub-band samples represented by each frame of data.

The fact that the sub-bands are multiplexed in frames does not necessarily imply that a complete frame of data must be stored before transmission. By careful design of the multiplexer, it is possible to synchronously encode the sub-bands and multiplex them without buffering the data. One scheme for doing this, for coder example A, is illustrated in Table VII. The table depicts the bit allocation for one frame (180 bits)

Table VII — Synchronous multiplexing of coder example A

Bit	Band					Bit	Band					Bit	Band				
No.	1	2	3	4	S	No.	1	2	3	4	S	No.	1	2	3	4	S
1	1		1		X	31				X		61	4			X	
2					X	32		X				62		X			
3					X	33		X				63		X			
4				1	X	34		X		7		64		X	8	13	
5					X	35				X		65				X	
6				X		36				X		66				X	
7		1		X		37	X	4	5			67		7	X		
8			X			38	X					68			X		
9			X	2		39	X			8		69		X		14	
10			2	X		40				X		70		X			
11		X				41	3		X			71		X			
12		X				42			X			72				X	
13		X				43		X				73	X		9		
14				3X		44		X		9		74	X			15	
15				X		45		X				75	X				
16			X			46			6	X		76				X	
17		2	X			47		5		X		77		8		X	
18	X					48			X			78			X		
19	X		3	4		49			X	10		79			X	16	
20	X					50		X				80		X			
21	2			X		51		X				81	5	X			
22			X			52		X				82		X	10		
23			X			53				X		83				X	
24		X		5		54	X			11		84			X	17	
25		X				55	X					85			X		
26		X				56	X		7			86				X	
27		3		X		57		6		X		87		9		X	
28			4	X		58				X		88		X			
29			X	6		59			X	12		89		X		18	
30			X			60			X			90		X			

Table VII — Continued

Bit	Band				
No.	1	2	3	4	S
91			11	X	
92	X				
93	X				
94	X			19	
95				X	
96				X	
97		10	X		
98			X		
99		X		20	
100		X	12		
101	6	X			
102				X	
103			X		
104			X	21	
105				X	
106				X	
107		11X			
108		X			
109		X	13	22	
110				X	
111			X		
112			X		
113	X				
114	X			23	
115	X				
116				X	
117		12		X	
118			14X		
119			X	24	
120		X			

Bit	Band				
No.	1	2	3	4	S
121	7	X			
122		X			
123				X	
124	X			25	
125	X				
126	X				
127		13	15	X	
128				X	
129			X	26	
130			X		
131				X	
132		X			
133		X			
134		X		27	
135				X	
136			16	X	
137		14	X		
138			X		
139		X		28	
140		X			
141	8	X			
142				X	
143	X				
144	X			29	
145	X		17		
146				X	
147		15		X	
148			X		
149			X	30	
150		X			

Bit	Band				
No.	1	2	3	4	S
151		X			
152		X			
153				X	
154			18X	31	
155			X		
156				X	
157		16		X	
158		X			
159		X		32	
160		X			
161	9			X	
162	X				
163	X		19		
164	X			33	
165			X		
166			X		
167		17		X	
168				X	
169		X		34	
170		X			
171		X			
172			20	X	
173			X		
174			X	35	
175				X	
176				X	
177		18X			
178		X			
179		X		36	
180				X	

of data. The first column gives the bit or clock number (at a clock rate of 9.6 kb/s), the next four columns represent sub-bands, the last column represents the synchronization channel and the X's represent allocated bits. Numbers and partitions in each sub-band column represent coder sampling times and sampling intervals. For example, in the first sub-band, nine samples of data (see Table VI) are coded with 3 bits/sample at appropriate clock times, 1, 21, 41, · · ·, 161. This corresponds to one sample every 20 clock times, which is the decimation ratio of sub-band 1 (see Table III). Within each sampling interval three slots (X's) are allocated for transmission of these three bits and, therefore, they do not have to be stored for more than one sampling interval. In the fourth sub-band, bit allocations alternate between two slots and one slot per sampling interval according to the needs of the 1½-bit coder. A frame sequence begins with the transmission of five synchronization bits. The sampling intervals of the sub-bands are offset in time so that these five bits can be transmitted together without conflict. The scheme could easily be implemented with the aid of a read-only memory (ROM).

The synchronous multiplexing scheme is also useful as a means for conveniently ordering bits in a frame even if frames must be buffered for other purposes. Another potentially useful application of synchronous multiplexing occurs in an all-digital implementation, where coder hardware and possibly filter hardware can be shared between sub-bands.

V. DESIGN AND IMPLEMENTATION OF THE FILTERS

The parameters of the bandpass filters are depicted in Fig. 7. The sub-band covers the frequency range from $m_i f_i$ to $(m_i + 1)f_i$. For practical reasons the filter passband must have a slightly narrower frequency range from $m_i f_i + \Delta f$ to $(m_i + 1)f_i - \Delta f$. A transition region, Δf, on the order of 50 to 60 Hz was used in simulations with good results. Filters are 175 to 200-tap FIR designs. If wider transition regions are allowed, lower-order filters can be used at a cost of an increased reverberant quality of the coder. A passband ripple of ±0.5 dB gives satisfactory results in simulations.

Signal frequencies outside of the sub-band are aliased into the sub-band by the decimation process in the transmitter. This aliasing is illustrated by the dotted line in Fig. 7. With a filter stop-band attenuation on the order of 45 dB, this aliasing is not detectable. Near the sub-band edges, a slightly larger amount of aliasing can be allowed, as shown in Fig. 7, in order to keep the filter passbands as wide as possible. Filter attenuations of 12 dB at sub-band edges were used in simulations. Since two such filters are cascaded in the sub-band coder (see Fig. 1), this aliasing is reduced by 24 dB at sub-band edges. It occurs only over a very narrow frequency range (a few Hz) and is not detectable. If lower filter

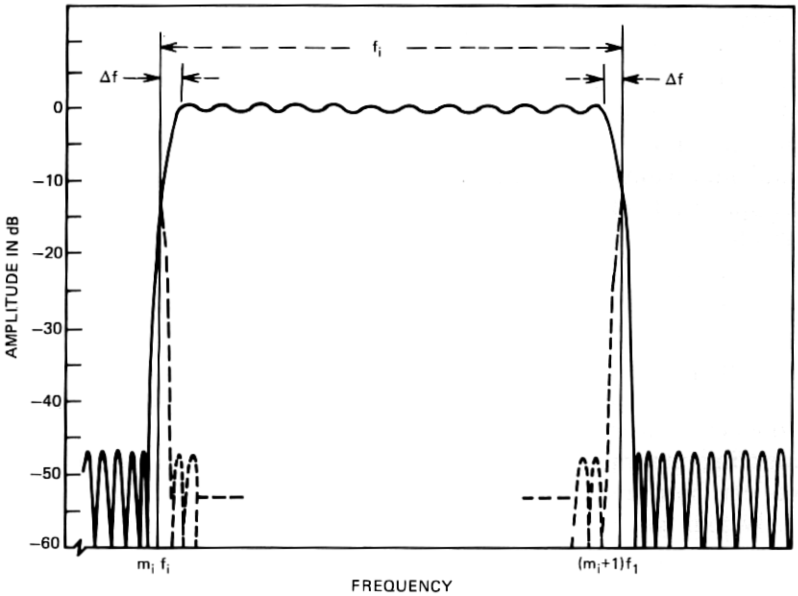

Fig. 7—Parameters of the bandpass filters.

orders (i.e., wider transition bands) are used, correspondingly larger attenuations should be used at band edges to compensate for the smaller slope of the filter roll-off in the transition regions.

The overall frequency response of the sub-band coders was measured by computer simulations. Figure 8a shows results for a 175-tap FIR filter implementation of sub-bands in Table III. Similar results are observed for IIR elliptic filters of order 6, 6, 8, 8 for bands 1 to 4, respectively. Phase distortions introduced by the IIR filters are not perceptible. In fact, the "smearing" of the phase helped to reduce the peak factor of the speech waveforms and led to a slightly improved performance (0.5 dB) in the adaptive coders. Figure 8b shows results of a 200-tap FIR filter implementation of the five-band coder in Table V.

In the receiver, the interpolating filters must have additional passband gain in order to restore the signal energy lost by decimation. The gains are equal to the decimation ratios. For example, if the sampling rate in the transmitter is decimated by 20, the interpolating filter must have a gain of 20 to account for signal energy lost in samples discarded in the decimation process.

Several hardware technologies are amenable to the implementation of sub-band coders. An attractive emerging technology, already mentioned in Ref. 1, is the charge-coupled-device (CCD) technology.[7] It offers possibilities for one or more filters on a chip with analog-to-discrete-time conversion accomplished essentially automatically. Filter outputs can

Table VIII — Comparison of sub-band coders vs ADPCM and ADM (1-bit ADPCM) coding.

Coder Comparison	Preference for Sub-band Coder (%)	Preference for ADPCM (%)
A. 16 kb/s sub-band coder (Example D)		
(1) 24 kb/s ADPCM (3 bit)	58	42
(2) 32 kb/s ADPCM (4 bit)	34	66
B. 9.6 kb/s sub-band coder (Example B)		
(1) 10.2 kb/s ADM	96	4
(2) 12.9 kb/s ADM	82	18
(3) 17.2 kb/s ADM	61	39
C. 7.2 kb/s sub-band coder (Example C)		
(1) 12.9 kb/s ADM	79	21
(2) 17.2 kb/s ADM	56	44

be offered in a convenient sample-and-hold format. The technology may also be tractable for the implementation of the coders.

All-digital technologies also offer many attractive possibilities for the sharing of hardware between sub-bands. Efficient computational methods are possible for implementing filters for decimating and interpolating digital signals.[8] Since digital or CCD filter cutoff frequencies are normalized to the filter-sampling frequencies, the bit rates of the coders can be varied over a limited range by simply varying the master clock frequency—a feat that cannot easily be accomplished with continuous-time filter technologies.

VI. SUBJECTIVE COMPARISONS WITH OTHER WAVEFORM CODING METHODS

Further subjective comparisons have been made at 16 kb/s and 7.2 kb/s in addition to comparisons reported in Ref. 1. Thirteen listeners were asked to compare pairs of sentences for quality and indicate which was better. Two speakers were used in the experiment and several comparisons of the same sentence pairs were made by each listener at different randomly selected times during the test. The results are summarized in Table VIII.

In part A of Table VIII, the quality of the 16-kb/s sub-band coder (Example D) is compared against the quality of 24- and 32-kb/s ADPCM. It was preferred in 58 percent of the sentence pair comparisons against 24-kb/s ADPCM and in 34 percent of the comparisons against 32-kb/s ADPCM. If the results are linearly extrapolated, the quality of the 16-kb/s sub-band coder can be said to be comparable to approximately 26.5-kb/s ADPCM. This is a significant improvement over earlier results reported in Ref. 1. It was obtained by allowing less overlap of the sub-bands and trading the extra bandwidth for more bits/sample in the lower sub-bands.

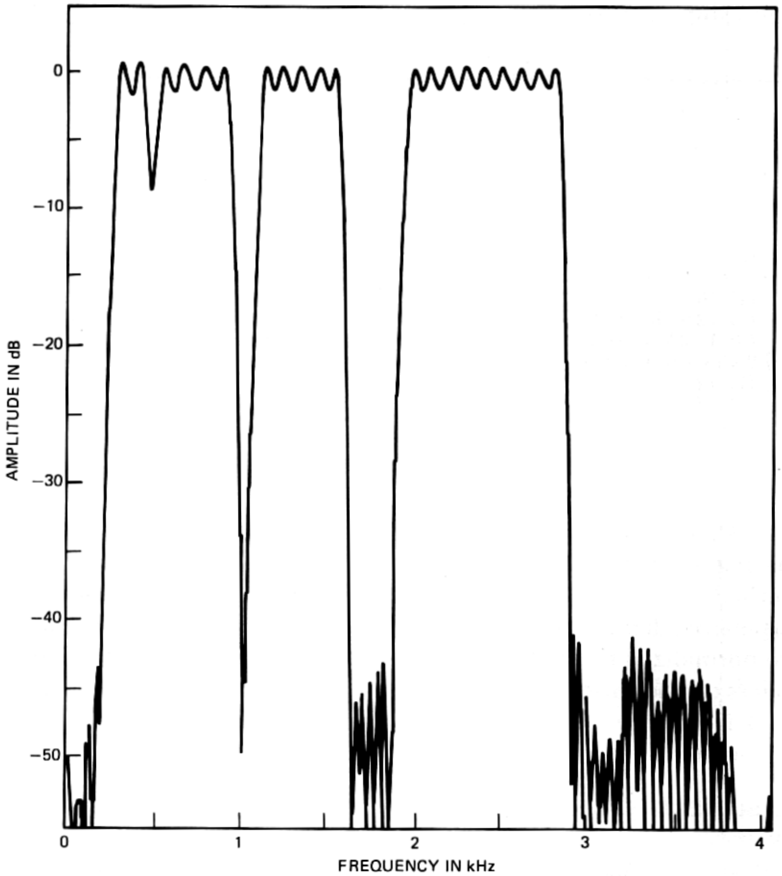

Fig. 8(a)—Measured frequency responses for 7.2- and 9.6-kb/s coders.

The 9.6-kb/s sub-band coder (Example B) is the same coder that was used for comparisons in Ref. 1. It is comparable to 19-kb/s ADM in quality. A slight improvement on this quality was observed from the sub-band coder in Example A.

In part C of Table VIII, the 7.2-kb/s sub-band coder (Example C) is compared against 12.9- and 17.2-kb/s ADM. The quality is preferred over that of 17.2-kb/s ADM and, if the results are linearly extrapolated, it is found to be comparable to approximately 18-kb/s ADM.

As seen by the above comparisons, a consistent advantage of about 10 kb/s in transmission rate is obtained by the sub-band coder over ADPCM or ADM for the same quality. Alternatively, at the same bit rate an improved quality is possible with the sub-band coder.

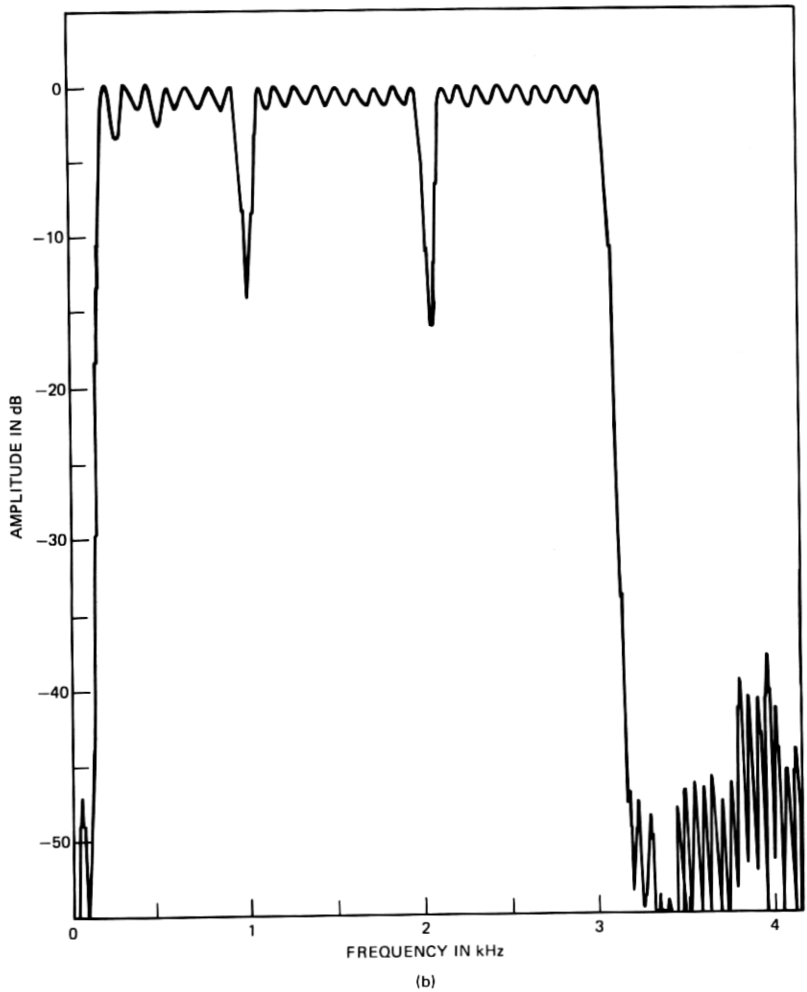

Fig. 8(b)—Measured frequency response for 16-kb/s coder.

VII. CONCLUSIONS

The design of sub-band coders involves the consideration of a large number of parameters and "trade-offs." For many of these parameters, no analytical means exist for choosing them in an optimal way. Consequently, in this paper we have attempted to provide some useful guidelines and insight for selecting parameters of sub-band coders. The guidelines are based on extensive computer simulations and subjective comparisons.

A number of practical considerations involved in selecting sub-bands, multiplexing sub-band data, and implementing the filters have also been discussed. Several sub-band coder designs have been proposed for bit rates of 7.2, 9.6, and 16 kb/s, and their performances have been compared with those of other waveform coding techniques.

VIII ACKNOWLEDGMENTS

The author gratefully acknowledges many valuable discussions at various stages of this work with J. L. Flanagan, D. J. Goodman, N. S. Jayant, and A. J. Presti.

REFERENCES

1. R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Sub-bands," B.S.T.J., *55* (October 1976), pp. 1069–1085. See also Proc. 1976 IEEE Int. Conf. on Acoust. Speech, and Sig. Proc., pp. 233–236.
2. L. L. Beranek, "The Design of Communication Systems," Proc. IRE, *35* (September 1947), pp. 880–890.
3. H. K. Dunn and S. D. White, "Statistical Measurements on Conversational Speech," J. Acoust. Soc. Amer., *11* (1940), 278–288.
4. N. S. Jayant, "Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantizers," Proc. IEEE, *62* (May 1974), 611–632.
5. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J., *52* (September 1973), 1105–1118.
6. D. J. Goodman, private communication.
7. C. H. Sequin and M. F. Tompsett, *Charge Transfer Devices*, suppl. 8 for *Advances in Electronics and Electron Physics*, New York: Academic, 1975.
8. R. E. Crochiere and L. R. Rabiner, "Optimum FIR Digital Filter Implementations for Decimation, Interpolation, and Narrowband Filtering," IEEE Trans. Acoustics, Speech, Signal Process., *ASSP-23*, No. 5 (October 1975), pp. 444–456.