Sub-band Coding of Speech

with

Dynamic Bit Allocation

by

Rafi Rabipour, B. Eng. (McGill University, Montreal)

Submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Engineering.

Department of Electrical Engineering,

McGill University,

Montreal, Canada.

August 1982

ABSTRACT

The results of the investigation of an 8-band sub-band coder with dynamic bit allocation at the rate of 16 Kbits/sec is presented. The band-partitioning technique adopted is the Quadrature Mirror Filtering (QMF) method and particular attention is paid to the problems involved with the use of such filters in a sub-band coder.

The sub-band signals are quantized using Adaptive Pulse Coded Modulation (APCM) quantizers. The number of quantization levels assigned to the sub-bands is revised regularly to adapt the coder to the changes in time of the spectral properties of speech. A simple algorithm is developed for the optimal (in the mean square error sense) assignment of the bits to the sub-bands based on the relative magnitudes of the sub-band energies.

The computer simulation of the coder produces fully intelligible speech of reasonably good quality. This study includes a discussion of the simulation of the coder as well as suggestions for improvements based on perceptual criteria.

Finally practical issues involved with the implementation of a real-time coder are considered.

RÉSUMÉ

Un procédé de codage par découpage en 8 sous-bandes, avec "affectation" dynamique des bits à un taux de 16 Kbits/sec est étudié. La technique de découpage est basée sur l'utilisation de filtres miroirs; une attention particulière est apportée aux problèmes relatifs à l'emploi de tels filters dans le codage par découpage en sous-bandes.

Le signal de chaque bande est quantifié par des quantificateurs MIC adaptés. Le nombre de niveaux de quantifications alloccé à chaque sous-bande s'adapte dans le temps aux propriétés spectrales de signal d'entrée.Un algorithme simple pour determiner l'affectation optimale (par rapport à l'erreur quadratique moyenne) des bits aux sous-bandes, selon le niveau d'energie dans chacune d'elles est présenté.

Le codeur simulé sur ordinateur produit un signal de parole très intelligible et de qualité acceptable. La simulation du codeur est discuté; quelques possibilités d'amélioration basées sur des critères de perception sont suggérées.

Finallement quelques considerations pratiques relatives à la realisation en temps réel du codeur proposé sont examinées.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my academic advisor Dr. P. Kabal for his guidance and support throughout this work.

I also wish to thank members of INRS staff Dr. P. Mermelstein and D. Sloan for helpful discussions and valuable advice.

I am also very grateful to my friends L. and H. Yerushalmi for their aid in the preparation of this report.

Table of Contents

1. INTRODUCTION

The recent advances in the digital hardware technology have enhanced the benefits of digital processing techniques over analog systems. The existence of a highly developed theory accompanied by the availability of increasingly reliable digital signal processing equipment at decreasing costs has made the representation of analog signals in a digital format more feasible than ever.

There are numerous advantages in the choice of transmission of voice signals in a digital format rather than in the analog form. Digital speech can be transmitted through noisy channels without any degradation. Also, in the digital form there is no distinction between speech and data signals, making it possible to use the same communication network for the transmission of both types of signals. This is particularly important at this time when a lot of interest has been shown in the development of integrated voice and data communication systems.

Several special applications increase the significance of digital representation of speech. In transmission of voice signals requiring secrecy, encryption may be performed most effectively in a digital form; bits of infomation can be scrambled according to a scheme known only to the transmitter and the receiver. Also, voice messages in the digital format can be stored efficiently on random access storage devices to be retrieved later.

Another important application is in the area of mobile telephone systems. The state of digital circuit technology is at a level where complex systems can be implemented on very few chips requiring little space. Also logic speeds are fast enough to handle real-time processing of voice signals. Privacy may be maintained by means of encryption.

The conversion from an analog to a digital signal requires periodic sampling of the analog waveform. The Sampling theorem states that bandlimited signals can be sampled without any loss of information if the sampling rate is higher than twice the bandwidth of the signal.

Samples of a continuous signal such as voice have a continuous range of amplitudes. In other words, there are an infinite number of amplitude levels. Transmission of such signals would require a communication channel with infinite bandwidth. Because of the finite capacity of the transmission channel it is necessary to approximate the signal samples by values selected from a discrete set with a finite number of amplitude levels. The selection of the proper level is based on the minimization of a measure of the error.

The process of transformation of samples from a continuous set to a discrete set is called quantization. The error incurred as the result of this approximation is called the quantization error.

The quantized samples are usually represented using binary numbers. With B-bit binary codewords it is possible to represent 2^B different quantization levels.

Given the probability density function of the input and the number of quantization levels it is possible to design the optimal "mean-square error" quantizer [1]. Mean-square error (M.S.E.) is a very common measure of distortion. It is defined to be the variance of the error signal, i.e. the difference signal between the input and the output waveforms :

$$M.S.E. \equiv E[(X(n) - \hat{X}(n))^2]$$

where $X(n)$: input sample
$\hat{X}(n)$: output sample

The distortion of speech caused by the quantization process is typically broad and flat in spectrum. However, due to the characteristics of the speech signal spectrum and the limited detection ability of the auditory system, the quantization noise is not perceived equally at different frequencies. Therefore, the quality of the reproduced speech is expected to improve if the spectrum of the quantization noise is "shaped" in such a way as to exploit the properties of the auditory system.

This motivates the quantization of speech in sub-bands, known as sub-band coding; the speech spectrum is partitioned into a number of sub-bands and each of the sub-bands is quantized and transmitted individually and independently of the others. By controlling the number of quantization levels assigned to each of the sub-bands it is possible to shape the distribution of the quantization noise across the spectrum. The total number of bits allocated to the sub-bands is subject to the constraint set by the finite capacity of the transmission channel.

The idea of sub-band coding of speech was advanced by Crochiere, Webber and Flanagan [2]. The use of Quadrature Mirror Filters as a band-splitting technique was studied by Esteban and Galand [3] who later described the real-time implementation of a sub-band coder [4]. Krasner [5] investigated the choice of sub-bands based on perceptual criteria, and Ramstad and Foss [6], and Barnwell [7] studied the use of recursive Quadrature Mirror Filters. Crochiere and Sambur [8] proposed movable sub-bands and Adaptive Bit Allocation was considered by Grauel [9] and Cheung et. al. [10]. In a recent paper by Makhoul, Berouti and Krasner [11] the shaping of noise in both time and frequency has been discussed.

The work presented here is a study of sub-band coding of speech with Dynamic Bit Allocation (DBA); the bits are re-assigned to the sub-bands at fixed intervals of time to adapt the coder to the changes in time of the spectral properties of speech. The sub-band signals are generated using the Quadrature Mirror Filtering technique. Particular attention is paid to the problems involved with the use of such filters in a DBA sub-band coder.

This report consists of five chapters. After the introduction, Chapter 2 contains a discussion of the theoretical aspects of sub-band coding as well as the description of the basic structure of the coder and its elements. In Chapter 3 the Dynamic Bit Allocation scheme is described, the issues involved are discussed and a bit assignment algorithm is developed. Chapter 4 describes the implementation of the coder on a computer and presents the results of the simulation. Practical considerations are addressed at the end of this chapter. Finally, in the fifth chapter the conclusions of the experiment are discussed and suggestions for future work are given.

2. THE THEORY OF SUB-BAND CODING

In this chapter the principles of sub-band coding are introduced,the basic structure of the coder is presented and its components are described.

2.1 - Structure of the Coder

The basic structure of the coder is shown in Fig. 2.1. Bandlimited speech signal is sampled to produce digital speech signal. This signal is then passed through a bank of filters which partition the speech into N sub-bands. Afterwards the sub-band signals are frequency translated to baseband, compressed and then quantized individually. The quantized samples (actually the indices of the quantizer levels) are transmitted to the receive , where each of the sub-bands is decoded and frequency shifted to its original position in the spectrum. The sub-band signals are then summed up to reproduce speech. The components of the coder are further discussed in the following sections.

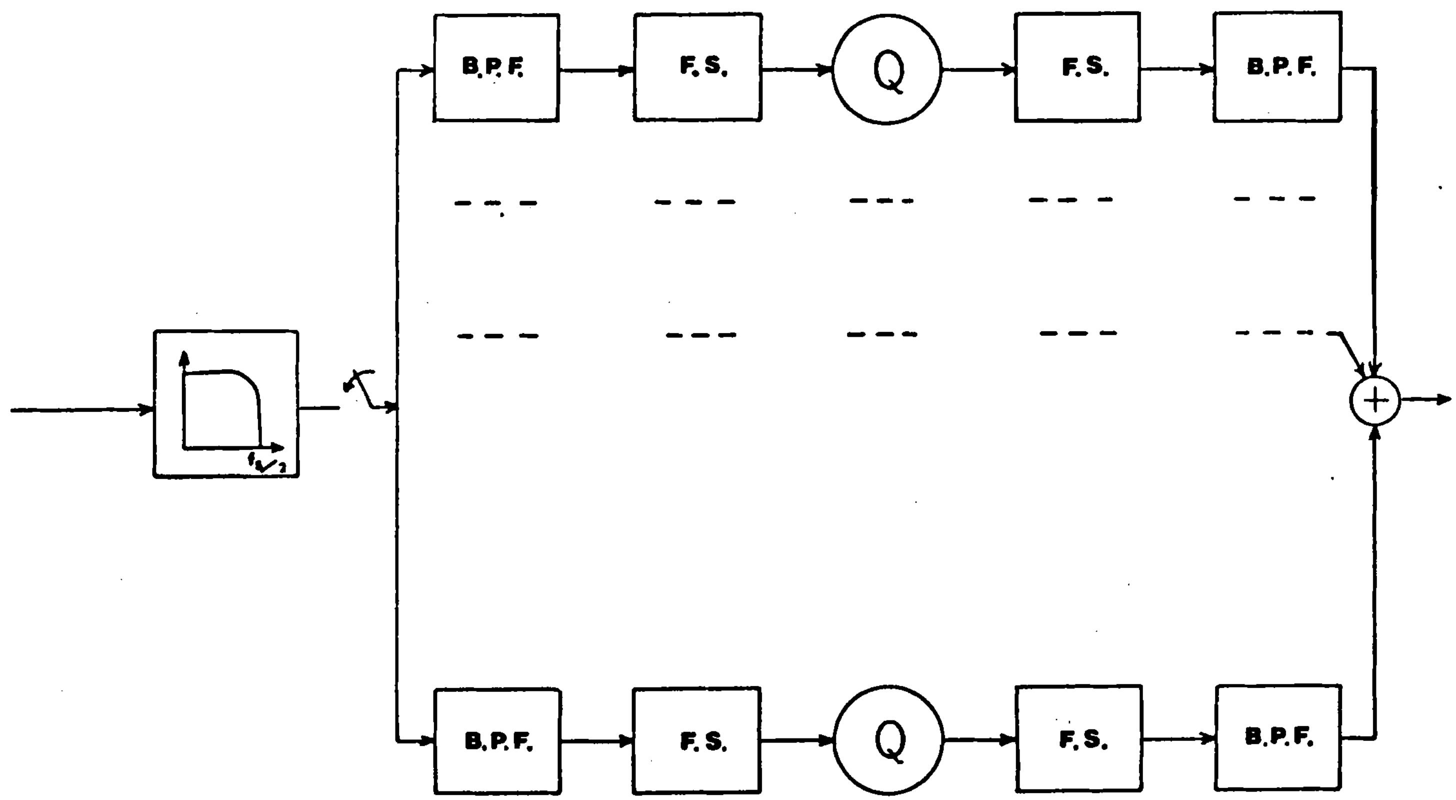

Figure 2.1 - Basic structure of a sub-band coder; bandpass filtering(BPF), frequency shift (FS) to baseband, quantization frequency shift to the original band, bandpass filtering.

2.1.1 - Band Partitioning Techniques

Three of the more important methods for dividing the speech spectrum into sub-bands are the following :

1. The most straightforward method is to utilize a bank of bandpass filters to obtain the sub-band signals. Each sub-band signals then frequency translated to baseband through the use of cosine wave modulation followed by an appropriate lowpass filter. The signal is then compressed by decimating it to its minimal rate of representation, i.e. twice the width of the lowpass filter [2] (Fig. 2.2). Note that if ω_1 and ω_2 are, respectively, the lower and the upper edges of a sub-band, they must satisfy the constraint $3\omega_1 \geq \omega_2$ in order to prevent aliasing as a result of the baseband translation.

At the receiver the complementary operation is performed; the signal is interpolated to its original rate, frequency translated back to its original band (again by cosine wave modulation) and bandpass filtered by a similar bandpass filter (Fig. 2.3).

2. Integer-band sampling - The division of the speech spectrum is again carried out by bandpass filters, but by a proper choice of the sub-band boundaries and its bandwidth the baseband translation can be performed by the decimation of the output of the bandpass filters [2].

This is in fact a special case of the method described previously. The sub-bands $S_n(t)$ are chosen to have a lower cut-off frequency of $m \times f_n$ and an upper cut-off frequency of $(m+1) \times f_n$ where m is an integer. This bandpass signal is sub-sampled

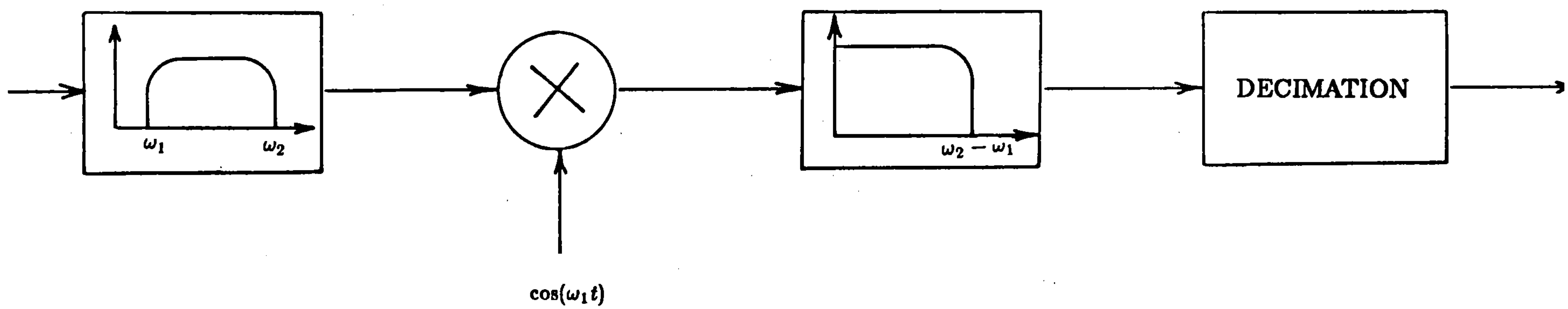

Figure 2.2 - The structure of the transmitter.

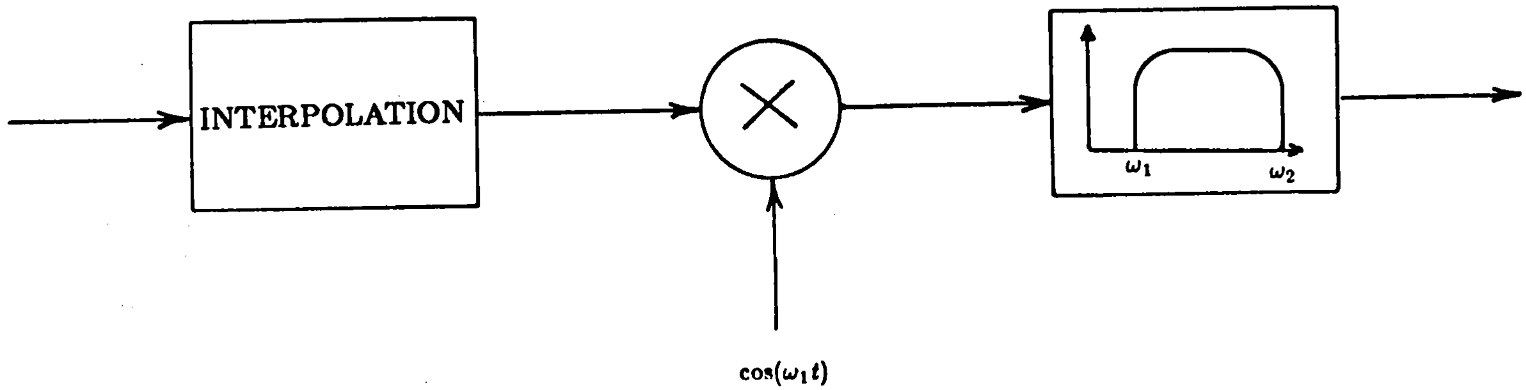

Figure 2.3 - The structure of the receiver.

to $2f_n$ and filtered by an appropriate lowpass filter to produce the compressed sub-band signal. By choosing the upper edge of the sub-band to be an integer multiple of its bandwidth, it is possible to accomplish the modulation essentially "for free"; the value of the cosine wave $x(t) = \cos(2\pi m f_n t)$ sampled at the rate of $2f_n$ is ± 1. Note that the decimation operation may have to be accompanied by alternate sign changes, depending on the value of m.

At the receiver the signal is interpolated to its original rate and bandpass filtered by a similar bandpass filter.

This technique offers some computational savings by forgoing the cosine wave modultations at the cost of further constraints on the position and the width of the sub-bands.

3. Quadrature Mirror Filtering (QMF) - Due to its rather lax constraints on the required filters and also its computational economy, the QMF is an important partitioning technique. The basic QMF structure is shown in Fig. 2.4 [3]. In this figure H_1 is a half-band lowpass filter with (finite) impulse response $h_1(n)$ of length N and H_2 is the corresponding half-band mirror filter, i.e. :

$$|H_1(e^{j2\pi fT})| = |H_2(e^{j\pi(f_s-2f)T}|$$

where f_s is the sampling frequency and $T = \frac{1}{f_s}$. This requirement can be satisfied easily by choosing H_2 such that :

$$h_2(n) = (-1)^n h_1(n) \qquad n = 0, 1, \ldots, N-1$$

After passing through the filters H_1 and H_2, the signals are sub-sampled to half their original sampling rates (by discarding every second sample) and transmitted to the receiver. At the receiver the signals are restored to their original rate by inserting

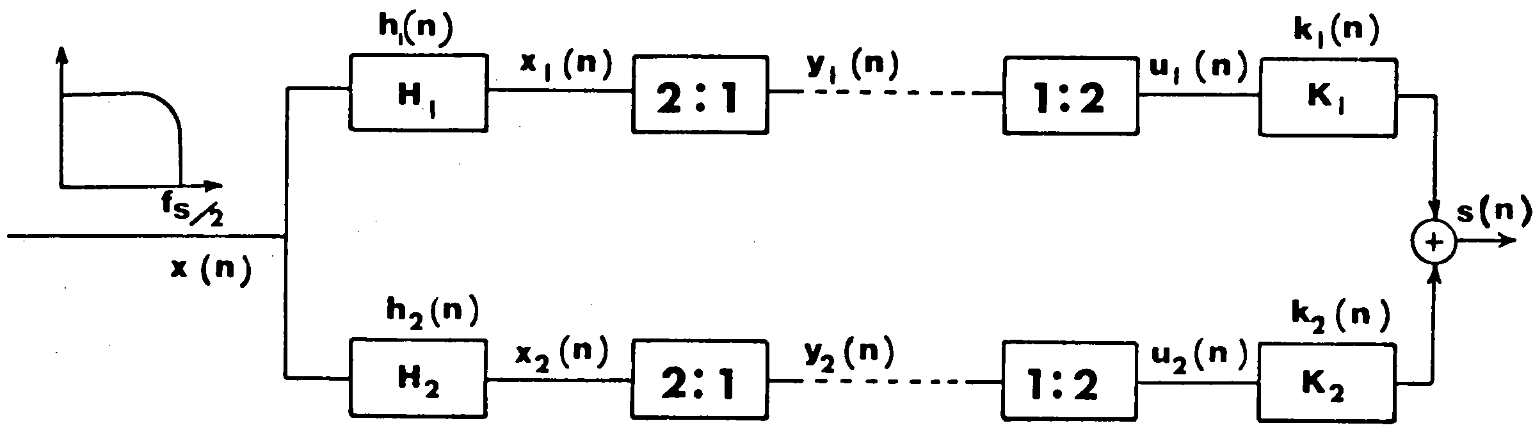

Figure 2.4 - The basic structure of a Quadrature Mirror Filtering (QMF) stage.

a null sample between each pair of signal points, operated on by the filters K_1 and K_2 and then summed up.

The Z-transform of the resulting stream $s(n)$ in terms of the Z-transforms of the input $x(n)$ and the filters is as follows [3]:

$$S(z) = \frac{1}{2}[H_1(z)K_1(z) + H_2(z)K_2(z)]X(z) + \frac{1}{2}[H_1(-z)K_1(-z) + H_2(-z)K_2(-z)]X(-z)$$

The second term in this expression represents the aliasing due to the decimation of the sub-band signals in the transmitter. Due to the symmetry requirement mentioned earlier, we have :

$$H_2(z) = \sum_{n=0}^{N-1}(-1)^n h_1(n)z^n = H_1(-z)$$

Now, if K_1 and K_2 are chosen such that :

$$K_1(z) = H_1(z)$$

$$\text{and} \quad K_2(z) = -H_2(z) = -H_1(-z)$$

then the aliasing terms cancel each other out and we will have :

$$S(z) = \frac{1}{2}[H_1^2(z) - H_1^2(-z)]X(z)$$

With the additional requirements that N be even and H_1 be linear phase, i.e. :

$$H_1(e^{j\omega T}) = H_1(\omega) \times e^{-j(N-1)\pi\frac{\omega}{\omega_s}}$$

where $\omega_s = 2\pi f_s$

then the Z-transform of $s(n)$ evaluated on the unit circle becomes :

$$S(e^{j\omega T}) = \frac{1}{2}e^{-j(N-1)\omega T}X(e^{j\omega T})$$

$$\Rightarrow \quad s(n) = \frac{1}{2}x(n - N + 1)$$

That is, the input signal can be reconstructed exactly with a delay of N-1 samples*. Fig. 2.5 shows the frequency domain representation of the signal in all the stages of the coder.

* Half-band filters are a special case of what are known as fractional-band filters. A detailed analysis of such filters can be found in [12].

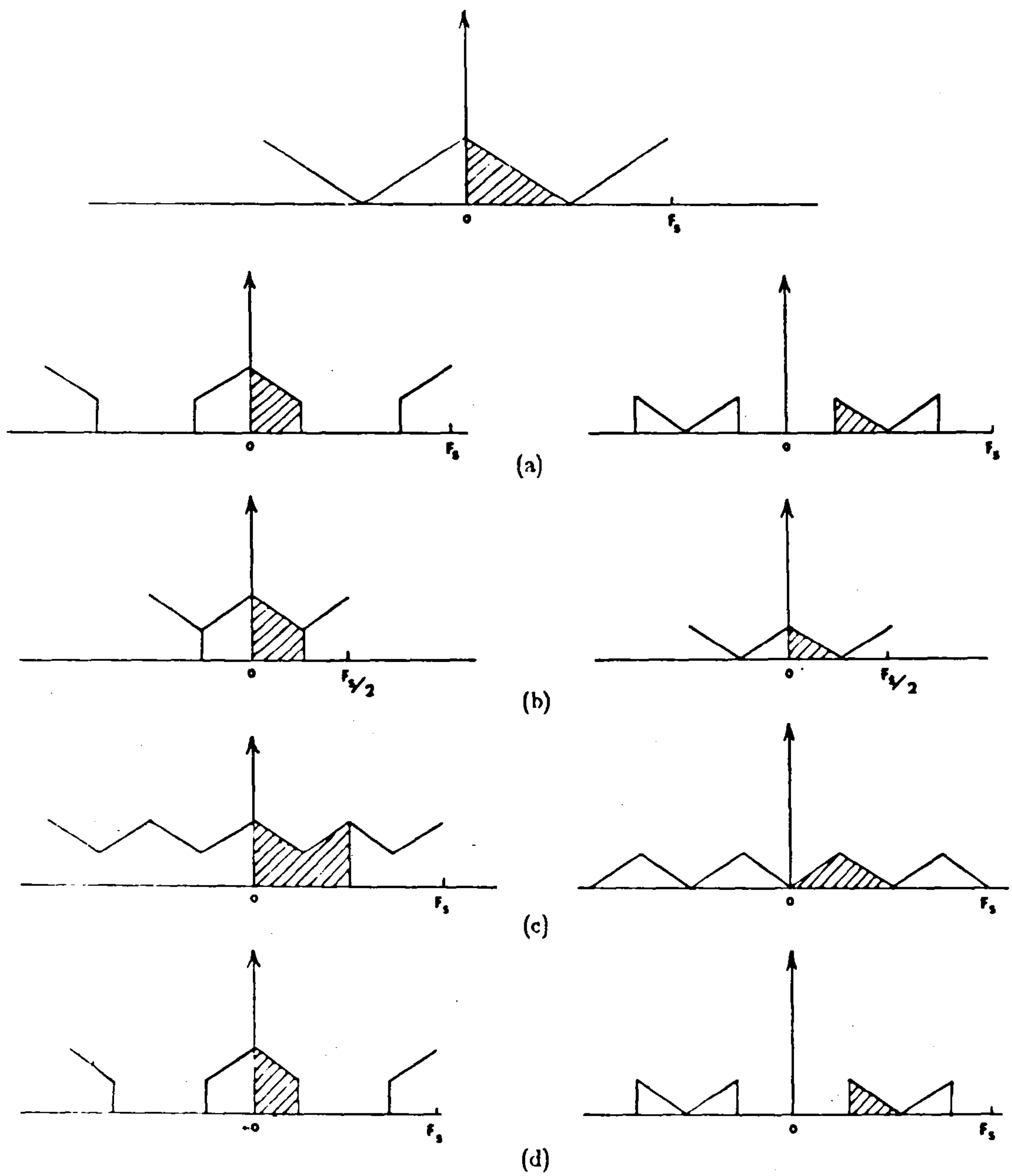

Figure 2.5 - Various stages of sub-band coding with quadrature mirror filters: (a) half-band filtering (b)2:1 decimation (c) 1:2 interpolation (d) half-band filtering.

The basic block described above may be used again to further sub-divide each of the sub-bands. Thus it is possible to form a tree structure to decompose the speech spectrum into narrower bands (Fig. 2.6). Equivalently, a parallel filtering structure may be constructed consisting of filters derived from the QMF tree structure by successive convolution of half-band filters in the appropriate paths, taking into account the decimation operations.

There are advantages in the choice of quadrature mirror filtering over other schemes. There are no stringent requirements on the filters; because of the exact cancellation of the aliasing there is no need for an extremely steep roll-off or a very high stop-band rejection. Therefore, it is possible to employ low-order filters. Also, due to the same reason there is no need for oversampling to prevent the aliasing caused by imperfect filters. Other virtues of the QMF technique include simplicity, ease of implementation and computational economy as a result of short and symmetric filters.

The main disadvantage of the QMF scheme is that there is little control over the position and the width of the sub-bands. The only flexibility afforded is that at any given stage one may further sub-divide any of the existing sub-bands (not all necessarily) by appending to them additional QMF stages.

It must also be realized that the sub-bands cannot be regarded as being completely independent and immune from errors in the transmission of each other. Each sub-band signal contains both a useful component necessary for the correct reconstruction of speech, and some aliasing energy. The aliasing is eliminated by cancellation during the process of reconstruction. Therefore any error in the representation of a sub-band signal, be it caused by channel noise or crude quan-

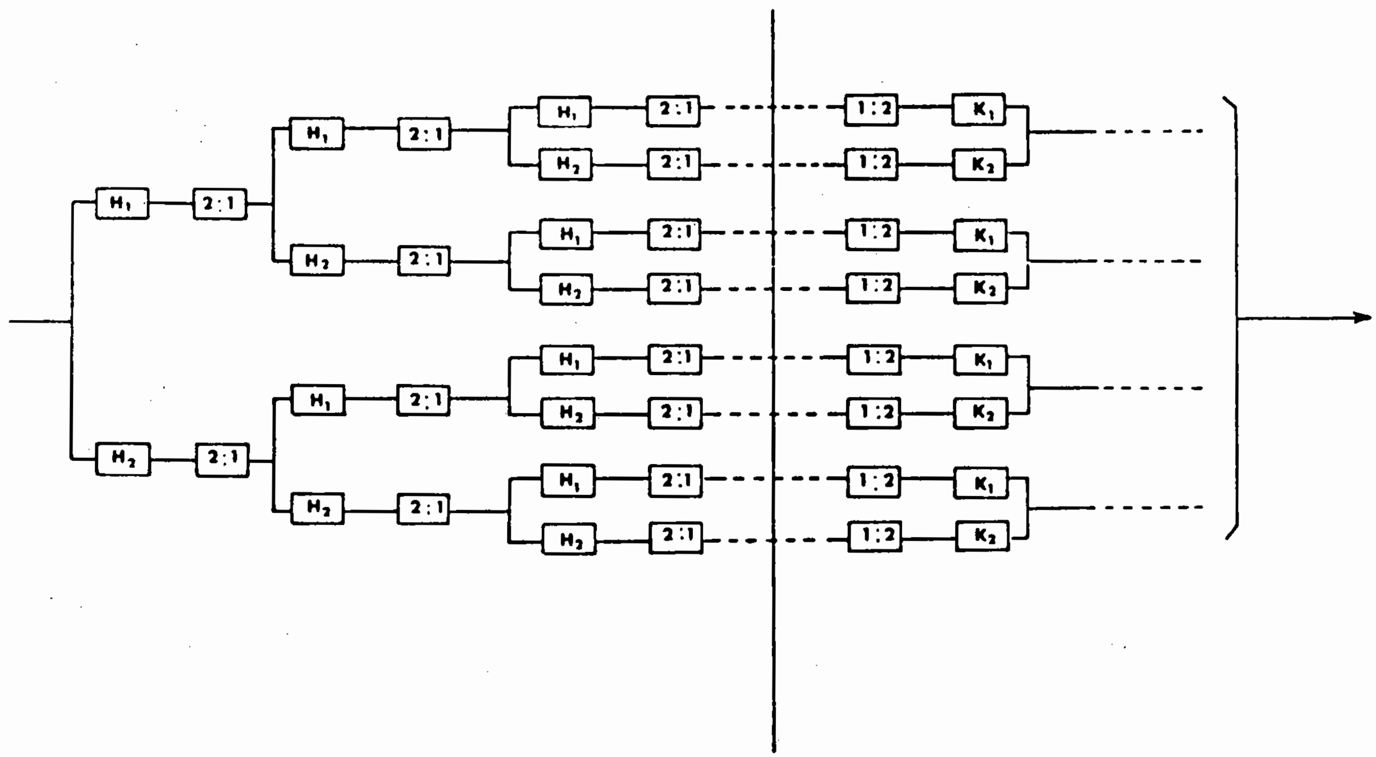

Figure 2.6 - A three-stage quadrature mirror filter tree structure.

tization, will lead to imperfect cancellation of the aliasing components and will therefore result in the presence of noise in other bands.

Due to the computational savings mentioned earlier, the QMF scheme seems to be a well-suited candidate for the real-time implementation of a sub-band coder. Thus it was adopted as the band-splitting method in this study which in fact focuses on the issues involved with the application of the QMF technique to the sub-band coding of speech with dynamic bit allocation.

2.1.2 - Quantization

The process of quantization degrades the signal. In order to reduce the perception of the noise due to the quantization, the quantizer characteristics must be designed so as to minimize a function of the error. A reasonable error function should be monotonically increasing. Two simple distortion measures are the sum of the absolute value of the error and the mean-square error (M.S.E.). Of these two, M.S.E. has been the most popular, mainly due to its mathematical tractability.

Given the distribution of the input and the number of quantization levels, the optimum M.S.E. quantizer can be designed for a unit-variance signal [1]. A quantizer designed in this manner can be used for any input stream with the same probability density function, through a simple scaling of the quantizer by the standard deviation of the signal. One may alternatively scale the signal to unit variance and then re-scale the output of the quantizer to the original level.

Now a suitable quantization technique must be selected. The sub-band signals are narrow-band and their spectrum is roughly flat across any given band.

Since such signals are sampled at their minimal rate, the sample-to-sample correlation is expected to be low. Therefore, PCM quantization techniques seem to be most appropriate as opposed to differential methods. Further reduction of the quantizer error is possible if strategies are adopted whereby the quantizer characteristics are updated on the basis of the short-term rather than the long-term signal statistics. Two such methods were considered; Block Companded Pulse Coded Modulation (BCPCM) and Adaptive Pulse Coded Modulation (ADPCM). These methods are explained below.

a - Block Companded PCM :

In this scheme the input stream is quantized on a block-by-block basis. The updating of the quantizer can be implemented in two ways:

i - The quantizer characteristics are updated at the end of each data block, so as to accomodate the largest sample in the next input block without overload [13].

ii - The quantizer is scaled with a scale-factor which varies during the quantization of a block of data, with a constant slope. The slope is chosen so as to reflect the change of the signal power from one block to the next.

The BCPCM method requires the transmission of extra information, termed "side-information", to the receiver; the scale factor in the first implementation and the slope in the second. Also, buffering of data is necessary to compute these parameters.

b - Adaptive PCM :

In this scheme the quantizer is scaled by a scale-factor on a sample-by-sample basis [14]. The scale-factor is updated by a multiplier whose value is dependent upon the index of the last quantization level. Thus depending on the last index of quantization, the scale-factor acts to expand or shrink the overall range of the quantizer in order to match the quantizer to the input signal. Alternatively, one may think of the action of the scale-factor to be one of normalization of the input signal in order to match the signal amplitude to a fixed quantizer. Since the scale-factor is modified on the basis of past output levels it is possible for the receiver to follow the changes in its magnitude without any side-information.

The APCM has the advantage of having a dynamic range. This is particularly useful in the case of the systems with low rates of transmission, where there are only a small number of quantization levels. This and the fact that APCM requires no side-information make it ideal for low-rate systems. Due to these properties APCM was selected as the quantization strategy in this study.

2.2 - Choice of the Sub-bands

Ideally, better results are expected with increasing number of sub-bands since the quantization noise can be shaped better. Practical limitations such as finite computational speed and low rates of transmission restrict the number of the sub-bands. Also, as the number of the sub-bands increases, due to the decimation in each QMF stage the distance between successive samples increases. This could have an adverse effect on the step-size adaptation of the APCM quantizers.

In choosing the position and the width of the sub-bands it is reasonable to seek combinations that would preserve important speech characteristics such as the formant structure and the pitch information. An additional requirement may be that the sub-bands should occupy portions of the speech band equally important in its perception. This can be attempted by selecting bands which make equal contribution to the Articulation Index [2].

In this study the speech spectrum is divided into eight equal bands. The width of each sub-band is 500 Hz (the input speech is sampled at 8000 samples/sec.). This choice is mainly based on practical considerations and so it is not necessarily the best, perceptually. The performance of the coder is expected to improve if the high-energy lower bands were to be half as wide. The filtering delay, however, would be prohibitively high. On the other hand the less energetic bands higher up in the frequency range need not be as narrow. However, due to practical reasons to be explained in a later chapter, it is preferable not to combine them into wider bands.

2.5 - Allocation of Bits

Once the quantization strategy is determined, it remains to assign the number of quantization levels of each of the sub-bands. The more quantization levels assigned to a band, the more accurately it may be reconstructed and the smaller the error shall be.

The total number of bits allocated to the bands is bound by the capacity of the transmission channel. So, a strategy must be developed to make the best use of the total allocation.

In this section we discuss the bit assignment problem and examine general guidlines for allocation of the number of quantization levels. The discussion in this section is limited to systems with fixed bit assignment, i.e. systems in which the distribution of the bits among the bands remains constant throughout the transmission. The problem of dynamic bit allocation is addressed in the next chapter.

The objective quality of coding systems is conventionally measured by the Signal to Noise Ratio (SNR). SNR is defined to be the ratio of the power of the input signal to the (quantization) noise power. The SNR figure is usually expressed in dB, i.e. :

$$SNR = 10 \log_{10} \frac{\sum_i X_i^2}{\sum_i (X_i - \hat{X}_i)^2}$$

where X = input stream
$\hat{X}$ = output stream

With increasing accuracy, the mean-square error of the reproduced signal diminishes and the SNR figure becomes larger.

To boost the overall SNR of the reconstructed signal it is reasonable to assign more quantization levels to the bands with higher energies at the expense of the less energetic ones. However, the subjective quality of speech does not always correlate well with the SNR measure. A system with the highest SNR is not necessarily judged to be the best, subjectively. This is mainly due to the phenomenon of masking, because of which an audio signal inhibits the perception of another audio signal. If the signal-to-quantization-error level is high enough, the perception of noise may be suppressed by the speech signal itself. Experiments have been carried out to estimate the value of SNR needed to mask the quantization noise [5,15]. Fig. 2.7 is a plot of such SNR values as a function of frequency [15].

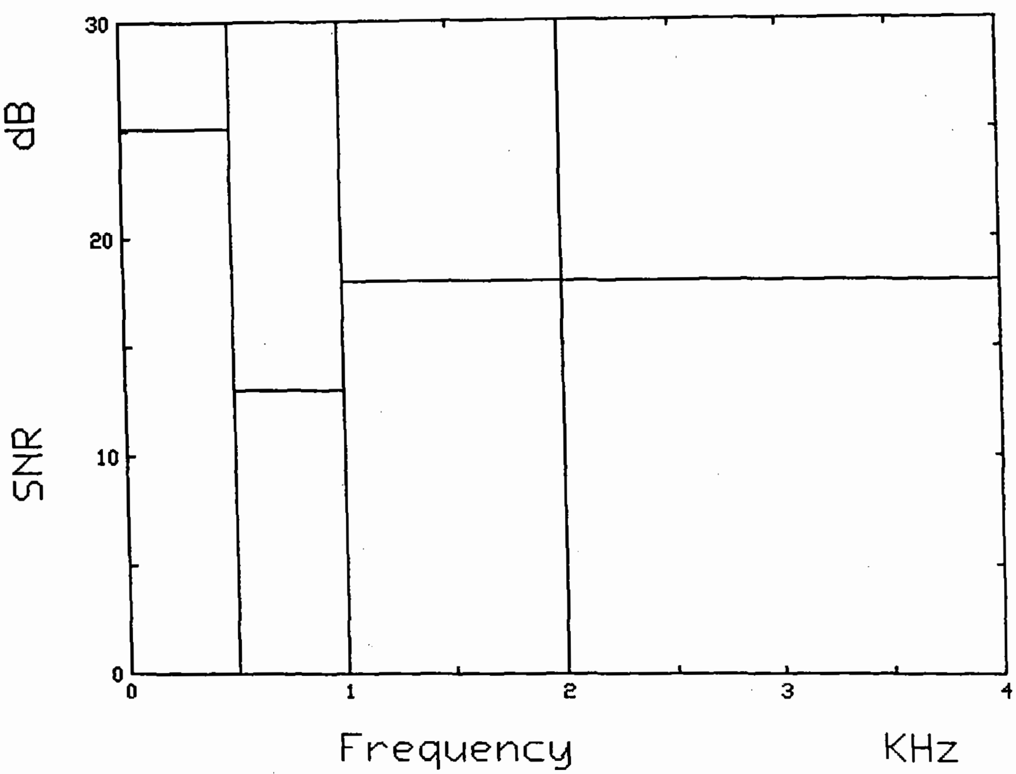

Figure 2.7 - SNR values at just-noticable noise level versus frequency for a sub-band coder with bands : 0-500 Hz, 0.5-1 KHz, 1-2 KHz, 2-4 KHz.

Based on a 17-band sub-band coder, Krasner [5] reports complete masking of the quantization noise by the signal at the rate of 34.4 Kbits/sec. Although at rates as low as 16 Kbits/sec it is not possible to sustain the signal-to-noise levels of Fig. 2.7, it can still be used as a guidline in the allocation of the quantization levels.

Other general rules include the assignment of bits on the basis of the contribution of the bands to the Articulation Index [16] and the requirement that the bands in the lower range of the spectrum be given sufficient weight since they contain most of the energy as well as important information such as the pitch and the first formants.

3. DYNAMIC BIT ALLOCATION (DBA)

This chapter addresses the issues involved in sub-band coding with dynamic bit allocation. The motivation for DBA is given and a procedure is developed for the optimal (in the M.S.E. sense) allocation of bits to the sub-bands.

3.1 - Motivation

The idea of coding of the speech in sub-bands was motivated by the non-flat spectral density of speech signals. Given a fixed rate of transmission, it is sought to exploit this characteristic of speech spectrum in order to improve the quality by assigning more quantization levels to the more important bands. The importance of a band is usually associated with the amount of energy it contains. Thus, in a fixed bit-assignment system the allocation of the quantization levels is based on the statistics gathered from long utterances.

However, speech is not a stationary signal and its characteristics vary considerably depending on the type of sound being uttered; the concentration of the energy in the sub-bands differs greatly from the noise-like fricatives to the quasi-periodic vowels with definite resonances. The varying position of the formants as well as the changes from speaker to speaker affect the speech spectrum. So, it is reasonable to expect improvements in the performance of the coder if the distribution of the bits among the sub-bands is varied in regular intervals so as to quantize the more energetic bands with more fidelity.

At this point it is necessary to select a fidelity criterion and develop a procedure to allocate the available bits to the sub-bands on the basis of energy estimates obtained from a (fixed-length) block of input data. One such scheme is described below.

3.2 - Bit Allocation Algorithm

The objective in allocation of bits to the sub-bands is to maximize a measure of quality given a fixed rate of transmission. Unfortunately, a function to give an objective measure of the quality of speech is yet to be found. Therefore, the mean-square error (MSE) is selected as the fidelity criterion. The minimization of MSE is equivalent to the maximization of the signal to noise ratio (SNR). Since a dynamic bit allocation algorithm operates on frames of data, this criterion amounts to maximizing what is known as the Segmental Signal to Noise Ratio (SSNR) [17] of the overall signal. SSNR is defined to be the average of the SNR (in dB) of fixed length segments of data. That is, if SNR_i is the SNR of the i^{th} frame of data, then:

$$SSNR = \frac{1}{N} \sum_{i=1}^{N} SNR_i \tag{3.1}$$

Compared to SNR, the segmental SNR is known to correlate better with subjective performance. Still, SSNR may not necessarily be an adequate measure; highest SNR or SSNR ratings do not always imply best perceptual quality. So, the MSE criterion is used only as a starting point. After the system which reconstructs speech with maximum SSNR is attained, the parameters must be perturbed and modifications must be made so as to accomodate the perceptual criteria.

An optimal MSE procedure for the assignment of bits to the sub-bands does exist which requires the knowledge of the tabular form of the rate-distortion function of the sub-band signals [18]. This algorithm is described below:

Determine the MSE due to a j-bit quantizer (j varies from 1 to the maximum number of bits allowed) with a unit variance input and a given distribution. The MSE of a sub-band signal applied to a j-bit quantizer can be computed by multiplying its variance by the MSE of the unit variance source with the same probability density function. Now the bits are distributed among the bands, one-by-one so as to maximize the decrease in the MSE.

As simple as it is, this algorithm is computationally burdensome since at each and every step one needs to sort the MSE's. Seeking a more efficient method, it is attempted to find an analytical solution to the bit assignment problem.

Assuming the sub-band signal has a rate-distortion function $K(B_i)$, where B_i is the number of bits assigned to the i^{th} sub-band, then given the MSE criterion, the total distortion is equal to the sum of the errors due to each band:

$$D = \sum_{i=1}^{N} D_i = \sum_{i=1}^{N} \sigma_i^2 K(B_i) \tag{3.2}$$

where N = total number of sub-bands
σ_i = standard deviation of the i^{th} sub-band
D_i = error in the i^{th} sub-band

The total distortion is to be minimized with respect to the B_i, subject to the capacity constraint:

$$\sum_{i=1}^{N} M_i B_i = C \tag{3.3}$$

where C = Capacity of the channel
M_i = the relative rate of transmission of the i^{th} band

Using the method of the Lagrange multiplier we get:

$$F = \sum_{i=1}^{N} \sigma_i^2 K(B_i) + \lambda\left(\sum_{i=1}^{N} M_i B_i - C\right) \tag{3.4}$$

$$\Rightarrow \frac{\partial F}{\partial B_i} = \sigma_i^2 \acute{K}(B_i) + \lambda M_i = 0 \tag{3.5}$$

$$\Rightarrow \frac{\sigma_i^2 \acute{K}(B_i)}{M_i} = -\lambda \tag{3.6}$$

Note the implicit assumption that all sub-bands have the same rate-distortion function.

To carry on with the development of an algorithm with an analytic solution for the B_i's it is necessary to find the rate-distortion function $K(B_i)$. In order to determine the rate-distortion function we need to know the distribution of the sub-band signals.

Histograms of the sub-band signals were produced by dividing a given range into a large number of "bins" and counting the number of times a sample falls into each of the bins. According to the histograms, the Gamma distribution seems to be a satisfactory statistical description of the sub-band signals, with the exception of the first sub-band which can also be described by a Gaussian distribution (Fig. 3.1).

The next step was to design uniform MSE quantizers for an input stream with Gamma distribution and unit variance [19]. The plot of the mean-square error

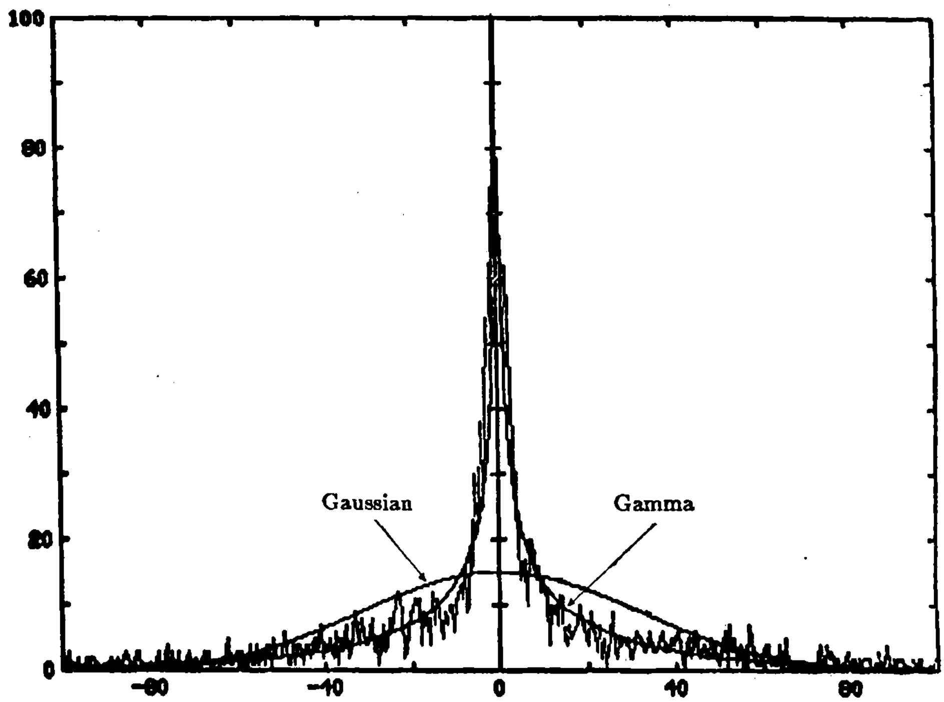

Figure 3.1 - Histogram of a sub-band signal with Gamma and Gaussian fits.

versus the number of the quantizer levels shows that the rate-distortion function can be adequately modelled by an exponential curve. The best MSE exponential fit decays with about 5.5 dB/bit (Fig. 3.2). So, K(B) may be replaced by its model ae^{-gB}, where $g \approx 1.27$.

Now it is possible to work out a solution to the bit assignment problem. Going back to (3.6) and replacing $K(B)$ by its model we get:

$$\frac{-g\sigma_i^2 a e^{-gB_i}}{M_i} = -\lambda \tag{3.7}$$

Combining (3.7) with the constraint equation (3.3), the solution is:

$$B_i = \frac{C}{\sum_{k=1}^{N} M_k} + \frac{1}{g}\left(\ln\left(\frac{\sigma_i^2}{M_i}\right) - G\right) \tag{3.8}$$

where

$$G = \frac{\ln\left[\prod_{k=1}^{N}\left(\frac{\sigma_k^2}{M_k}\right)^{M_k}\right]}{\sum_{k=1}^{N} M_k} \tag{3.9}$$

Another point worth noting is that from (3.7):

$$\frac{D_i}{M_i} = \frac{\lambda}{g} \tag{3.10}$$

As mentioned before, the M_i express the relative rate of transmission of the bands. Since the sub-bands are sampled at their Nyquist rate, M_i also represent the relative widths of the sub-bands. With this interpretation of M_i , according to (3.10) the optimal (in the MSE sense) allocation of bits should result in equal distortion per unit of frequency, i.e. the quantization error is expected to have a flat spectrum. Note that this stems from the exponential characteristic of the rate-distortion function.

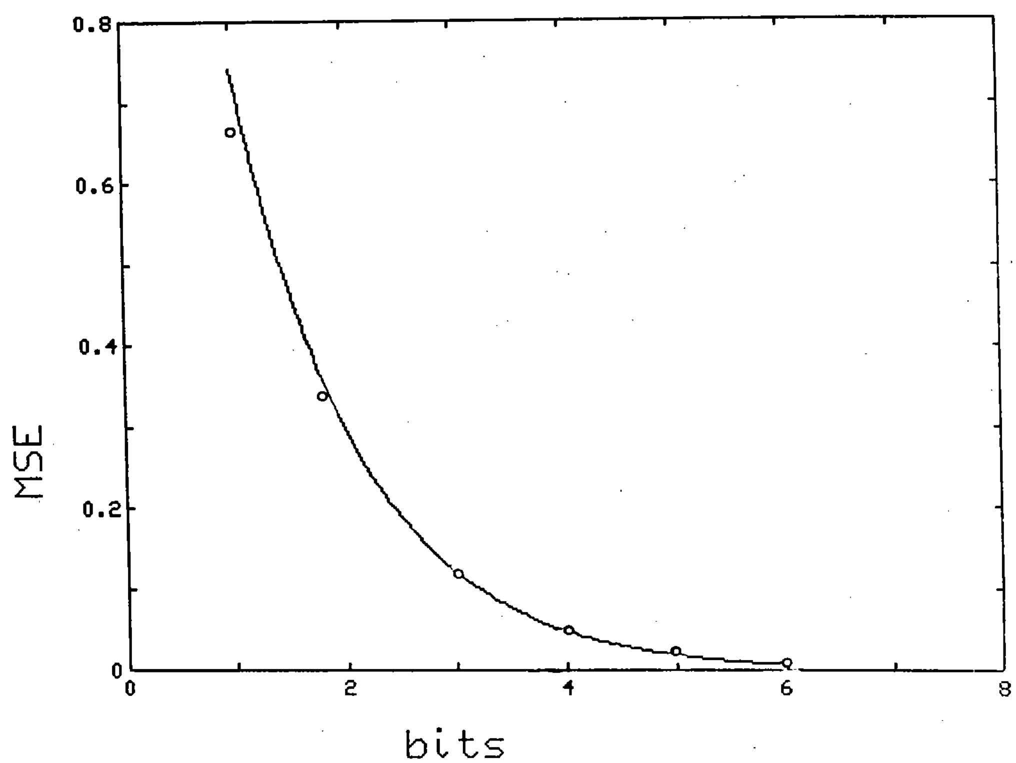

Figure 3.2 - Plot of MSE values of optimal uniform quantizer with Gamma-distributed input and its best MSE exponential fit.

Given the solution above, it is easy to determine the magnitude of error per unit frequency. This can be done by noting that if the energy per unit frequency of a band k is at this "critical" level, then it will be assigned zero bits. Therefore, from (3.8) we get:

$$0 = \frac{C}{\sum_{i=1}^{N} M_i} + \frac{1}{g}(\ln(\frac{{\sigma_k}^2}{M_k}) - G) \tag{3.11}$$

$$\Rightarrow \frac{{\sigma_k}^2}{M_k} = \exp[-g\frac{C}{\sum_{i=1}^{N} M_i} + G] \tag{3.12}$$

Using (3.9):

$$\alpha = \frac{{\sigma_k}^2}{M_k} = \exp[-\frac{gC}{\sum_{i=1}^{N} M_i}][\prod_{i=1}^{N}(\frac{{\sigma_i}^2}{M_i})^{M_i}]^{\frac{1}{\sum_{i=1}^{N} M_i}} \tag{3.13}$$

where $\alpha =$ "critical distortion level"

The solution given in (3.8) needs some refinements to be of use in a real coder; practical considerations require that B_i's be integer. Then, it may be preferable to limit the number of bits which can be assigned to a band in order to prevent the allocation of too many quantization levels to a single sub-band. Also, it often happens that the energy of one or more bands is lower than the critical distortion level, resulting in negative bit assignments.

A procedure is given to provide for the practical requirements mentioned above. Three cases are considered depending on the value of M_i.

3.2.1 - Equal Bandwidths ($M_i = 1 \quad i = 1, \ldots, N$)

From (3.8) and (3.9) the solution becomes:

$$B_i = \frac{C}{N} + \frac{1}{g} \ln \frac{\sigma_i^2}{(\prod_{k=1}^{N} \sigma_k^2)^{\frac{1}{N}}} \tag{3.14}$$

And the distortion per sub-band (3.13) is:

$$\alpha = (\prod_{i=1}^{N} \sigma_i^2)^{\frac{1}{N}} e^{-\frac{gC}{N}} \tag{3.15}$$

The bit allocation algorithm consists of two passes. In the first pass an integer number of bits (IB_i) is computed for each band, where IB_i is given below:

$$IB_i = [B_i] \tag{3.16}$$

where $[\,]$ = nearest integer

Also, IB_i is limited to the range $[0,L]$, where L is the maximum number of bits a band is allowed to be assigned. In the same pass we also compute a quantity A_i (the Residual Claim of a band to the bit resource) according to:

$$A_i = B_i - IB_i \tag{3.17}$$

At the end of this pass another parameter, T is computed as:

$$T = C - \sum_{i=1}^{N} IB_i \tag{3.18}$$

A positive T is the number of bits not yet assigned (due to the clamping operation and/or the nearest integer assignment). These can now be distributed, one-by-one, to the sub-band with the highest claim excluding those that have reached the clamping level L. To show that this is indeed the optimal assignment, consider two sub-bands i and j which have already been given b_i and b_j bits, respectively. Furthermore assume that each of these bands has a residual claim, say a_i and a_j. We want to show that if a single leftover bit is to be allocated to one of these bands, it should be given to the one with the highest residual claim, regardless of the relative magnitudes of the sub-band energies, σ_i^2 and σ_j^2.

With the allocation of the b_i and b_j bits to the bands i and j, the total distortion due to these bands is:

$$\begin{aligned} D &= D_i + D_j \\ &= \sigma_i^2 K(B_i) + \sigma_j^2 K(B_j) \\ &= a\sigma_i^2 e^{-gB_i} + a\sigma_j^2 e^{-gB_j} \end{aligned} \tag{3.19}$$

The effect of assigning an additional bit to one of these bands is to multiply and hence, reduce the distortion due to that band by the fraction e^{-g}. So, it is intuitively clear that the additional bit should be given to the band with the largest distortion. Since the bands i and j have the residual claims a_i and a_j, respectively, their original assignment (i.e. the assignment computed in the first pass of the bit allocation algorithm) has been $b_i + a_i$ and $b_j + a_j$, respectively. From (3.8) we have:

$$\sigma_i^2 e^{-g(b_i+a_i)} = \sigma_j^2 e^{-g(b_j+a_j)} \tag{3.20}$$

$$\Rightarrow \frac{\sigma_i^2 e^{-gb_i}}{\sigma_j^2 e^{-gb_j}} = \frac{D_i}{D_j} = \frac{e^{ga_i}}{e^{ga_j}} \tag{3.21}$$

From (3.21) we conclude that if $a_i > a_j$, then the distortion due to the band i (D_i) is greater than D_j , the distortion due to the band j. So, the extra bit should be allocated to band i.

A negative T represents the number of extra bits assigned to the sub-bands (due to negative B_i's and/or the nearest integer assignment). These must be taken away, one-by-one, from the sub-band with the lowest claim.

Excluding the effect of clamping the number of bits, this procedure gives the optimal (in the MSE sense) integer solution to the bit assignment problem.

In the above procedure the number of operations in the second pass is proportional to the magnitude of T given in (3.18). Minor modifications may be introduced to reduce the absolute value of T and hence the number of operations. The following are examples of such modifications:

i - The number of negative assignments can be reduced by comparing the sub-band energies with α (3.15) and leaving out bands with energies lower than or equal to the critical distortion level. Needless to say, the variances corresponding to such bands must be excluded from (3.14) once the procedure is started.

ii - The assumption can be made that if the length of the block of data is short enough, the variations in the energies of the sub-bands are small enough from one block to the next so that the value of T in the next computation will change only slightly. This assumption can be put to use by replacing the capacity C in (3.14) with the variable $W(k)$, where k is the index of time and $W(.)$ is given by:

$$\begin{aligned} &W(1) = C \\ &W(k) = W(k-1) + T(k-1) \qquad 1 < k \end{aligned} \tag{3.22}$$

3.2.2 - Unequal Bandwidths

Two cases are considered:

(a) - $M_i \leq 2 \quad i = 1, \ldots, N$:

The first pass of the algorithm given in **3.2.1** remains the same, with the slight modification of (3.18) to:

$$T = C - \sum_{i=1}^{N} (M_i \times IB_i) \tag{3.23}$$

The second pass, however, is not as straightforward as in the previous case; the residual claim of the bands is no longer a valid basis for the addition or the subtraction of the "T" bits and a new criterion is necessary. The reason for this is that a bit assigned to a band i, where $M_i = 2$, represents two bits of the total allocation.

We will first consider the case of leftover bits, i.e. $T > 0$. If $T = 1$ it is obvious that only bands can be considered whose relative rates of transmission (M) are one. So, the arguments of case **3.2.1** are applicable. However, if $T > 1$, the decision must be based on the assignment of two bits at a time. It is therefore necessary to determine whether it is best to:

1 - Give a single bit to a band with $M = 2$ (recall that this is equivalent to taking two bits away from T).

2 - Give two bits to a band with $M = 1$.

3 - Give one bit to each of two bands with $M = 1$.

Consider three sub-bands, indicated by the sub-scripts i, j and k, with the residual claims a_i, a_j and a_k, respectively. Furthermore, assume that a_i is the greatest among the sub-bands with $M = 2$, and a_j and a_k are the greatest among the bands with $M = 1$, such that $a_j \geq a_k$. It may be shown* that for the optimal (in the MSE sense) assignment of the bits to the sub-bands:

* See Appendix-A for the proof

i - 1 bit should be given to the band i if:

$$a_i + \frac{\ln 2}{g} > \frac{1}{g} \ln\left(e^{g a_j} + e^{g a_k}\right) \tag{3.24}$$

ii - 1 bit must be assigned to the band j and the procedure must be repeated.

The extension of the discussion above to the case $T < 0$ is obvious; the only difference is that the sub-bands with the least residual claims should be compared. The results are the same with the direction of the inequalities reversed.

(b) - $M_i > 0 \quad i = 1, \ldots, N$:

In this case the first pass can take place as in (a), but the combinations to be considered in the second pass are so numerous as to make the procedure more burdensome, computationally, than even an exhaustive search. Therefore, it will not be discussed.

3.3 - On the MSE criterion

The optimal (in the MSE sense) bit allocation results in quantization noise with a flat spectrum across the speech band (Fig. 3.3). However, given a fixed rate of transmission, flat noise distribution does not conform with perceptual criteria [15]. An adverse result of flat noise is the following. The position of the quantization noise relative to the sub-band energies depends on the transmission rate (3.13). At low rates, the energy of the bands higher up in the frequency range will be below the critical distortion level and hence these bands will not be transmitted, resulting in muffled and unnatural sounding speech.

One possibility for alleviating this problem is the use of weighting. Weighting factors can be assigned to each of the bands depending on the desired shape of the quantization noise. These factors would weight the sub-band energies during the bit-assignment computations.

However, the usefulness of fixed weights is expected to be limited; for instance weights fit for voiced utterances are not suitable for unvoiced speech, since the perceptual importance of the sub-bands is not the same in these cases. A functional form of weighting may be employed instead [20], where the weights are of the form:

$$\omega_i = \sigma_i^{-2\gamma} \tag{3.25}$$

and γ is a parameter whose optimum value is to be determined subjectively (through listening tests).

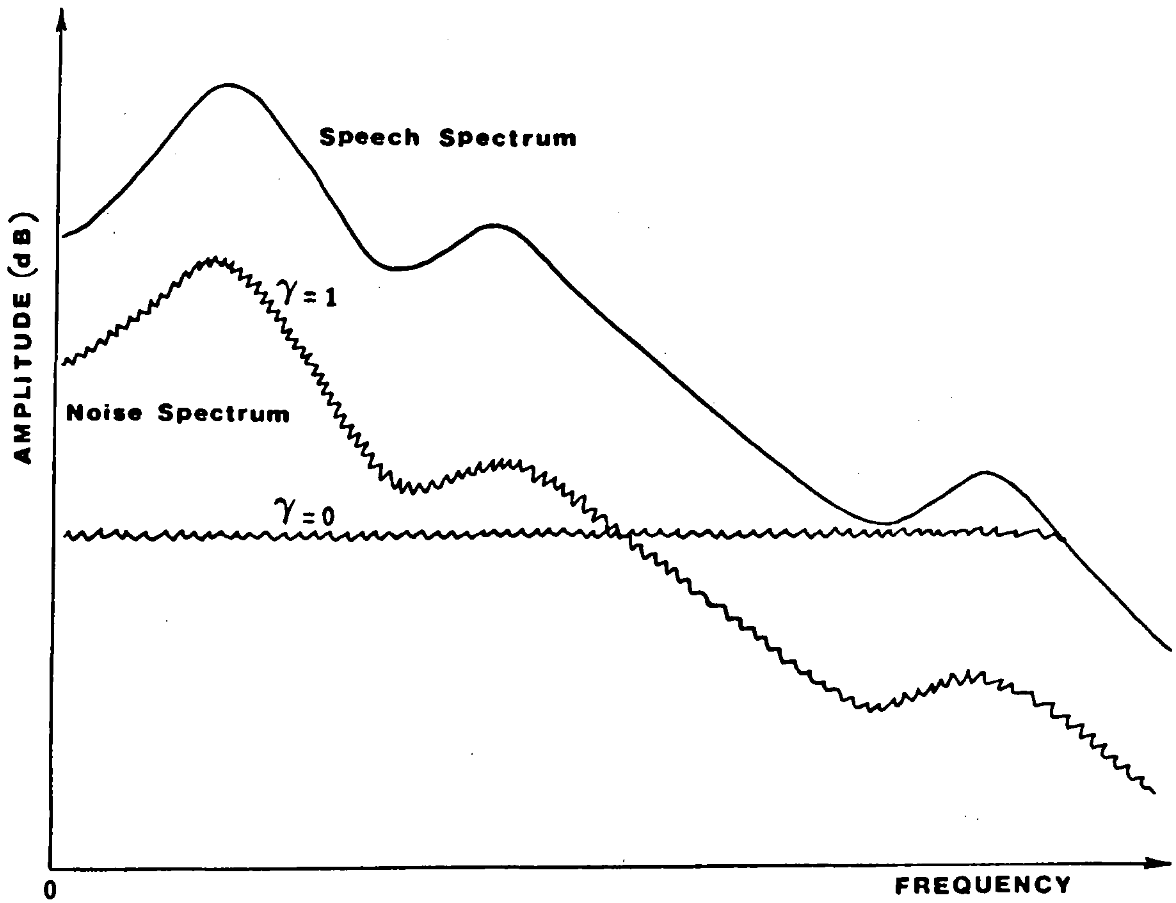

Figure 3.3 - Noise shaping with the use of weights.

Special cases occur for $\gamma = 0$ and $\gamma = 1$. The case $\gamma = 0$ amounts to no weighting and results in a flat quantization noise. The case $\gamma = 1$ gives equal importance to every band and results in uniform distribution of the bits. In this case the signal to noise ratio is the same for every band and the noise spectrum will shadow that of the speech signal (Fig. 3.3). The optimum value of γ is expected to lie in between these two extremes.

4. SIMULATION OF THE CODER

This chapter describes the simulation of the coder and discusses the results of the experiments. The practical issues involved with the implementation of a QMF sub-band coder are considered at the end of the chapter.

4.1 - Introduction

The coder was investigated through computer simulation. High level programming languages allow quick and easy implementation of the coder. A computer simulation is also a flexible means through which a system can be thoroughly studied, since it facilitates speedy modification of parameters and provides easy access to each and every component of the coder.

The simulation was carried out on a Vax 11/780 computer, in Fortran. The audio facilities included a 15-bit A/D, D/A converter, analog filters and amplifiers.

The examination of the coder requires digitization of the speech waveforms. High quality speech signal is passed through an appropriate lowpass fiter and then sampled at a rate of 8000 samples per second. The digitized speech signal is then stored on a computer disk for use as the input to the coder. The coder processes the digital waveform in non-real-time and the output is stored on a disk. The performance of the system is assessed by retrieving the input and the output signals and playing them back in real-time.

4.2 - Coder Implementation

Several half-band filters were designed using filter design programs. The filters were linear phase (i.e. the filter coefficients are symmetric) and their orders ranged from a minimum of 22 to a maximum of 64 taps* .

As mentioned earlier, Adaptive Pulse Coded Modulation (APCM) was selected as the quantization strategy. To prevent the allocation of too many levels to a single band, it was decided to limit the number of bits assigned to each sub-band to 5. Allocation of more bits to the same sub-band will in effect "rob" the other bands of quantization levels, diminishing the overall subjective quality of the output.

At this point it was necessary to find proper multipliers for 2-bit through 5-bit quantizers. These were determined experimentally and are listed in Appendix-B.

The uniform quantizers used in the APCM module were designed to be optimal (in the M.S.E. sense) for a unit variance input with a Gamma distribution

---

* A listing of the filter coefficients can be found in Appendix-B

[19]. The underlying assumption is that the APCM technique will normalize the input to these quantizers. This assures reasonable continuity across frame boundaries where changes in the number of quantization levels occur.

Although the transmission rate of interest in this work is 16 kbits/sec, some experiments were performed and data were produced using other rates. The quoted rates, however, do not include the extra bit-rate necessary for the transmission of side-information. The issue of side-information will be addressed separately.

To summarize the operation of the sub-band coder, the input signal (speech) is divided into frames. Each frame is passed through the QMF structure to produce the sub-band signals. The bit allocation algorithm determines the number of quantization levels for each of the sub-bands. The sub-band signals are then quantized by the appropriate APCM quantizers and transmitted to the receiver. At the receiver the quantized samples are decoded and put through the inverse filtering operation to reproduce speech.

The length of the frames of data was chosen to be 32 milliseconds. This length seems to be short enough to justify the implicit assumption of the stationarity of the input signal (the same frame-length is used in Linear Predictive Coding of speech), and long enough to require minimal overhead transmission of side-information (bit assignment information).

The coder was examined using 5 sentences from a list of phonetically balanced sentences, each spoken by a male and a female speaker.

4.3 - Experiments

A QMF tree structure with three stages was utilized to produce eight sub-bands each 500 Hz wide (recall that the sampling rate is 8000 samples/sec). The half-band filter used was a 36 tap filter with a stop-band rejection of 33 dB. The frequency response of this filter is given in Fig. 4.1. A trade-off is involved in the choice of the length of the filter; a longer filter can be designed to have a sharper roll-off and/or a higher stop-band rejection, resulting in reduced aliasing energy and hence better separation of the sub-bands. The price paid is increased complexity, and a longer filtering delay (the delay of a symmetric filter is proportional to its length). Given a QMF structure, the overall filtering delay is directly proportional to the length of the half-band filter. For a filter bank using 36 tap half-band filters this delay is 245 samples (at sampling rate of 8000 samples/sec the delay due to the filtering operation is about 30.6 milliseconds).

The overall frequency response of the system **without** quantization is presented in Fig. 4.2 . The amplitude of the ripples is less than 0.5 dB. Table 4.1 gives the SNR and the segmental SNR (SSNR) of the test utterances processed by the coder without quantization. Listening tests show that it is impossible to distinguish the input speech signal with the speech produced by the system without quantization.

The test utterances were processed by the coder. Fig. 4.3 displays the spectrograms of one of the input signals and the output speech. The SNR and the SSNR figures are given in Table 4.2 .

It is observed from Table 4.2 that for a given speaker the SNR values vary

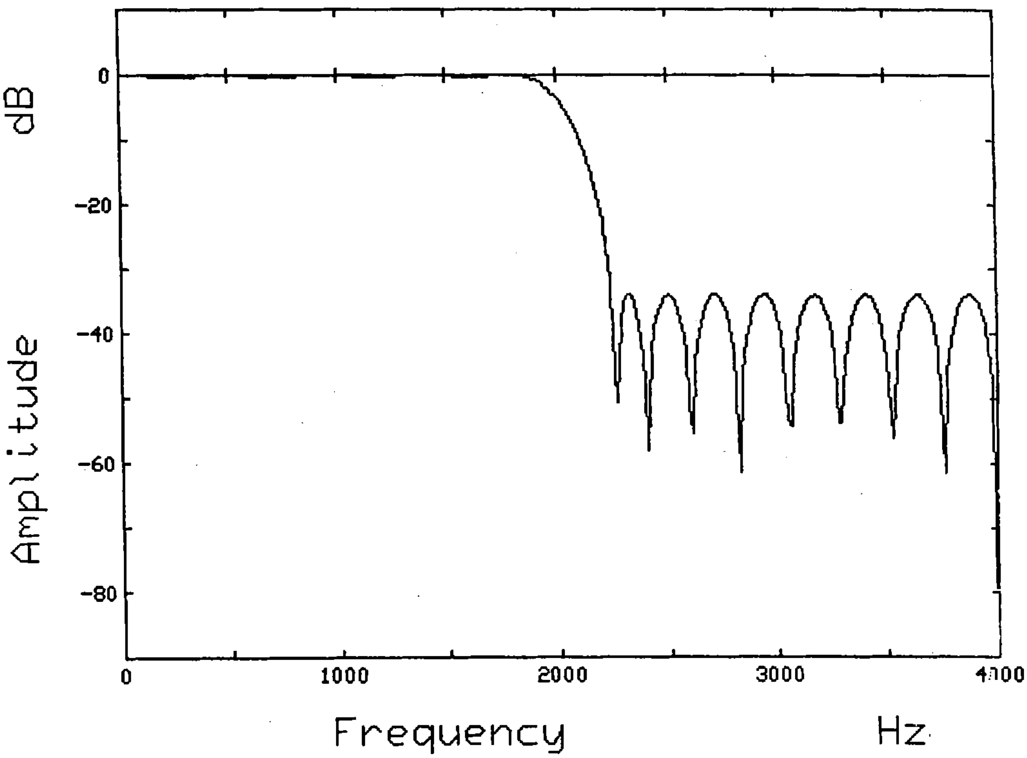

Figure 4.1 - Frequency response of the 36-tap half-band filter.

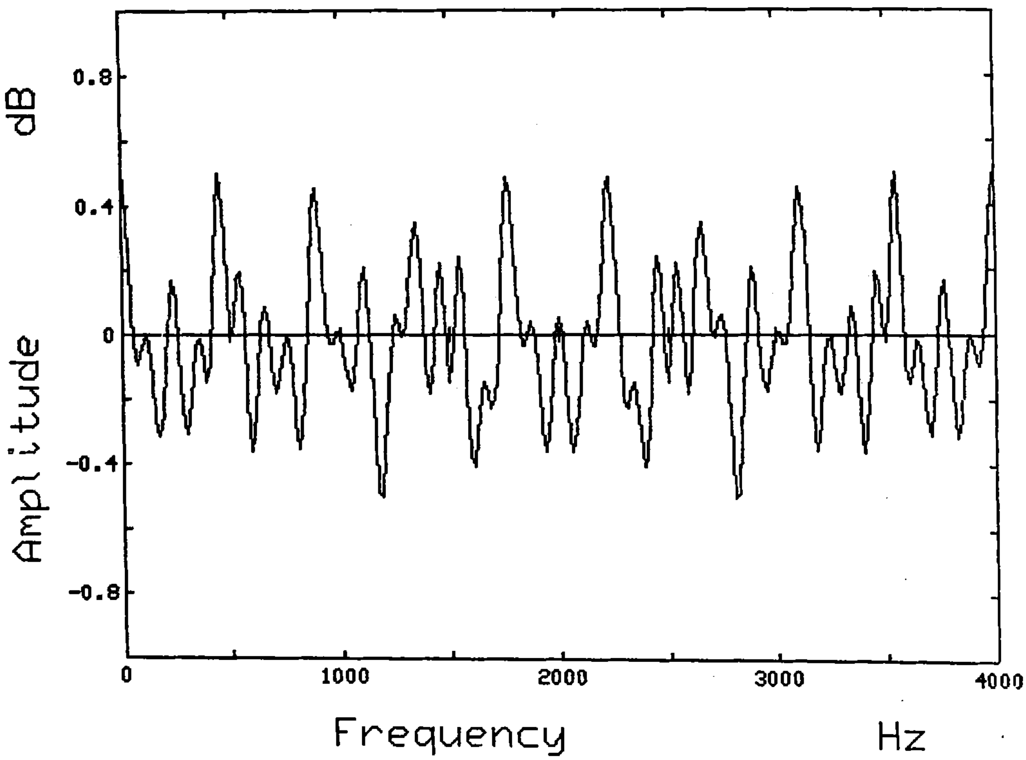

Figure 4.2 - Overall frequency response of the coder without quantization.

sentence	speaker	
	female	male
A	26.766/26.107	26.598/25.499
B	26.350/26.564	27.450/25.270
C	25.875/25.900	24.745/24.841
D	27.779/26.247	24.852/24.880
E	27.349/26.433	25.926/26.654

Sentence A:	"Add the sum to the product of these three."
Sentence B:	"Cats and dogs each hate the other."
Sentence C:	"Oak is strong and also gives shade."
Sentence D:	"Open the crate but don't break the glass."
Sentence E:	"Thieves who rob friends deserve jail."

Table 4.1 - SNR (dB)/seg.SNR (dB) values of the coder without quantization (using 36-tap half-band filters).

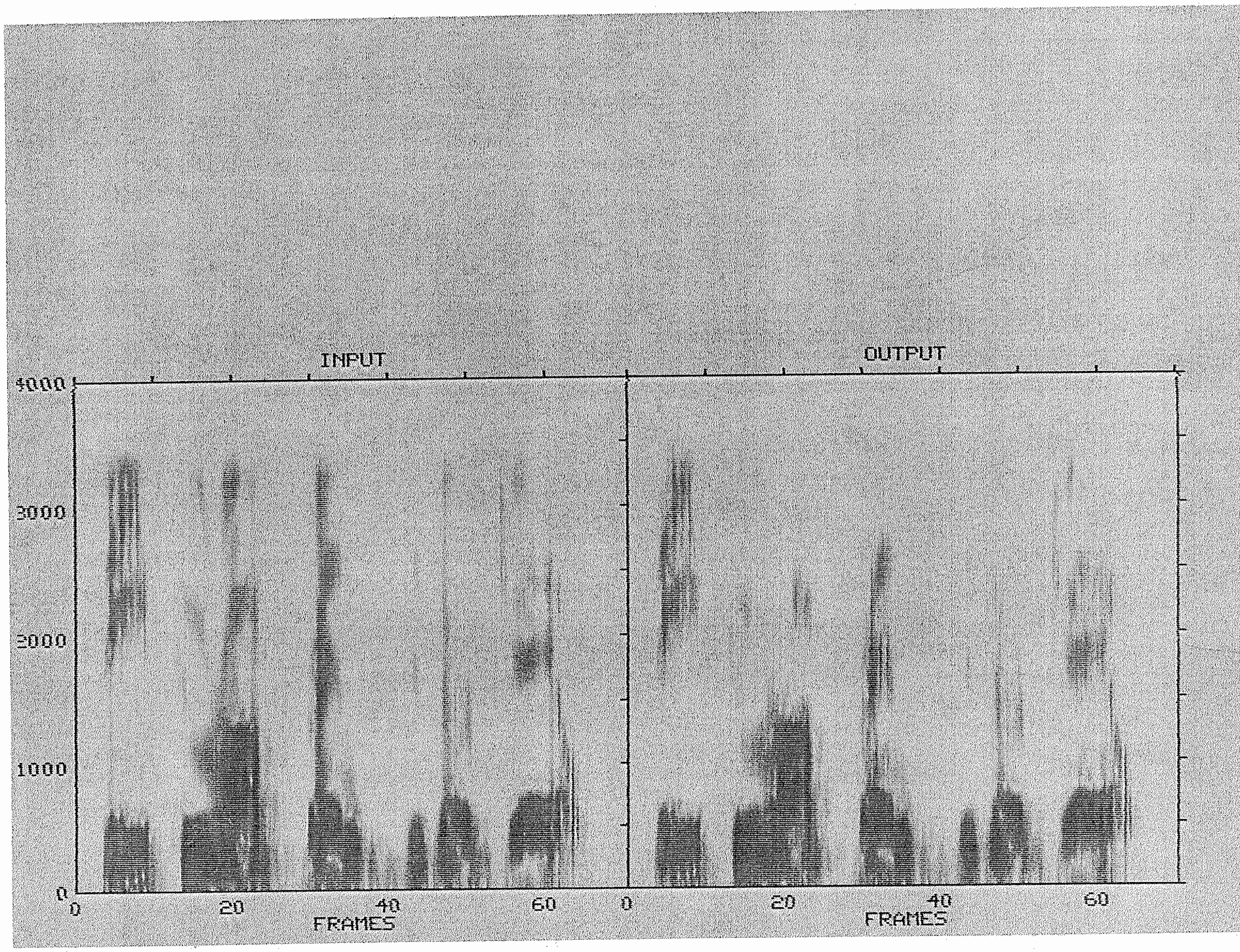

Figure 4.3 - Spectrograms of the input and the output of the coder.

sentence	speaker	
	female	male
A	17.305/15.825	11.911/14.617
B	15.373/15.500	12.326/13.872
C	19.608/16.645	16.204/15.078
D	18.883/16.300	15.925/14.812
E	19.385/16.723	13.786/15.788

Sentence A: "Add the sum to the product of these three."

Sentence B: "Cats and dogs each hate the other."

Sentence C: "Oak is strong and also gives shade."

Sentence D: "Open the crate but don't break the glass."

Sentence E: "Thieves who rob friends deserve jail."

Table 4.2 - SNR (dB)/seg.SNR (dB) values of reconstructed speech.

widely from sentence to sentence, while the SSNR seems to be more stable. The reason for this is that the signal to noise ratio is influenced more by high amplitude segments of the signal than by the lower energy segments. Therefore, the SNR figure mainly reflects the performance of the coder during the high amplitude segments of speech. The segmental SNR, on the other hand, is the average of the SNR of short segments of the signal and is therefore a more reliable measure of the coder performance. Accordingly, our conclusions will be based mainly on the segmental SNR rather than the SNR figures.

Theoretically, the spectrum of quantization noise is expected to be flat. While due to practical constraints, such as the integer bit assignment requirement, this is not expected to be exactly fulfilled, the average across the full utterance of the error spectrum should be flat. Fig. 4.4 shows the spectrogram of a processed test utterance and its average spectrum. The spectrogram of the corresponding error signal and its average spectrum are shown in Fig. 4.5. Note that with the exception of the first two sub-bands, the distortion is roughly flat. The deviation from flatness in the first two sub-bands is caused by the clamping of the allocated bits to a maximum of 5. This explanation has been verified by reducing the transmission rate, whereby the "claim" of the sub-bands to the bit resource may be lowered to below the clamping level.

Fig. 4.6 shows the spectrogram and the average power spectrum of the error due to the processing of the same input speech at the rate of 10 kbits/sec. At this rate the average power spectrum of the error is roughly flat with the exception of the higher bands whose energies are lower than the "critical" distortion level.

Now we focus our attention on the quality of the reconstructed speech at

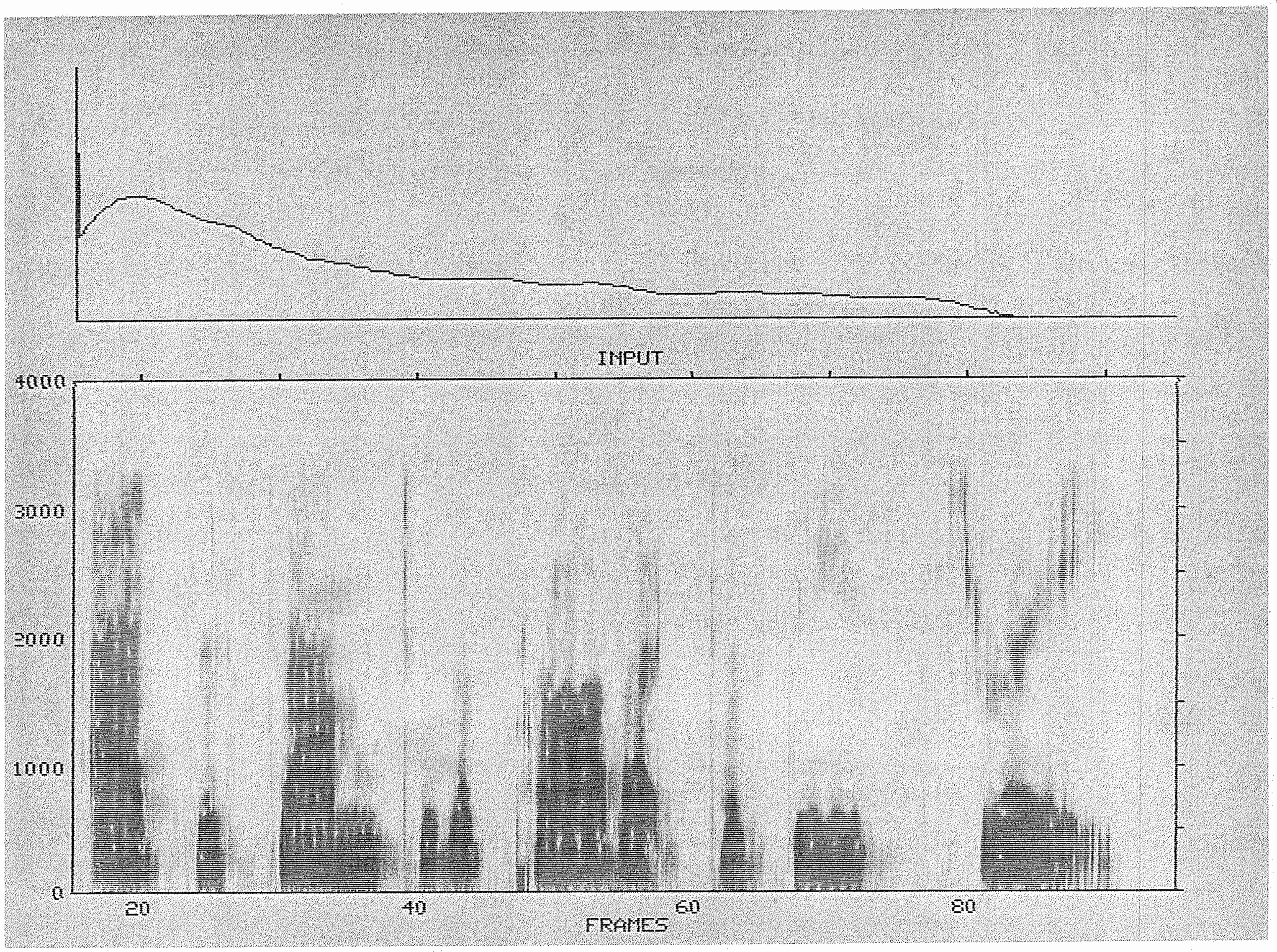

Figure 4.4 - Spectrogram of a sentence and its average power spectrum.

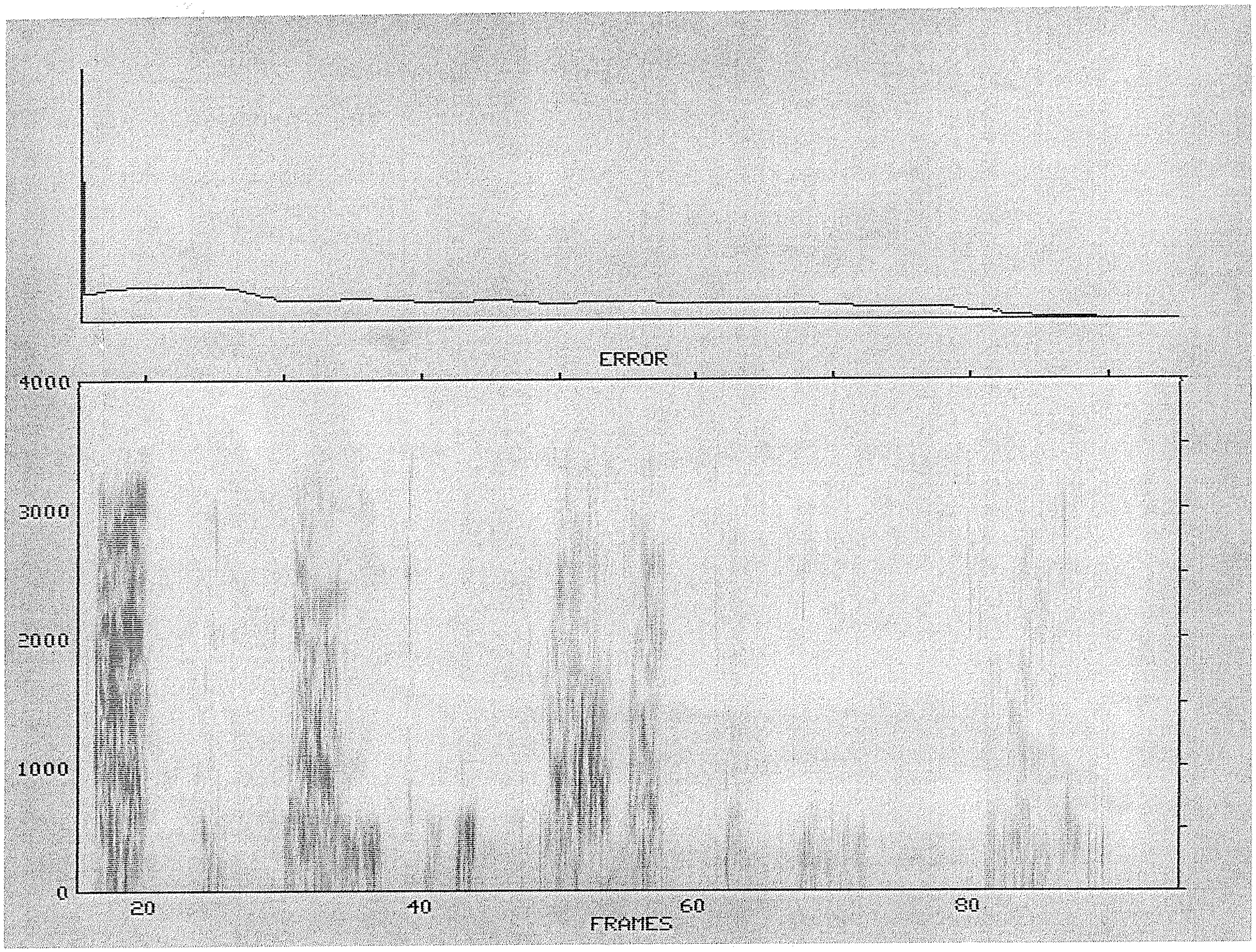

Figure 4.5 - Spectrogram of the error waveform and its average power spectrum.

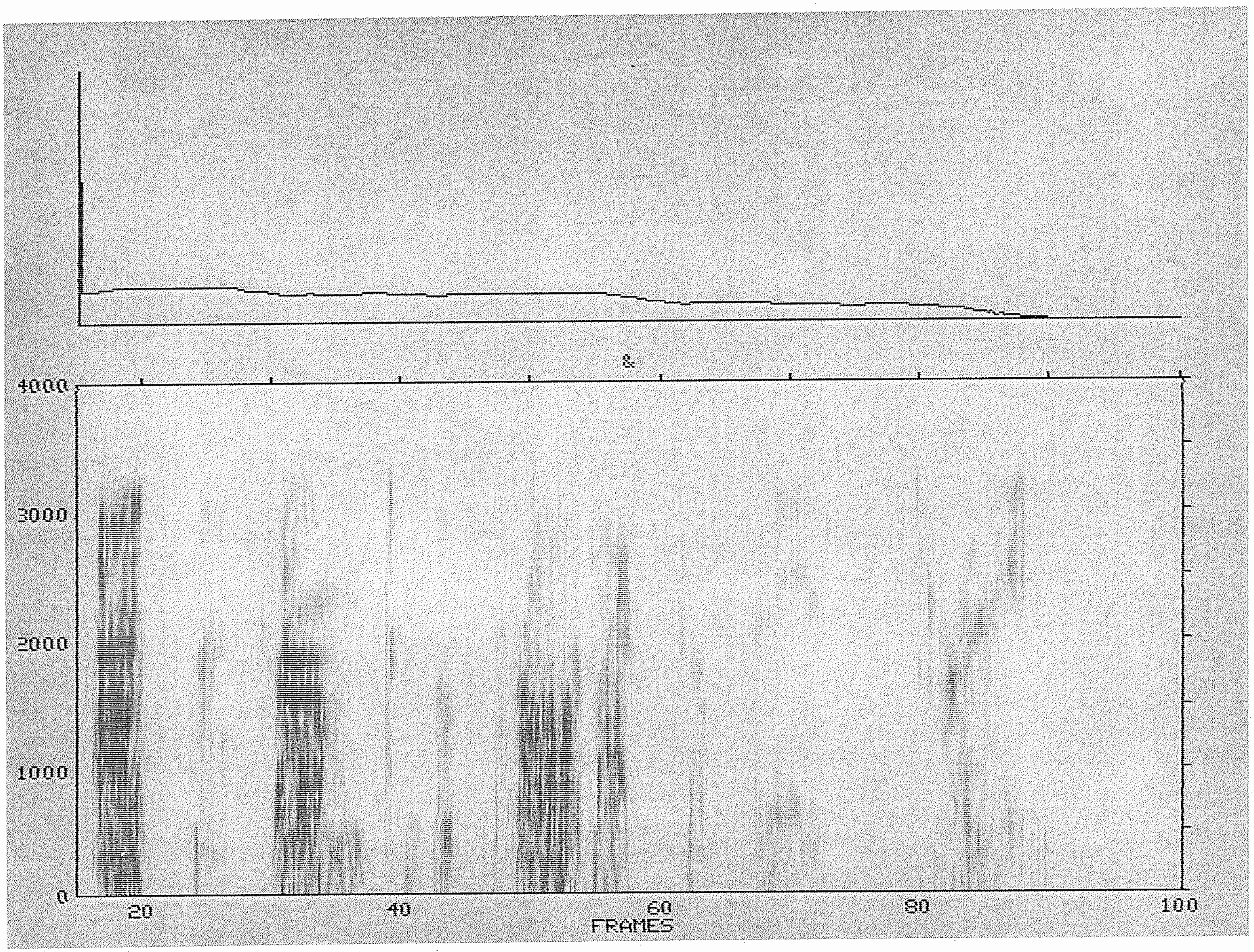

Figure 4.6 - Spectrogram of the error waveform and its average power spectrum at the rate of 10 Kbits/sec.

the transmission rate of 16 kbits/sec.

At this stage, the output speech is fully intelligible and sounds natural. Listening tests show that two types of noise are perceived in the coded speech. One is a kind of burbling noise which makes the speech sound harsh. The other type can be described as high-pitch swishing and aliasing noise. The latter is the more annoying of the two.

An experiment was conducted to investigate the nature of the distortion produced by the quantization of each sub-band, and the amount of the influence of such noise on the overall quality of the output. In this test the same input signal was processed by the coder repeatedly, and each time one of the sub-band quantizers was removed. The experiment shows that the burbling is caused by the quantization of the first few sub-bands. Examining typical bit-assignment patterns (Table 4.3) we find that these bands are represented by a large number of quantization levels most of the time. So it is suspected that such noise may be the result of discontinuities caused by the changes in the number of quantization levels across frames.

More importantly, the experiment also shows that in the absence of quantization in the last two sub-bands (3000-4000 Hz) the aliasing noise is in effect eliminated. At first glance this seems to be odd; these bands are so low in energy that they are often not transmitted at all, or at best assigned very few bits (Table 4.3). To explain this phenomenon, recall that in a QMF structure aliasing is suppressed by cancellation. Therefore, if a sub-band is quantized crudely or not transmitted at all (assigned zero bits), it will have a direct effect on its neighboring bands since the aliasing components do not cancel.

Data frame	Sub-band: 1	2	3	4	5	6	7	8
80	3	0	1	2	2	4	4	0
81	1	0	0	3	5	4	3	0
82	5	3	1	3	3	1	0	0
83	5	4	3	3	1	0	0	0
84	5	5	1	3	2	0	0	0
85	5	5	1	2	2	1	0	0
86	5	5	1	2	2	1	0	0
87	5	4	0	0	3	3	1	0
88	5	2	0	0	3	4	2	0
89	5	2	0	0	2	4	3	0
90	5	2	0	0	2	4	3	0
91	5	3	1	1	0	3	2	1
92	5	3	1	0	1	3	3	0
93	5	3	1	1	2	3	1	0

Table 4.3 - The bit assignment pattern for the utterance "three" spoken by a female speaker.

The spectrogram of the word "three" from one of the test sentences is displayed in Fig. 4.7 together with the spectrogram of the coded version of the same word. Noise is visible in the 3400-4000 Hz range of the spectrum. A higher resolution spectrogram of the same pair (Fig. 4.8) shows that the noise is aliasing, mainly due to leakage from the lowest band (0-500 Hz) into the highest (note that the noise is an attenuated mirror image of the energy in the lower bands).

To understand how this happens one must carefully examine the filtering operation. The QMF tree structure can be replaced by eight parallel filters operating on the input signal, followed by 8:1 sub-sampling. The coefficients of each of these filters can be computed by successive convolution of the tree-structure half-band filters in the appropriate paths, taking into account the decimation operations.

The equivalent parallel filters producing the first four sub-bands were computed using the 36-tap half-band filter mentioned earlier. The frequency response of these filters are displayed in Fig. 4.9. The frequency response of the filters producing the last four sub-bands is the mirror image of that of the first four.

The frequency response of the filter producing the highest band (mirror image of Fig. 4.9a) shows that the rejection of the lowest band (0-500 Hz) is around 35 dB. Since the energy in the 3.5-4 kHz band is at times up to 70 dB below the energy in the 0-500 Hz band, the 8:1 sub-sampling results in great distortion of the higher band by aliasing. Normally this would not degrade the quality of the output speech, since the QMF structure provides for the elimination of such noise at the reconstruction stage. However, in case of coarse quantization, or, as in the case at hand, complete loss of a sub-band signal, the aliasing remains.

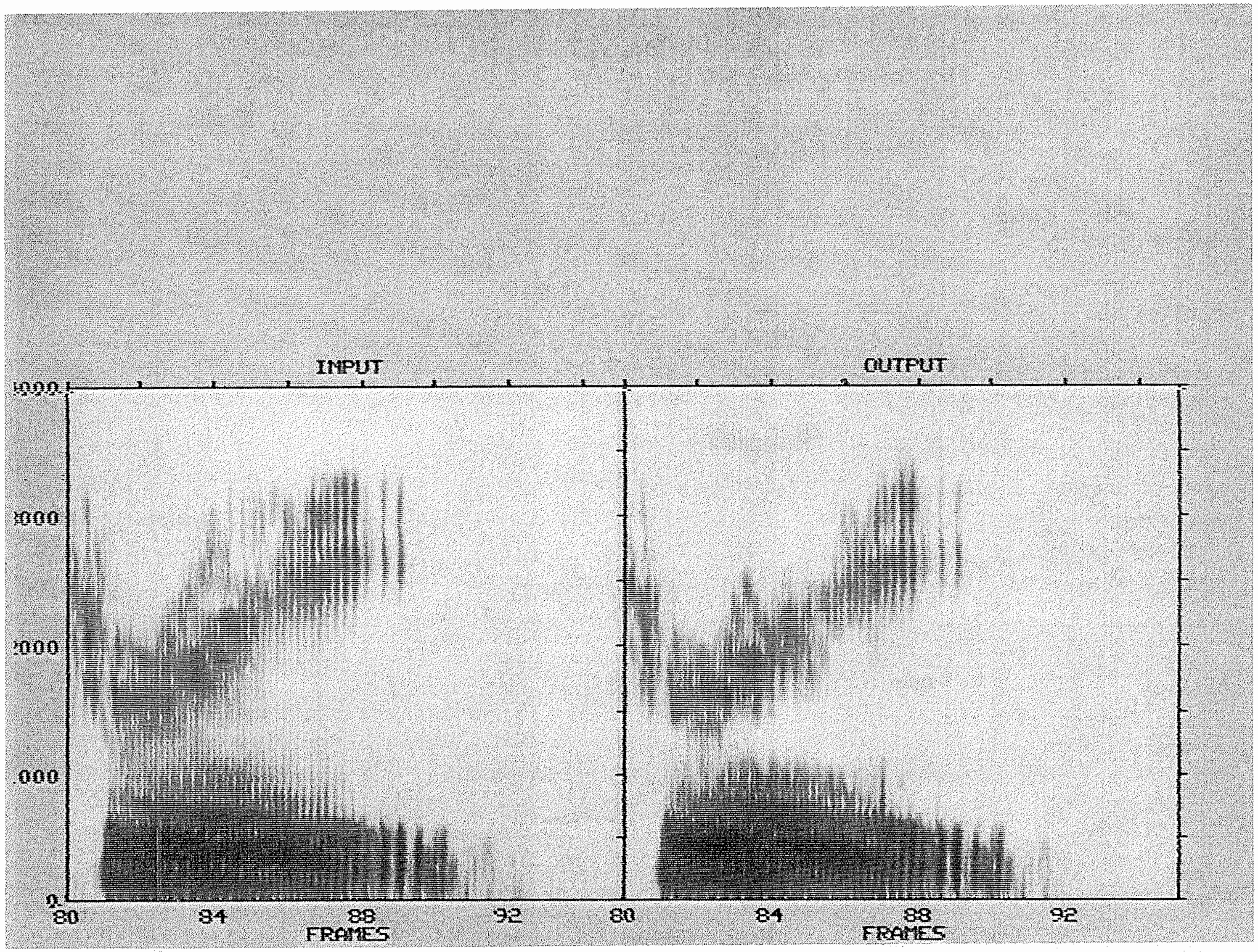

Figure 4.7 - Spectrogram of the word "three" spoken by a female speaker and its coded version.

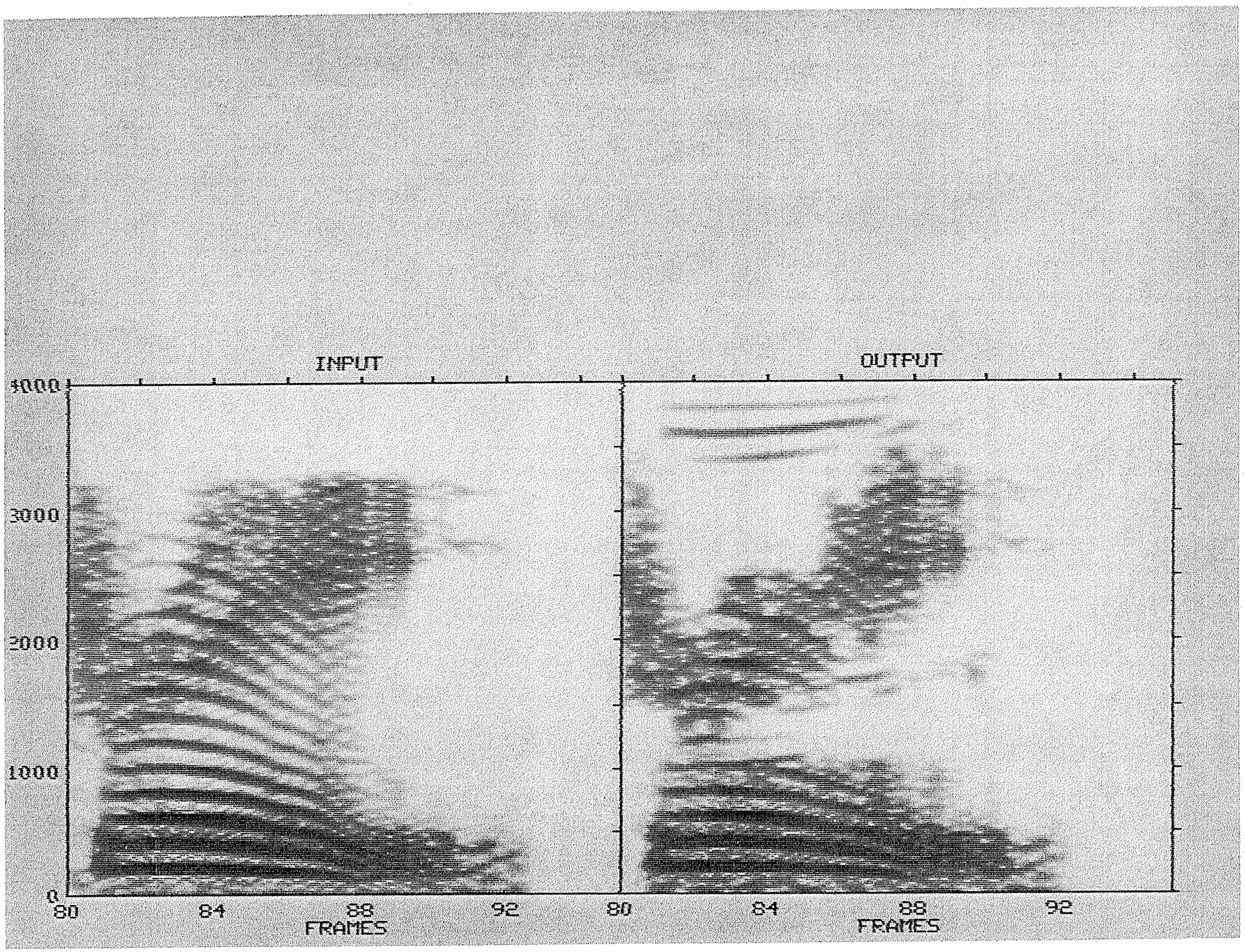

Figure 4.8 - High-resolution spectrogram of the word “three” spoken by a female speaker and its coded version.

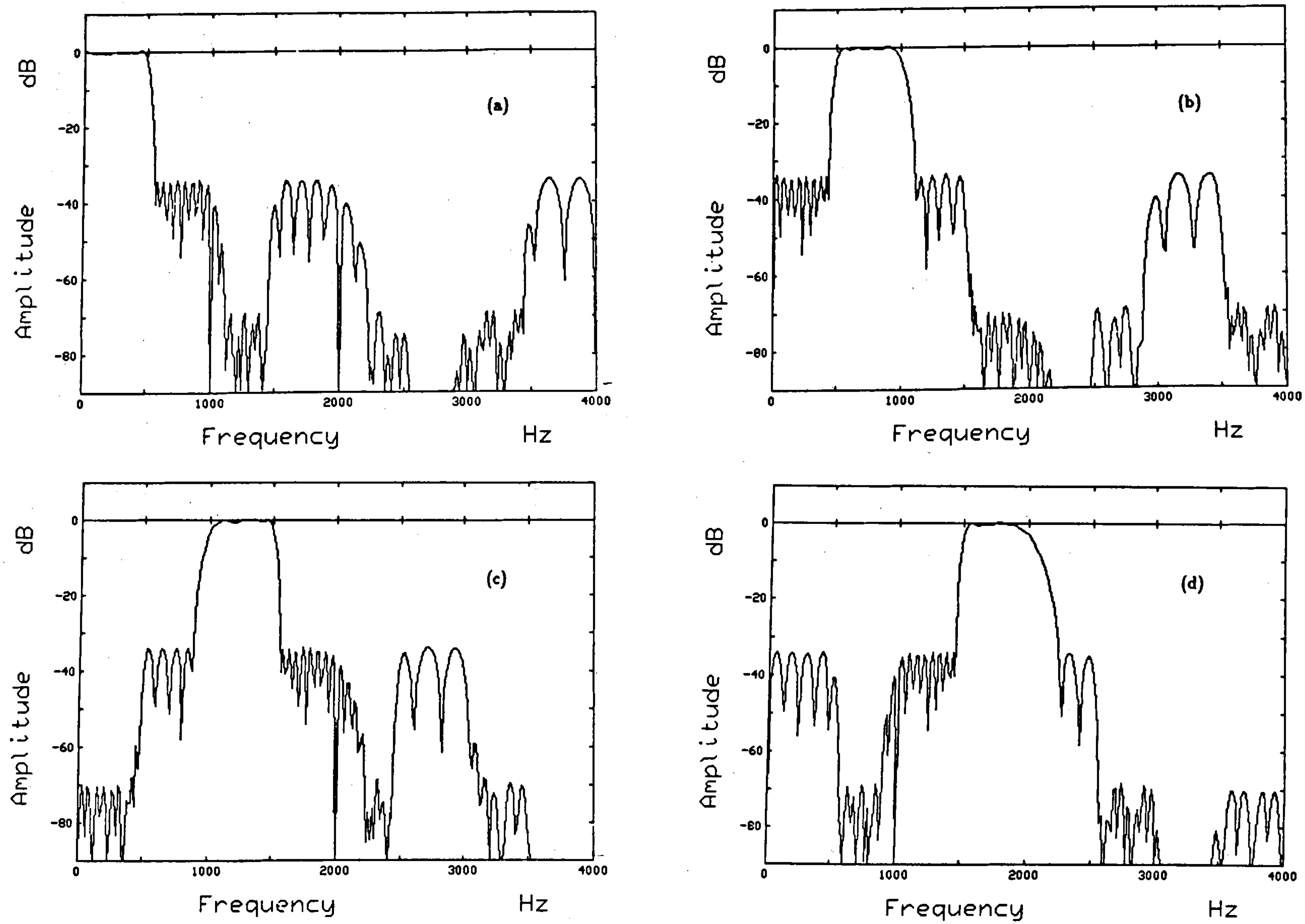

Figure 4.9 - Frequency response of parallel filters derived from the QMF tree structure :
(a) 0-500 Hz;(b) 500-1000 Hz;(c) 1000-1500 Hz;(d) 1500-2000 Hz

In order to examine the validity of these speculations, several sustained vowels were processed by the coder. The input vowels are "UH" as in "but", "U" as in foot and "IY" as in "beet". Fig. 4.10 shows the spectrogram of the input signals together with their average power spectrum.

Listening tests show that while the aliasing noise is present in all cases, it is heard most distinctly in the reconstructed "U" sound. In the case of the vowel "IY" the noise is least perceivable.

From the average power spectrum of the input (Fig. 4.10) it is observed that the differences in the energies of the 0-500 Hz band and the 3.5-4 kHz band are roughly 38 dB, 45dB and 60dB for the vowels "IY", "UH" and "U", respectively. Note that the higher the difference between the energy levels of the two bands, the more audible the aliasing noise. This is in agreement with the earlier conclusions. The spectrograms of the processed vowels and their average power spectrums are shown in Fig. 4.11.

4.4 - Elimination of the Aliasing Noise

In this section ways of reducing the aliasing noise are considered. Three of the more simple solutions are the following :

1- Better quantization of the highest band: At the transmission rate of 16 kbits/sec it is impossible to just allocate more bits to the highest sub-band without seriously compromising the quality of the coder. One possibility is to pass the input speech through a bandpass filter to eliminate the energy in the 0-250 Hz range. Then

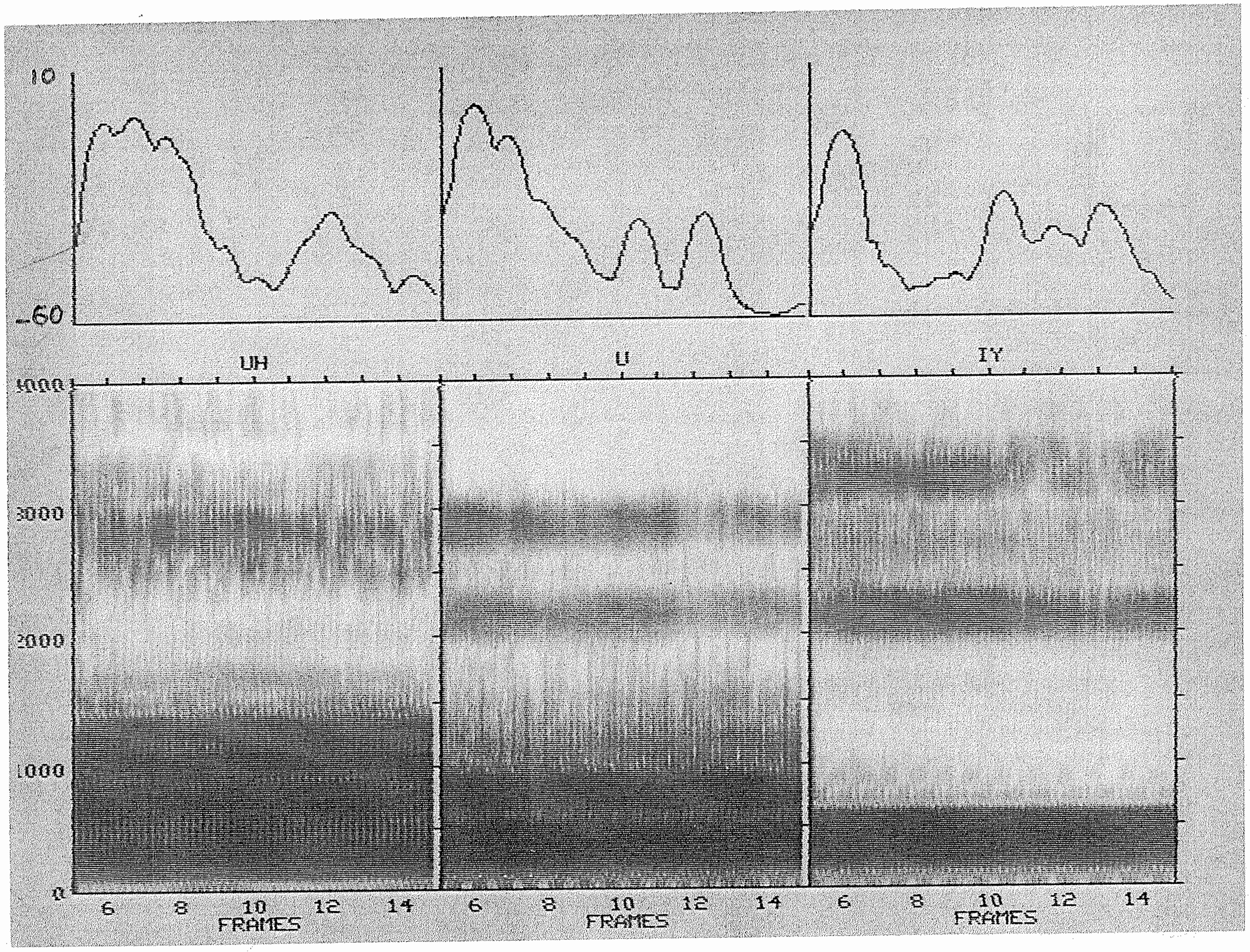

Figure 4.10 - Spectrogram of sustained vowels and their average power spectrum.

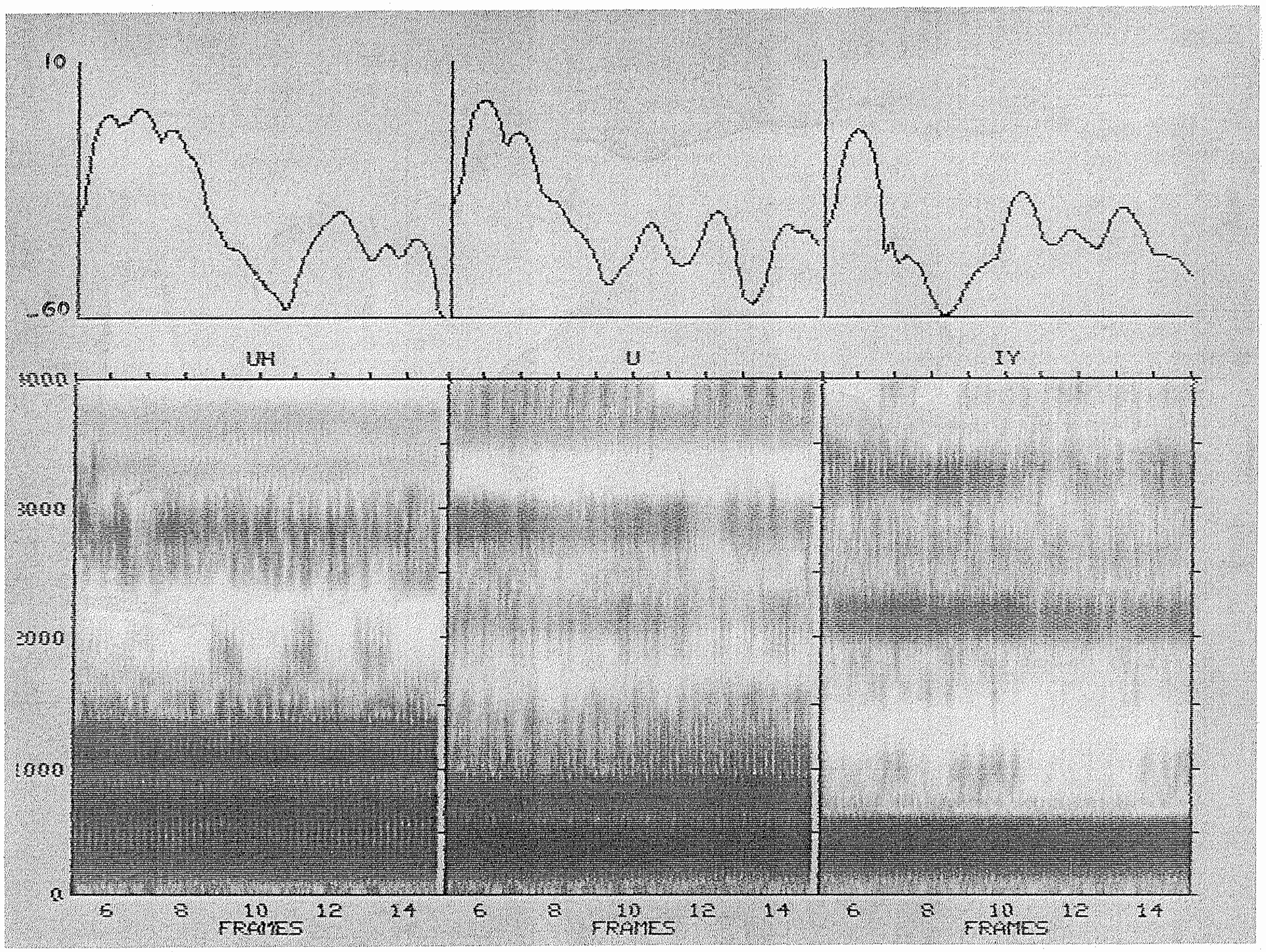

Figure 4.11 - Spectrogram of sustained vowels (processed by the coder) and their average power spectrum.

the lowest sub-band in the QMF structure (0-500 Hz) can be further divided in two, the 0-250 Hz band may be discarded and the 250-500 Hz band can be quantized at a reduced rate. This can free up to 2 bits per frame, which may now be assigned, permanently, to the highest band.

Such a coder was implemented and used to process the test sentences. Listening tests showed that although the aliasing was reduced, the overall quality of the reconstructed speech had deteriorated; it sounded harsher and somewhat unnatural. This effect is due to the discarding of the 0-250 Hz band and the aliasing caused by the loss of its signal.

The obvious disadvantage of this scheme is that it is necessary to add another filtering stage to perform the sub-division of the 0-500 Hz band, resulting in the doubling of the filtering delay, not to mention the added complexity.

2- Higher order filters: Higher order half-band filters can be used in the basic QMF structure to achieve better separation of the sub-bands, in order to reduce the inter-dependence of the sub-band signals. However, depending on the order of the new filters, the complexity, delay and memory requirements could be increased drastically.

3- Use of different order filters in different stages: A look at the average power spectrum of speech shows that while the energy level in adjacent bands does not differ greatly, the difference is much larger when the bands in the lower frequency range are compared to those in the higher end of the spectrum.

The idea in using different order filters for different stages is to free the higher bands from the leakage of aliasing energy from the lower ones, through the use of a large order filter with higher stop-band rejection in the first filtering stage. Lower order filters may be used in the subsequent stages, where the sub-band energies are of comparable magnitudes.

This seems to be a simple and inexpensive way of eliminating the elias-ing noise; the complexity is not expected to increase by much since the effect of employing a higher order filter at the first stage may be offset by the reduction of the order of the filters in the subsequent stages. Also note that because of the 2:1 sub-sampling in every stage of a QMF structure, the filtering delay due to a given filter is half the delay of the same filter employed in the next stage. Therefore, the use of a large filter followed by two short filters could in fact reduce the overall filtering delay of the coder.

This scheme was implemented using a 64-tap filter in the first stage and a 24-tap filter in the next two stages. The frequency response of these filters and the overall frequency response of the coder (excluding the quantizers) is illustrated in figures 4.12 to 4.14, respectively. The overall filtering delay due to this configuration is 201 samples (at the sampling rate of 8000 samples/sec this is equivalent to about 25 milliseconds).

The frequency response of the equivalent parallel filters producing the first four sub-bands is given in Fig. 4.15. note that in the case of the filter which produces the highest sub-band (mirror image of Fig. 4.15a) the rejection of the 0-500 Hz band has been increased from 35 dB to over 80 dB.

The spectrogram of the word "three" and its processed version are again displayed in Fig. 4.16, followed by a high resolution spectrogram in Fig. 4.17. Note

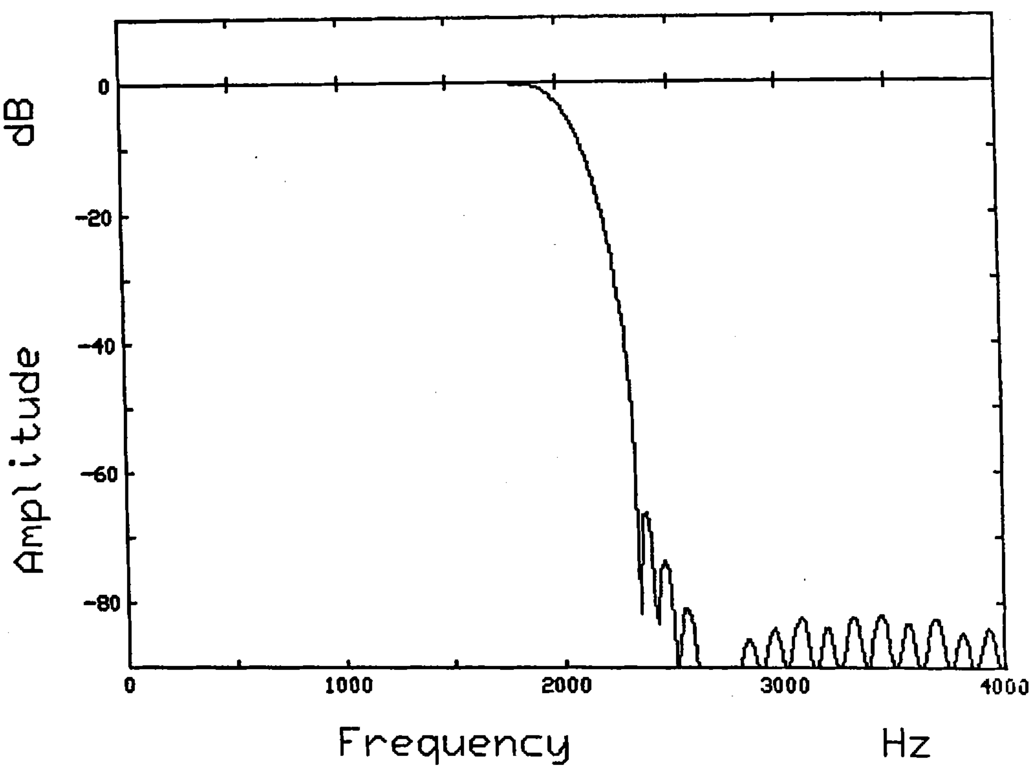

Figure 4.12 - Frequency response of the 64-tap half-band filter.

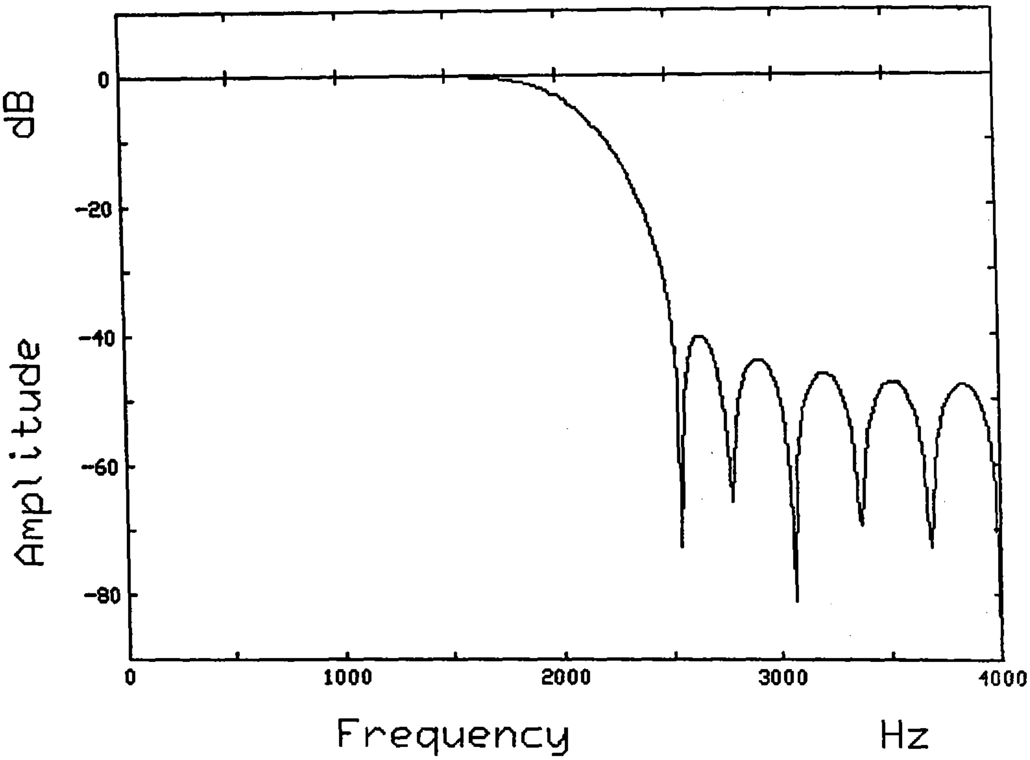

Figure 4.13 - Frequency response of the 24-tap half-band filter.

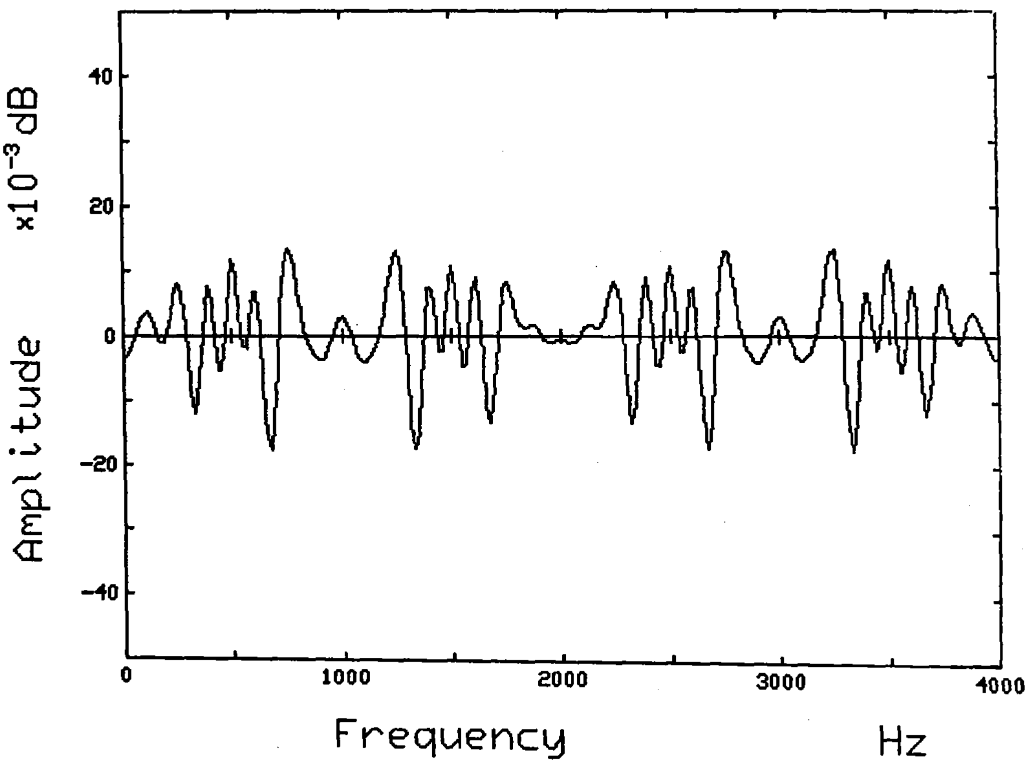

Figure 4.14 - Overall frequency response of the coder without quantization.

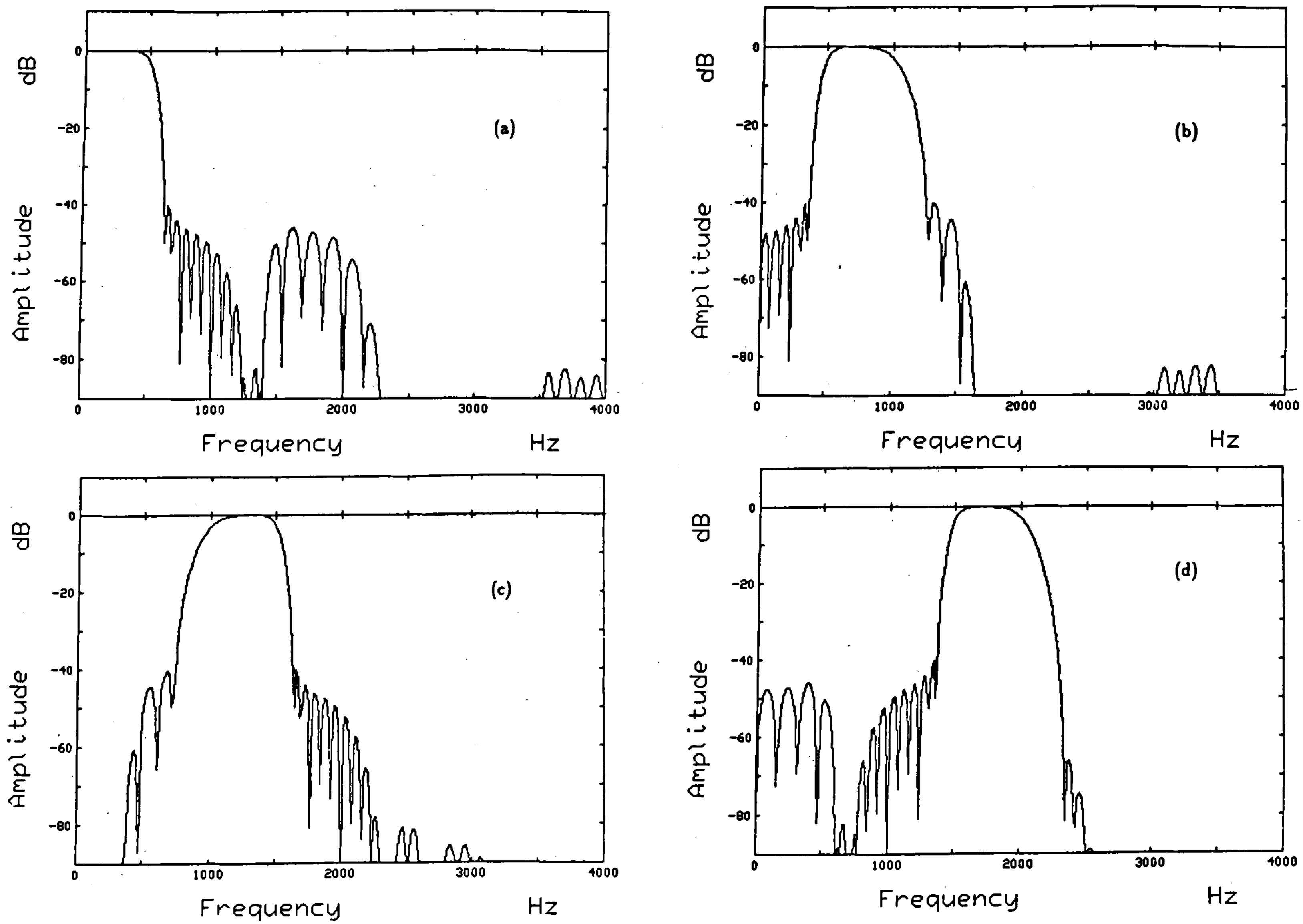

Figure 4.15 - Frequency response of parallel filters derived from the QMF tree structure : (a) 0-500 Hz;(b) 500-1000 Hz;(c) 1000-1500 Hz;(d) 1500-2000 Hz

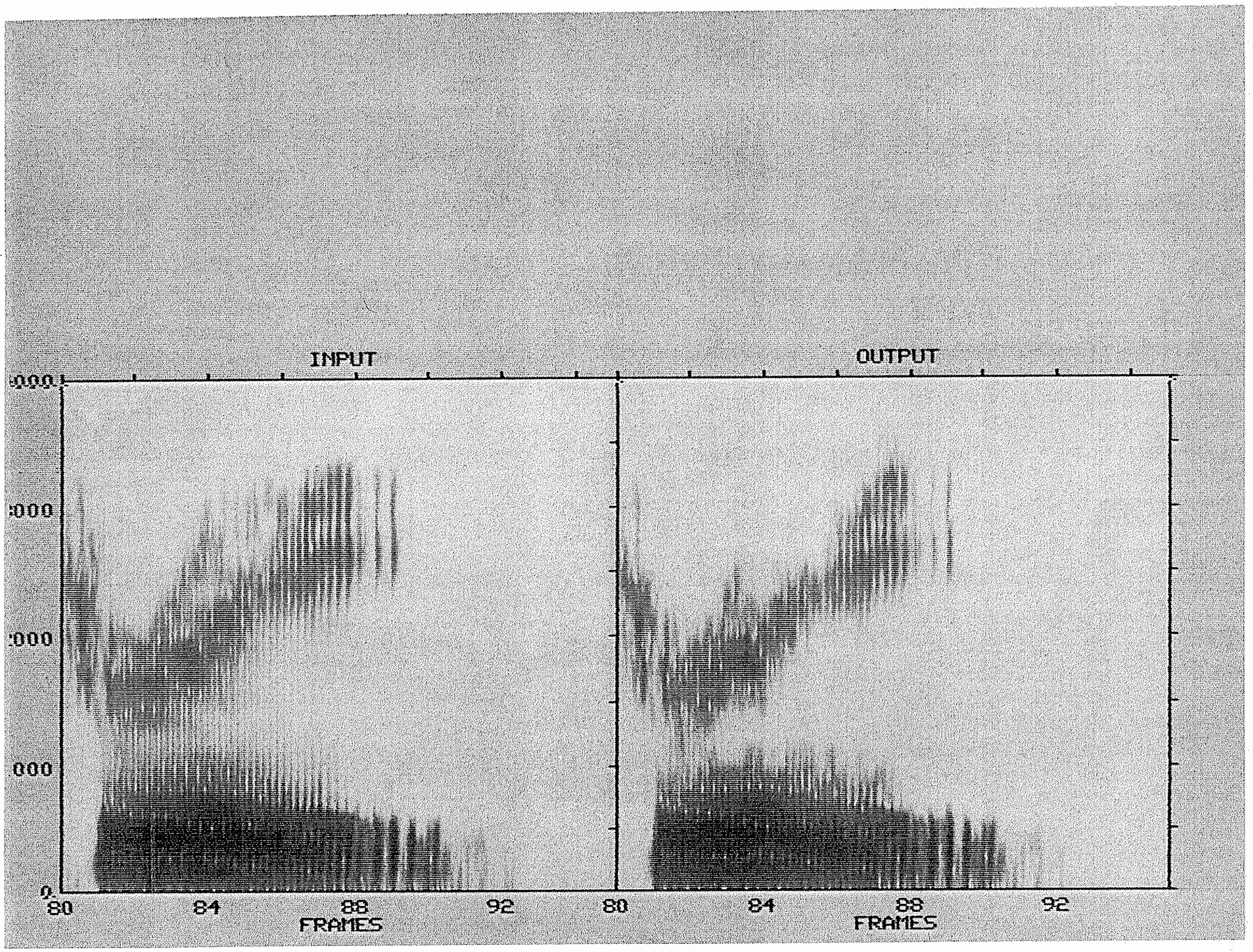

Figure 4.16 - Spectrogram of the word "three" spoken by a female speaker and its coded version.

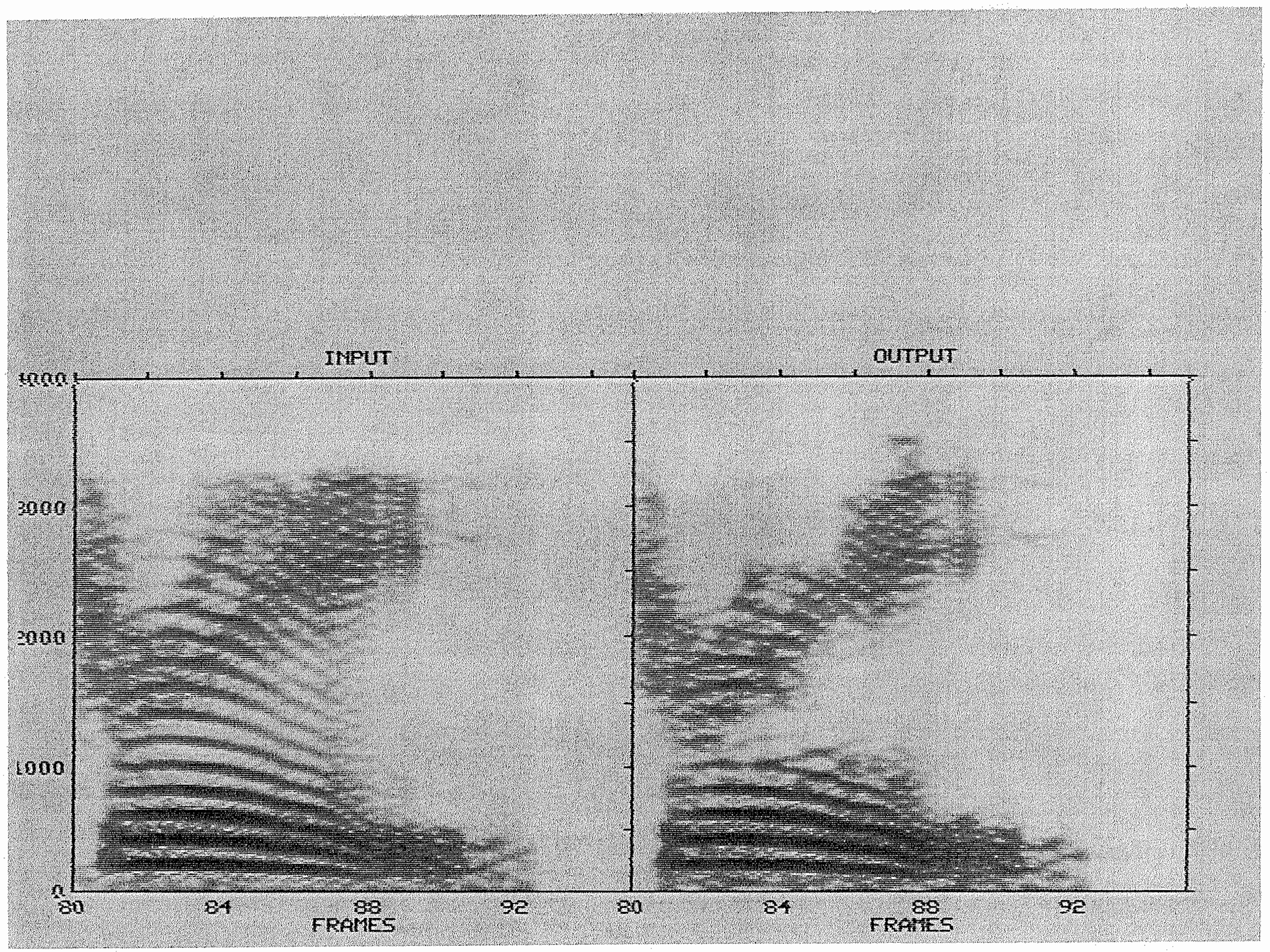

Figure 4.17 - High-resolution spectrogram of the word "three" spoken by a female speaker and its coded version.

that the aliasing has been eliminated, although some noise is still present.

The SNR figures are given in Table 4.4. Some improvement is achieved in the segmental SNR values. The increase in the SSNR values is small because the elimination of the aliasing noise affects the low energy bands only.

Subjective tests also confirm the elimination of the aliasing noise. However, annoying high-pitch noise can still be heard. This noise sounds like high frequency tones in the range of 3 to 3.5 kHz and is visible in the spectrograms of Fig. 4.16 and Fig. 4.17.

Looking at the bit assignment patterns corresponding to the regions where the high frequency noise is present, it was observed that one or both of the two highest bands were quantized with 1-bit quantizers. A series of tests were carried out to assess the value and the effect of the 1-bit quantizers. These are described in the following section.

4.5 - One-bit quantizer

The results of the tests discussed in the previous section pointed to the possibility of adverse effects by the 1-bit quantizers. To some extent this is to be expected; the only information provided by the 1-bit quantizer output is the sign of the input sample. No information is available on the amplitude of the input sample and since no step-size adjustment is possible, the step-size and hence the amplitude of the output remains at the level reached in the previous frame of data. Modifications can be made to provide for step-size adjustment in 1-bit quantizers.

sentence	speaker	
	female	male
A	16.855/16.369	11.642/15.176
B	15.288/15.776	11.901/14.460
C	19.658/17.271	15.626/16.075
D	18.809/16.700	16.166/15.338
E	18.802/17.317	14.449/16.443

Sentence A:	"Add the sum to the product of these three."
Sentence B:	"Cats and dogs each hate the other."
Sentence C:	"Oak is strong and also gives shade."
Sentence D:	"Open the crate but don't break the glass."
Sentence E:	"Thieves who rob friends deserve jail."

Table 4.4 - SNR (dB)/seg.SNR (dB) values of the coder with a 64-tap filter in the first stage and 24-tap filters in the subsequent stages.

One such scheme is known as the $1+1/k$-bit quantization [21] and is based on sending the amplitude information once every k samples, and the transmission of the sign of the input for the rest of the input samples. The step-size adjustment is based on the output amplitude level and therefore takes place every k^{th} sample. However, for short frames of data even such a scheme is not expected to be adequate since the step-size adjustment may not be fast enough to converge to the right range.

To shed light on the role of the 1-bit quantizers, the sub-band coder was tested in the following configurations:

a- Omission of the 1-bit quantizers:

In the first experiment the 1-bit quantizers were bypassed, i.e. no transmission of data took place in the case of the sub-bands which were assigned a 1-bit quantizer.

The output speech sounds somewhat harsher and less natural, because of the spectral nulls. However, the annoying high-pitch noise is reduced drastically and the overall quality of the output is preferable to that of the coder in the normal configuration.

b- Optimum step-size:

In another configuration the step-size of the 1-bit quantizers was set to the optimum (in the M.S.E. sense) for the given frame of data. Note that in practice extra information must be provided to adjust the step-size at the receiver and at low rates of transmission this may not be affordable.

Listening tests show that the reconstructed speech still features the high-pitch noise although it is somewhat reduced.

c- $1 + 1/2$-bit quantization:

In this scheme the function of the 1-bit quantizer is to code the amplitude and the sign of the input sample alternately.

Since the purpose of the $1 + 1/2$-bit quantization is to provide a way of adjusting the step-size towards its optimal value, this scheme is not expected to outperform quantization with optimum step-size (the previous configuration). This speculation is confirmed by subjective tests.

The experiments above lead us to believe that perceptually, the distortion caused by a 1-bit quantizer outweighs its benefit -filling the spectral gaps. The scheme involving optimum step-size quantizer is perceptually comparable to the one which omits the 1-bit quantizers, but the extra bit-rate necessary for the transmission of the step-size information is prohibitively high.

Statistics of the bit assignment patterns show that single bit assignments constitute an average of over 1.65 kbits/sec, i.e. over 10% of the overall transmission rate. To make better use of this fraction of the capacity it was decided to re-distribute the single bits in the following manner:

The bands which are allocated one bit each are singled out and sorted in the descending order of highest residual claim to the bit resource. Then the bit assigned to the sub-band with the lowest claim is given to the one with the highest. This procedure is repeated for the remaining bands. Thus the single bits are combined in pairs in favor of the sub-band with the highest residual claim. Notice that according to this scheme, if the number of sub-bands originally assigned one bit is odd, in the end a single band will remain with a 1-bit quantizer. Such a band is coded with a $1 + 1/2$-bit quantizer.

The SNR figures corresponding to this scheme are given in Table 4.5. A slight but definite improvement is observed in comparison with the results of Table 4.4. Subjectively the difference is noticable; the reconstructed speech is less noisy and sounds smoother. However, the noise is not eliminated completely and some burbling and swishing sounds are still audible. These are suspected to be caused by the shifting of the bits from frame to frame.

Listening tests involving sustained vowels also show some improvement in the quality of the coder output although an adverse effect is detected as well; some vowels, notably "OW" as in "bought" and "AE" as in "bat" become nasalized, undoubtedly due to the spectral nulls caused by the elimination of the 1-bit quantizers.

4.6 - Further Improvements in the Quality of the Coder

In this section we discuss modifications to the coder in order to improve the quality of the reconstructed speech. These modifications include the introduction of perceptual criteria as well as better quantization of the sub-band signals.

sentence	speaker	
	female	male
A	17.206/16.839	11.989/15.583
B	16.196/16.708	11.702/14.889
C	20.364/18.093	15.903/16.510
D	19.511/17.204	16.381/16.112
E	20.321/17.854	14.577/17.070

Sentence A:	"Add the sum to the product of these three."
Sentence B:	"Cats and dogs each hate the other."
Sentence C:	"Oak is strong and also gives shade."
Sentence D:	"Open the crate but don't break the glass."
Sentence E:	"Thieves who rob friends deserve jail."

Table 4.5 - SNR (dB)/seg.SNR (dB) values of the coder with the 1-bit quatizers combined.

4.6.1 - Non-uniform Quantization

The results so far were obtained using uniform quantizers inside the APCM module. The reason for this is that due to the large number of parameters involved, the analysis of APCM is a difficult task and hence optimal quantizers are not easy to design. Due to the lack of knowledge of the best quantizer, it was decided to use uniform quantizers since those are the easiest to implement (at least as far as software implementation is concerned).

It is possible to gain some insight into the operation of the quantizer block by examining the input to the quantizer inside the APCM module (Fig. 4.18). The quantizer should be designed based on the statistics of its input. However, this process must be applied iteratively since when the quantizer is changed, the step-size adaptation alters the signal statistics. In practice if a quantizer is reasonable, its replacement by the one designed to be optimal for the measured statistics does not appreciably affect the statistics of the input signal.

The histogram of the "real" quantizer input for 2-bit through 5-bit quantizers is given in Fig. 4.19. These histograms are given for the first sub-band. The histograms based on other sub-bands essentially display the same shape.

A point worth mentioning here is that these histograms also provide some information about the choice of the APCM multipliers. Optimal multipliers are expected to normalize the input to the quantizer inside the APCM module. Note that the standard deviation of each of the histograms is roughly equal to unity. We therefore conclude that the multipliers are not far off their optimal values.

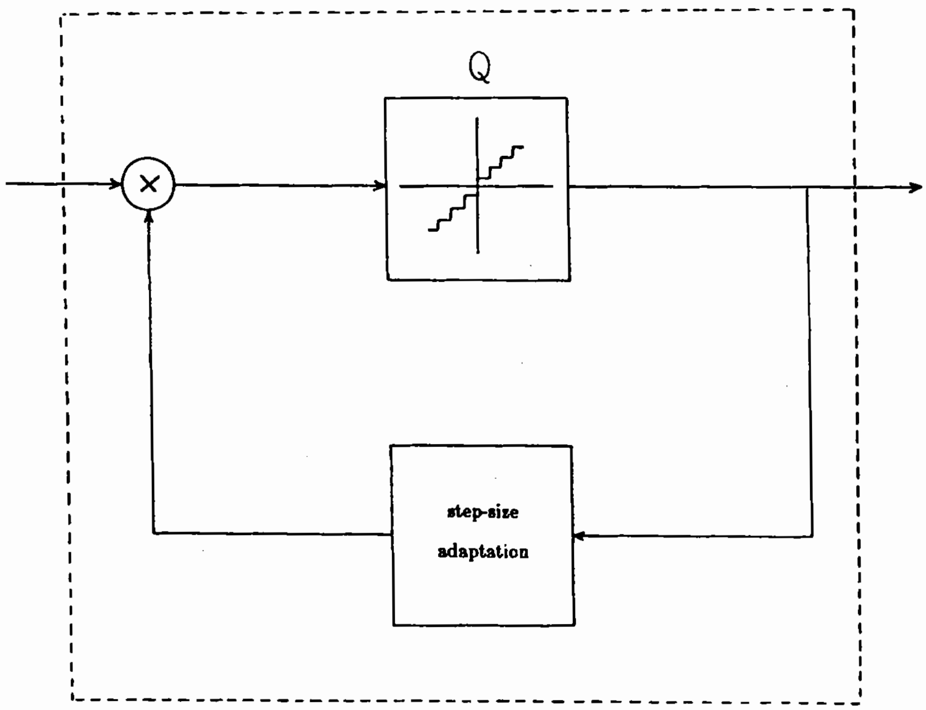

Figure 4.18 - APCM quantizer module.

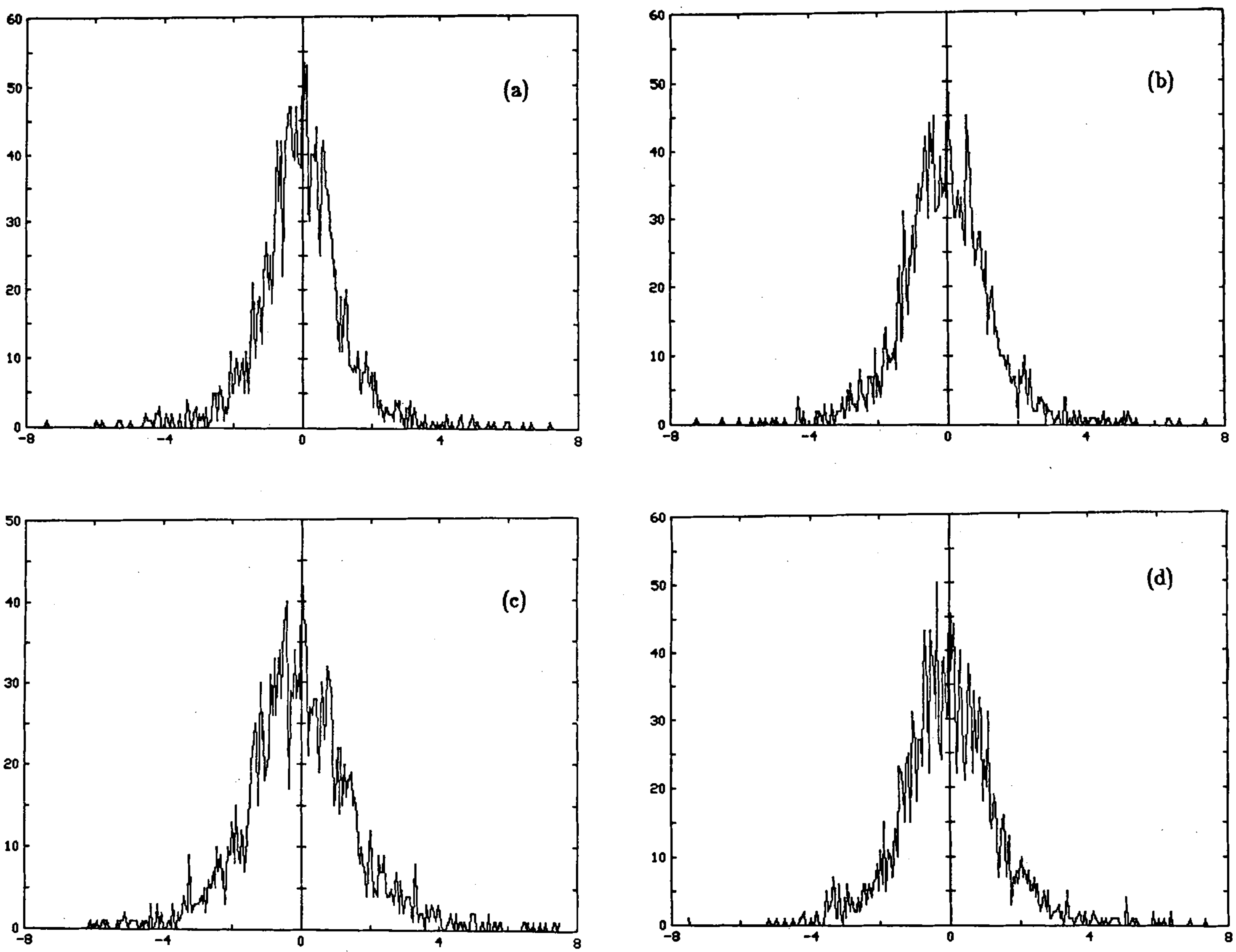

Figure 4.19 - Histogram of the input to the uniform quantizer inside the APCM module: (a) 2-bit quantizer; (b) 3-bit quantizer; (c) 4-bit quantizer; (d) 5-bit quantizer.

Notice from Fig. 4.19 that the distribution of the "real" quantizer input can be approximated by a Gamma function. So, the uniform quantizers of the APCM module were replaced by the optimal Gamma quantizers [19](the multipliers, however, remained unchanged). The histograms of the quantizer input signals produced using the **non-uniform** Gamma quantizers are shown in Fig. 4.20. It is seen that the histograms can still be approximated by the Gamma distribution. Hence we conclude that non-uniform Gamma quantizers are satisfactory.

The SNR figures corresponding to the coder employing non-uniform Gamma quantizers are given in Table 4.6. These are supplemented by the results of another set of five utterances (Table 4.7) each spoken by a male and a female speaker (not the same speakers as in the first set). These results show a modest improvement over the use of uniform quantizers. Listening tests show that in this configuration the reconstructed speech sounds a little richer and smoother. Informal subjective tests rate the performance of the coder as equivalent to 5.3-bit logarithmically-companded PCM.

4.7 - Perceptual Criteria

So far our efforts have been concentrated on minimizing the mean-square error of the output of the coder. Now we consider ways of improving the subjective quality of the coder.

4.7.1 - Weighting

As mentioned in chapter III, two forms of weights may be considered. One

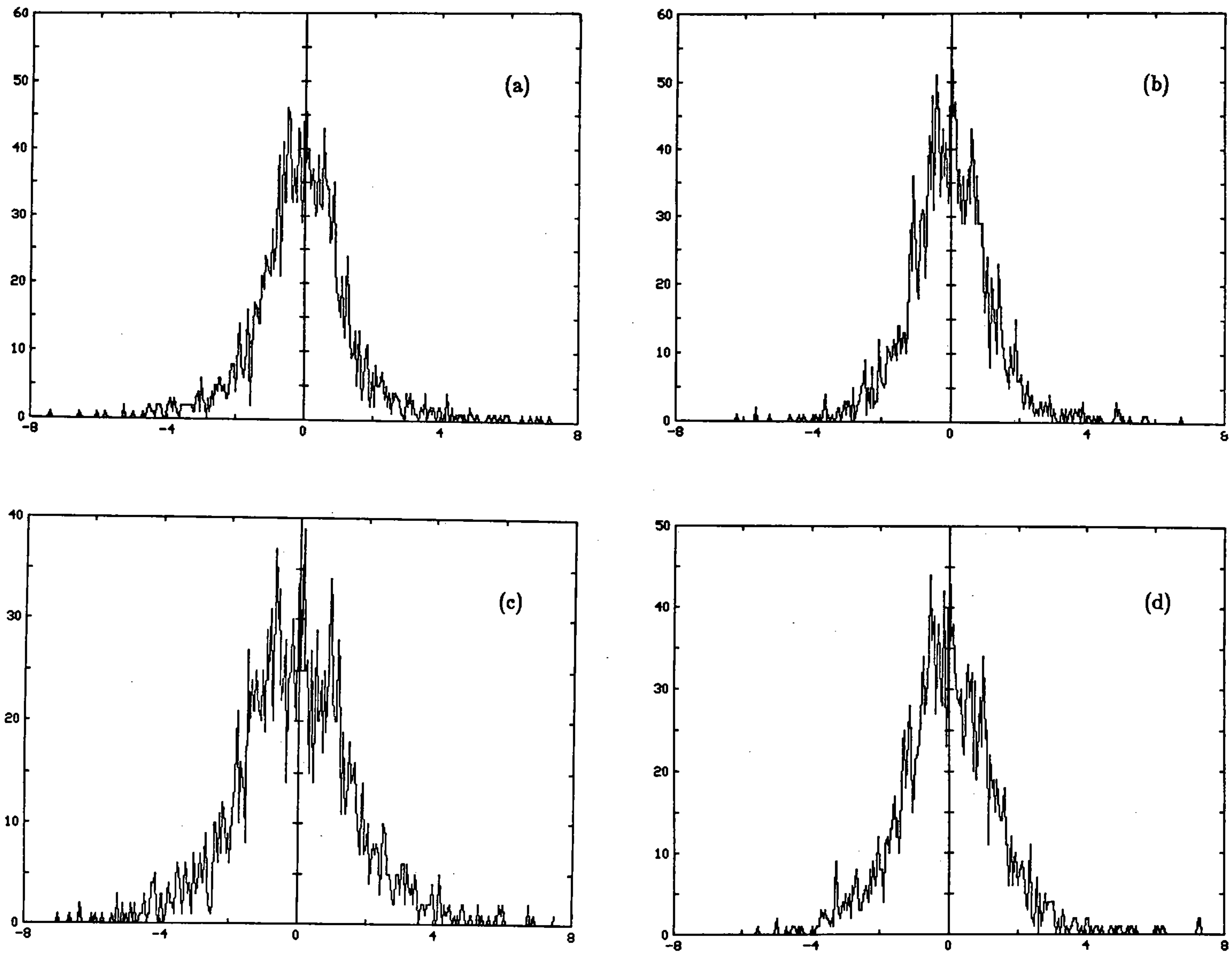

Figure 4.20 - Histogram of the input to the non-uniform quantizer inside the APCM module: (a) 2-bit quantizer; (b) 3-bit quantizer; (c) 4-bit quantizer; (d) 5-bit quantizer.

sentence	speaker	
	female	male
A	18.545/17.452	13.742/16.752
B	17.046/17.337	14.741/16.505
C	21.381/18.279	19.018/17.721
D	20.241/17.986	18.456/17.877
E	20.427/18.118	17.141/18.176

Sentence A: "Add the sum to the product of these three."

Sentence B: "Cats and dogs each hate the other."

Sentence C: "Oak is strong and also gives shade."

Sentence D: "Open the crate but don't break the glass."

Sentence E: "Thieves who rob friends deserve jail."

Table 4.6 - SNR (dB)/seg.SNR (dB) values of the coder utilizing non-uniform quatizers.

sentence	speaker	
	female	male
F	18.418/17.522	19.724/17.223
G	18.445/19.316	19.551/19.752
H	17.095/19.227	18.973/18.715
I	17.766/18.629	19.536/19.146
J	18.063/19.218	19.563/19.139

Sentence F: "The birch canoe slid on the smooth planks."

Sentence G: "Glue the sheet to the dark blue background."

Sentence H: "It's easy to tell the depth of a well."

Sentence I: "These days a chicken leg is a rare dish."

Sentence J: "Rice is often served in round bowls."

Table 4.7 - SNR (dB)/seg.SNR (dB) values of the coder utilizing non-uniform quatizers.

is to assign constant factors to each of the sub-bands. These would weight the sub-band energies so as to give varying degrees of importance to the sub-bands during the process of bit allocation. However, the usefulness of such weights is limited, since the perceptual significance of the sub-bands changes depending on the speech sound being uttered. Constant factors can be employed in situations where it is desired to bias the bit assignment algorithm in favor of a particular band. Such an application is exemplified by the final stage of the bit allocation algorithm, i.e. the elimination of the 1-bit quantizers; the highest bands are often assigned a single bit, which may be taken away and re-assigned to a lower band in the subsequent stage of the algorithm. The assignment of suitable weights to the higher bands can help control the direction of such transactions to insure sufficient representation of the higher sub-bands.

The other type of weighting uses as weights the sub-band energies raised to some power $-\gamma$. As the value of γ ranges from zero to one, the bit allocation algorithm undergoes changes from a strategy which results in equal distortion per sub-band to one which causes uniform distribution of the bits. As γ approaches unity, the low-energy sub-bands in the higher range of the frequency start to get a larger share of the overall bit-rate at the cost of less precision in the coding of the high-energy bands.

Subjective tests show that with increasing value of γ the reconstructed speech sounds less muffled and swishing noises are reduced. Also, since the difference in the bit assignment patterns across frames become smaller, the swishing sound dies out as well. However, the price paid is coarser quantization of the high-energy bands which results in harsher and less intelligible output. Subjectively it was found that the best value of γ is around 0.15. The speech quality deteriorated rapidly as

γ increased beyond 0.25.

Fig. 4.21 is a plot of the SNR and the segmental SNR of a test utterance as a function of γ.

4.7.2 - Addition of White Noise

Injection of white noise at the receiver is considered in order to fill in the spectral gaps caused by the assignment of zero bits to a sub-band.

A trade-off is involved in doing so. A positive effect is expected since the spectral nulls are eliminated. On the other hand, due to the interaction of the sub-bands in a QMF structure, some of the white noise is expected to leak into other bands resulting in further distortion.

Subjectively it was found that the injection of white noise does have some advantages; the output speech sounds more natural and the nasalization of the vowels is reduced. Also some of the burbling and swishing noise caused by the shifting of the bits is masked. Best results occurred when the white noise power was set to 75% of the energy of the sub-band into which it was injected.

The distortion caused by the introduction of white noise manifests itself in the form of background hissing. Intelligibility is slightly affected as well.

Since the injection of noise takes place at the receiver, it is obvious that extra side-information is necessary to transmit the magnitude of the power of the appropriate sub-bands to the receiver. However, it is possible to obtain an estimate

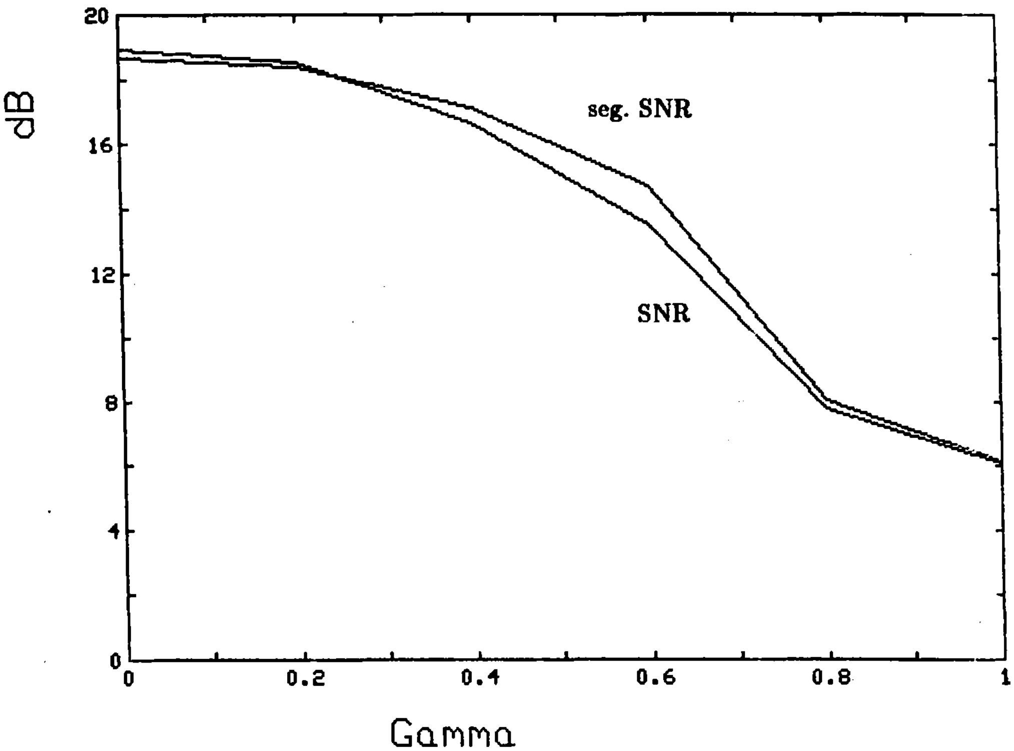

Figure 4.21 - Plot of the SNR and the segmental SNR of the sentence "it's easy to tell the depth of a well" as a function of the weighting parameter γ.

of the sub-band power without any side-information. From equation (3.14) we get the following relationship between the sub-bands i and j:

$$B_i - B_j = \frac{1}{g}\ln\left(\frac{\sigma_i^2}{\sigma_j^2}\right)$$

Now if band i is assigned zero bits, i.e. $B_i = 0$, then

$$\sigma_i^2 = \sigma_j^2 e^{-gB_j}$$

So, given the power of band j, it is possible to get an estimate of the power of sub-band i. To eliminate the need for extra side-information, the value of σ_j^2 used in the equation above can be computed from the previous frame of data. Because of such approximations as well as factors such as the integer bit assignment and the clamping of the bits, the estimate of σ_i^2 may not be accurate.

4.7.3 - Larger Bit Assignment Intervals

A problem inherent in a dynamic bit allocation system is the burbling and swishing noise caused by the shifting of bits across the frames. A simple way of reducing such noise is to increase the time-interval between successive bit allocation computations.

Recall that the bit assignment was performed every 32 milliseconds. In one experiment this interval was doubled and then tripled. In all cases the bit allocation computation was based on energy estimates of 32-millisecond frames of data. It was found out that the noise was reduced when the bit allocation computation was carried out at every other frame, i.e. at 64 msec intervals. Despite the sub-optimal configuration, no adverse effects were noticeable. However, the speech quality is degraded if the bit assignment takes place less often.

4.8 - Practical Considerations

In this section practical issues such as the side-information, the overall delay and the complexity of the coder are discussed.

4.8.1 - Side-information

For proper reconstruction of the input signal, the parameters relevant to the quantization process in the transmitter must be known to the receiver. The parameters necessary to completely describe an APCM quantizer -the quantization technique adopted in this work- are the number of the quantization levels and the step-size. In APCM quantization the step-size is modified according to the past output level (past index of quantization). So, if the initial step-size and the quantizer multipliers are known to the receiver, the step-size variations can be tracked exactly throughout the coding process. However, the number of quantization levels assigned to each band must be communicated to the receiver before each frame of data is transmitted.

There are a number of ways of providing this piece of information. One is to transmit the sub-band energies for each frame. Then the receiver could go through the same series of computations performed in the transmitter to obtain the bit assignment pattern. To do so, the sub-band energies have to be quantized for tranmission to the receiver. To get a good estimate of these parameters, each must be assigned at least 5 bits. In the case of a coder with eight sub-bands a total of 40 bits per frame of data is required. For a frame length of 32 msec this is equivalent to 1.25 kbits/sec.

Another possibility is to assign a codeword to each and every possible bit assignment pattern and transmit the appropriate codeword in the beginning of each frame of data. Statistics gathered using a total number of 20 utterances -spoken by two male and two female speakers- showed that only about 250 out of over 5000 possible bit patterns occurred. Therefore, presumably 9 bits, i.e. 512 codewords should be enough to represent the most frequently encountered patterns. For 32 msec-long frames the side-information takes up about 290 bits/sec. Of course a strategy must be adopted in case a computed bit pattern does not match any of those represented by the codewords. A simple strategy such as the use of the previous bit pattern is expected to be quite satisfactory.

4.8.2 - Delay

The overall coder delay is comprised of two components: the buffering and the filtering delays. The buffering delay is incurred since before a frame of data can be coded it must be stored, in a "buffer", to allow the computation of statistics necessary for the bit allocation procedure. So the buffering delay is equal to the length of the frame of data (32 msec).

To compute the filtering delay, recall that for a symmetric filter with N coefficients the delay is $\frac{N-1}{2}$ samples. The filtering delay due to the transmitter is computed as follows :

first filtering stage:	$\frac{N_1-1}{2}$
second filtering stage:	$2 \times \frac{N_2-1}{2}$
third filtering stage:	$4 \times \frac{N_3-1}{2}$

where N_i is the number of filter coefficients in the i^{th} stage. The factors 2 and 4 in the above expressions are due to the 2:1 decimations taking place in every filtering stage.

So the total transmission delay (D_t) is:

$$D_t = \frac{N_1 - 1}{2} + 2\frac{N_2 - 1}{2} + 4\frac{N_3 - 1}{2}$$

Since the filtering operation in the receiver is similar to the transmitter, the overall filtering delay (D_{total}) is twice the transmitter delay:

$$D_{total} = N_1 + 2N_2 + 4N_3 - 7$$

Notice that it is beneficial to have shorter filters in the second and the third stages.

For the case of a 64-tap filter in the first stage and a 24-tap filter in the subsequent stages the filtering delay is 201 samples, equivalent to about 25 milliseconds at the sampling rate of 8000 samples/sec. Hence the total delay is about 57 milliseconds.

The overall delay can be reduced by either shortening the length of the buffer, or eliminating the buffer altogether by basing the bit assignment computation on energy estimates obtained from the past frame of data (i.e. the one being transmitted). Obviously some degradation is to be expected due to the sub-optimal bit allocation

4.8.3 - Complexity

The complexity of a signal processing system is commonly measured in terms of the number of multiplication and addition operations required per output sample. One of the most important advantages of quadrature mirror filtering is its computational economy made possible because of the symmetric properties of such filters. To show how computational savings may be realized, consider a digital waveform $x(n)$, the input signal to a QMF block with the lowpass and highpass half-band filters $h_1(n)$ and $h_2(n)$, respectively. The sub-band signals, $x_1(n)$ and $x_2(n)$ are the results of the convolution of the input signal with each of the half-band filters:

$$x_1(n) = \sum_{j=0}^{N-1} h_1(n)x(j-n) = \sum_{j\,odd}^{N-1} h_1(n)x(j-n) + \sum_{j\,even}^{N-1} h_1(n)x(j-n)$$
$$x_2(n) = \sum_{j=0}^{N-1} h_2(n)x(j-n) = \sum_{j\,odd}^{N-1} h_2(n)x(j-n) + \sum_{j\,even}^{N-1} h_2(n)x(j-n)$$

Since $h_2(n) = (-1)^n h_1(n)$, then

$$x_1(n) = A(n) + B(n)$$
$$x_2(n) = A(n) - B(n)$$

where

$$A(n) = \sum_{j\,odd}^{N-1} h_1(n)x(j-n)$$
$$B(n) = \sum_{j\,even}^{N-1} h_1(n)x(j-n)$$

So, two filtering operations can be performed for the price of a single filtering operation plus an addition. Therefore the complexity for each QMF block is roughly equal to the length of its half-band filters. In the case of a QMF tree structure, in going from one level to the next the number of blocks is doubled, but due to the 2:1 decimation the overall number of computations remains constant. Thus in a three-level QMF tree structure the complexity C is approximately equal to the sum of the orders of the half-band filters, i.e.:

$$C \approx N_1 + N_2 + N_3$$

where N_i = the order of the half-band filter of the ith stage.

In the case of finite impulse-response (FIR) filters, due to the 2:1 sub-sampling after each stage of the QMF filtering, it is not necessary to compute the output samples which will be thrown away. So if FIR filters are used, complexity can be reduced by a factor of two. For the coder used in this study, the complexity in terms of the number of multiplies and additions is:

$$C \approx \frac{64}{2} + \frac{24}{2} + \frac{24}{2} = 56$$

which requires a computational speed of about 2.4 Microseconds per multiply and add.

The computational-speed requirement can be decreased by noting that a tree structure allows parallel processing of the sub-band signals; if the QMF tree is implemented using separate modules, the highest needed computational speed is that required of the longest filter -in this case the 64-tap filter. If a Finite Impulse Response (FIR) filter is employed, due to the 2:1 decimation, 32 multiplications and additions have to be performed per sample. Thus at the sampling rate of 8000 samples/sec, one needs a maximum computational speed of 3.9 microseconds per multiply and add.

Note that since the number of computations in the transmitter and the receiver are the same, the complexity of a transmitter-receiver pair is twice the figures given above.

5. CONCLUSIONS

In this chapter the contents of the previous chapters are summarized and the conclusions are discussed. At the end of the chapter areas of interest for future work are suggested.

The fundamentals of the sub-band coding theory have been presented. Various methods for the partitioning of the speech spectrum were considered and the Quadrature Mirror Filtering (QMF) technique was selected because of its computational efficiency and ease of implementation. The Adaptive Pulse Coded Modulation (APCM) strategy was adopted for the quantization of the sub-band signals since it offers a dynamic range and requires no overhead transmission of step-size information.

The selection of the sub-bands is dependent upon the band-splitting technique as well as the quantization strategy. The QMF method restricts the freedom of the choice of the sub-bands; only octave bandwidths are realizable. In this investigation the 0-4 kHz range was divided into eight 500 Hz sub-bands. Theoretically it is desirable to have narrow bands on the lower frequency range of the spectrum, and wider bands on the other end. But if another filtering stage is added to further split the lower bands, the extra 2:1 sub-sampling will double the distance between the samples of the sub-band signals. This increase will adversely affect the step-size adaptation in the APCM quantizers. Also, the addition of another QMF stage will increase the complexity and double the filtering delay.

A procedure is devised for the optimal (in the M.S.E. sense) assignment of the bits to the sub-bands. The development of this algorithm is based on the assumption of exponential rate-distortion function for the sub-band outputs. The optimal bit allocation results in uniform distribution of the quantization noise across the speech band. However, flat distortion is not perceptually desirable. Therefore, the parameters of the optimal coder must be perturbed through the use of weighting, in order to shape the noise spectrum.

The experimental results justify the assumption of the exponential rate-distortion function. The output of the sub-band coder is affected by two types of noise. One is a kind of burbling noise which makes the speech sound harsh. The harshness of the reconstructed speech can be reduced to a degree by employing non-uniform optimal Gamma quantizers. The other type of distortion is a high-pitch noise consisting of aliasing and swishing. The problem of aliasing can be effectively remedied through the utilization of a long half-band filter in the first stage of the QMF tree structure. A long filter with a high stop-band rejection ensures better separation of the sub-bands reducing the leakage of aliasing energy from the energetic lower half of the speech spectrum into the low-energy higher half.

The 1-bit quantizers are also found to contribute to the high-pitch noise. A simple means of overcoming this problem is to combine the single bits in pairs, in favor of the sub-bands with the highest residual bit-request. Subjectively, this strategy results in the reduction of noise. However, this leads to another form of distortion; the spectral nulls nasalize some vowels.

One conclusion that may be drawn from the observed effect of the 1-bit quantizers is that it is probably beneficial to have narrow bands in the high range of the spectrum as well. The rationale for this conclusion is that rather than having a single bit assigned to a, say, 1 kHz-wide band in the high frequency region, it is better to have 2 bits assigned to the more energetic half of such a band (recall that because of the 2:1 decimation, a single bit assigned to a 1 kHz band is equivalent to two bits assigned to a band half as wide).

The swishing noise is attributed to the shifting of bits across the frame boundaries and the discontinuities caused by it. Especially the transition to and from zero bits are suspected to be the main cause of such noise. It is therefore expected that further division of the sub-bands into narrower and hence more numerous bands should lead to an increase in the swishing noise.

The subjective quality of the coder output can be improved by the introduction of perceptual criteria. The shape of the noise spectrum can be controlled to some extent by imposing scale-factors on the sub-bands during the process of bit allocation. The weights can be used to give more significance to the bands in the higher range of the spectrum, which because of low energies are normally assigned very few quantization levels. The cost is coarser quantization of the energetic bands in the lower spectrum range, resulting in harsher reconstructed speech.

Injection of white noise is another way of improving the quality of the coder output. At the transmission rate of 16 kbits/sec the low-energy sub-bands are often assigned zero bits. The perceptual effect of such spectral gaps is less natural sounding speech and nasalization of some vowels. White noise generated and injected at the receiver is a way of alleviating this problem. The side-effects of the addition of noise are a background hissing and reduction of intelligibility. However, the background noise also serves to mask out some of the swishing noise discussed earlier.

The complexity of the coder with the configuration used throughout this work (a 64-tap half-band filter in the first filtering stage and 24-tap filters in the subsequent stages) is about 56 multiply and adds for each of the transmitter and the receiver devices. This is equivalent to roughly 2.4 microseconds per multiply and add. The digital hardware technology is at a stage where speeds as fast as 150 nanoseconds per multiply and add are realizable. Therefore, a real-time implementation of a QMF sub-band coder is feasible. Furthermore, if one opts for parallel processing of the sub-band signals in the tree structure, the computational-speed requirement can be reduced to 3.9 microseconds per multiply and add.

The overall delay of the coder is about 57 milliseconds. It may be desirable to reduce the delay because some of the presently available communication channels (notably the telephone links) cannot tolerate delays longer than about 45-50 milliseconds without echo suppression. The 57 millisecond delay mentioned above is comprised of a 32 millisecond buffering delay and a 25 millisecond filtering delay. The use of shorter filters for the purpose of reducing the overall delay may not be acceptable without seriously compromising the coder quality. However, shorter buffers can be tolerated to a degree, i.e. energy estimates needed for the computation of the bit assignments can be based on shorter data segments. Thus it is possible to decrease the buffering and hence the overall delay.

Sub-band coding of voice signals is known as a way of obtaining reasonably good quality speech at medium bit rates. It features attractive properties such as ease of implementation, robustness with respect to background noise and speaker variations and computational economy enhanced by the use of quadrature mirror filtering technique. The quality of sub-band coders is further improved by dynamic assignment of quantization levels to the sub-bands. Compared to fixed bit-assignment systems, sub-band coders with Dynamic Bit Allocation (DBA) sound smoother and more natural and offer higher intelligiblity. Despite the presence of some swishing noise the output quality of a DBA coder is distinctly better than a fixed bit-assignment one. The distinction becomes greater as the transmission rate decreases. The performance of the DBA coder was examined at rates as low as 9 kbits/sec. There is a sharp increase in the distortion at rates below 12 kbits/sec. At 9 kbits/sec the coder is noisy and intelligibility is reduced. Longer half-band filters must be employed in all stages to further reduce the aliasing energy which is more perceivable at low transmission rates.

Finally possibilities for further investigations are considered. At the rate of 16 kbits/sec the main source of distortion in the DBA coder output is high-frequency swishing and short bursts of high-pitch noise. This distortion is thought to be caused by quantizer transients at the frame boundaries, due to the changes in the number of quantization levels (especially when transitions to/from zero bits occur). A study of waveform characteristics corresponding to such distortions could lead to simple rules for the elimination (or at least reduction) of the noise.

The injection of white noise into sub-bands with zero bit-assignments was found to be beneficial. However, more investigation is necessary to find ways of communicating sub-band energy estimates to the receiver without incurring excessive overhead transmission of side-information.

Another area of work which is beginning to receive the attention of researchers is the development of recursive quadrature mirror filters. Recursive filters can be designed to have better performances like greater stop-band rejections and sharper roll-offs. Such characteristics lead to better separation of the bands and lower aliasing energies, making it possible to use QMF sub-band coders at lower transmission rates.

APPENDIX A

Unequal Bandwidths: The Optimal Bit Allocation

In this section the proof of the optimality of the second pass of the bit allocation algorithm presented in section 3.2.2 is given.

We will only consider the case of leftover bits, i.e. $T > 0$, since the case $T < 0$ follows intuitively. If $T = 1$ it is obvious that only bands can be considered whose relative rates of transmission (M) are one. So, the arguments of section 3.2.1 are applicable. However, if $T > 1$, the decision must be based on the assignment of two bits at a time. It is therefore necessary to determine whether it is best to:

1 - Give a single bit to a band with $M = 2$ (recall that this is equivalent to taking two bits away from T).

2 - Give two bits to a band with $M = 1$.

3 - Give one bit to each of two bands with $M = 1$.

Based on the discussion of the case $M_i = 1$ for $i = 1, \ldots, N$ it is obvious that in each case the selected bands must have the highest residual claims among the sub-bands of the same relative transmission rate (M). To select the optimal choice from the three listed above we must compare the reduction in the distortion caused by each.

Let us consider three sub-bands, indicated by the sub-scripts i, j and k, with energies σ_i^2, σ_j^2 and σ_k^2 such that $M_i = 2$ and $M_j = M_k = 1$. Now assume these bands have already been assigned b_i, b_j and b_k bits and that they have the residual claims a_i, a_j and a_k. Furthermore, assume that a_i is the greatest among the sub-bands with $M = 2$, and a_j and a_k are the greatest among the bands with $M = 1$, such that $a_j \geq a_k$.

The decrease in distortion due to each of the three choices above are :

$$\Delta_1 = \sigma_i^2 e^{-gb_i}(1 - e^{-g}) \tag{A.1}$$

$$\Delta_2 = \sigma_j^2 e^{-gb_j}(1 - e^{-2g}) \tag{A.2}$$

$$\Delta_3 = \sigma_j^2 e^{-gb_j}(1 - e^{-g}) + \sigma_k^2 e^{-gb_k}(1 - e^{-g}) \tag{A.3}$$

The choice number 1 will be made if $\Delta_1 > \Delta_2$ and $\Delta_1 > \Delta_3$. Let us examine each case.

If $\Delta_1 > \Delta_2$ then:

$$\sigma_i^2 e^{-gb_i}(1 - e^{-g}) > \sigma_j^2 e^{-gb_j}(1 - e^{-2g}) \tag{A.4}$$

From (3.10):

$$\frac{\sigma_i^2 e^{-g(b_i+a_i)}}{2} = \sigma_j^2 e^{-g(b_j+a_j)} \tag{A.5}$$

$$\Rightarrow \qquad \sigma_j^2 e^{-gb_j} = \frac{\sigma_i^2 e^{-gb_i}}{2} e^{g(a_j - a_i)} \tag{A.6}$$

Combining (A.6) and (A.4) we get:

$$\sigma_i^2 e^{-gb_i}(1 - e^{-g}) > \frac{\sigma_i^2 e^{-gb_i}}{2} e^{g(a_j - a_i)}(1 - e^{-2g}) \tag{A.7}$$

$$\Rightarrow \qquad e^{ga_i}2(1+e^{-g}) > e^{ga_j} \tag{A.8}$$

$$\Rightarrow \qquad a_i + \frac{\ln 2}{g} + \frac{1}{g}\ln(1+e^{-g}) > a_j \tag{A.9}$$

If $\Delta_1 > \Delta_3$ then:

$$\sigma_i^2 e^{-gb_i}(1-e^{-g}) > \sigma_j^2 e^{-gb_j}(1-e^{-g}) + \sigma_k^2 e^{-gb_k}(1-e^{-g}) \tag{A.10}$$

From (3.10):

$$\frac{\sigma_i^2 e^{-g(b_i+a_i)}}{2} = \sigma_i^2 e^{-gb_i} e^{-g(b_k+a_k)} \tag{A.11}$$

$$\Rightarrow \qquad \sigma_k^2 e^{-gb_k} = \frac{\sigma_i^2 e^{-gb_i}}{2} e^{g(a_k-a_i)} \tag{A.12}$$

Combining (A.10) with (A.6) and (A.12) we get:

$$2\sigma_i^2 e^{-gb_i}(1-e^{-g}) > \sigma_i^2 e^{-gb_i}(1-e^{-g})e^{g(a_j-a_i)} + \sigma_i^2 e^{-gb_i}(1-e^{-g})e^{g(a_k-a_i)} \tag{A.13}$$

$$\Rightarrow \qquad 2e^{ga_i} > e^{ga_j} + e^{ga_k} \tag{A.14}$$

$$\Rightarrow \qquad a_i + \frac{\ln 2}{g} > \frac{1}{g}\ln(e^{ga_j} + e^{ga_k}) \tag{A.15}$$

Note that if the inequality (A.15) holds, then we will also have:

$$a_i + \frac{\ln 2}{g} > \frac{\ln e^{ga_j}}{g} = a_j \tag{A.16}$$

and if (A.16) holds, then so does (A.9). So, we conclude that choice number one must be made if (A.15) is true. Otherwise, a single bit must be given to the band j and the procedure must be repeated until $T \leq 1$.

A special case occurs if $a_k = a_j$; the inequality (A.15) becomes $a_i > a_j$. This may also be used as an approximation to (A.15), since if $a_i > a_j$ then (A.15) holds as well.

APPENDIX B

The Coder Parameters

The experiments described in this report were carried out using 24, 36 and 64-tap filters. The 24 and the 64-tap filters are from [22]. The first half of the (symmetric) filter coefficients are given below.

24-tap	36-tap	64-tap
2.3292659×10^{-3}	4.62690077×10^{-4}	3.5961890×10^{-5}
-5.1829782×10^{-3}	1.47141311×10^{-2}	-1.1235150×10^{-4}
-2.2731449×10^{-3}	-1.29800290×10^{-3}	-1.1045870×10^{-4}
1.3540120×10^{-2}	-1.10011958×10^{-2}	2.7902771×10^{-4}
-6.5046689×10^{-4}	2.67996965×10^{-3}	2.2984380×10^{-4}
-2.7551951×10^{-2}	1.51144369×10^{-2}	-5.9535628×10^{-4}
1.0046210×10^{-2}	-5.28850593×10^{-3}	-3.8236310×10^{-4}
5.0881620×10^{-2}	-2.03917548×10^{-2}	1.1382600×10^{-3}
-3.4641430×10^{-2}	9.42546874×10^{-3}	5.3085393×10^{-4}
-9.9878848×10^{-2}	2.74511818×10^{-2}	-1.9861769×10^{-3}
0.1246452	-1.60941090×10^{-2}	-6.2437239×10^{-4}
0.4686479	-3.76989730×10^{-2}	3.2358770×10^{-3}
	2.77153254×10^{-2}	5.7431590×10^{-4}
	5.51907532×10^{-2}	-4.9891472×10^{-3}
	-5.21572791×10^{-2}	-2.5847671×10^{-4}
	-9.72526073×10^{-2}	7.3671709×10^{-3}
	1.39277145×10^{-1}	-4.8579351×10^{-4}
	4.59261596×10^{-1}	-1.0506890×10^{-2}
		1.8947140×10^{-3}
		1.4593960×10^{-2}
		-4.3136738×10^{-3}
		-1.9943651×10^{-2}
		8.2875602×10^{-3}
		2.7160550×10^{-2}
		-1.4853970×10^{-2}
		-3.7649728×10^{-2}
		2.6447000×10^{-2}
		5.5432450×10^{-2}
		-5.0954871×10^{-2}
		-9.7790956×10^{-2}
		0.1382363
		0.4600981

The step-size of each of the quantizers has a minimum, a maximum and a mid-rise/mid-tread threshold [23]. These are as follows:

sub-band	minimum	maximum	threshold
1	10	4500	15
2	10	4500	15
3	10	4500	15
4	10	4500	15
5	10	2500	15
6	10	2500	15
7	6	1200	9
8	1	100	1.5

The quantizer step-size multipliers are given below:

output index	number of output levels: 2	4	8	16	32
1	0.92	0.76	0.8	0.77	0.8
2	1.4	1.9	0.95	0.8	0.8
3			1.6	0.8	0.85
4			1.9	0.9	0.85
5				1.2	0.94
6				1.6	0.98
7				2.2	1.0
8				2.8	1.0
9					1.2
10					1.4
11					1.6
12					1.8
13					2.0
14					2.3
15					2.6
16					3.0

REFERENCES

[1] - Max J., "Quantizing for minimum distortion," IRE Trans. Inform. Theory, Vol. IT-6, March 1960.

[2] - Crochiere R.E., S.A. Weber and J.L. Flanagan, "Digital coding of speech in sub-bands," Bell Syst. Tech. J., Vol. 55, October 1976.

[3] - Estaban D. and C. Galand, "Application of Quadrature Mirror Filters to split band voice coding schemes," in Int. Conf. ASSP, May 1977.

[4] - Galand C. and D.J. Estaban, "16 KBPS real time QMF sub-band coding implementation," in Int. Conf. ASSP, April 1980.

[5] - Krasner M.A., "The critical band coder - Digital encoding of speech signals based on perceptual requirements of the auditory system," in Int. Conf. ASSP, April 1980.

[6] - Ramstad T.A. and O. Foss, "Sub-band coder design using recursive Quadrature Mirror Filters," Signal Processing: Theories and Applications, 1980.

[7] - Barnwell T.P., "An Experimental study of sub-band coder design incorporating recursive Quadrature Filters and optimum ADPCM," in Int. Conf. ASSP, March 1981.

[8] - Crochiere R.E. and M.R. Sambur, "A variable-band coding scheme for speech encoding at 4.8 Kb/s," Bell Syst. Tech. J., Vol. 56, May-June 1977.

[9] - Grauel C., "Sub-band coding with Adaptive Bit Allocation," Signal Processing, Vol. 2, January 1980.

[10] - Cheung R.S. and R.L. Winslow, "High quality 16 Kb/s voice transmission: The sub-band coder approach," in Int. Conf. ASSP, April 1980.

[11] - Makhoul J., M. Berouti and M. Krasner, "Time and frequency domain noise shaping in speech coding," in Int. Conf. ASSP, March 1981.

[12] - Kabal P., "Feasibility study of a hardware implementation of a 4 8 kb/s RELP speech coder," INRS-Telecommunications Technical Report, No. 81-08, May 1981.

[13] - Croisier A., "Progress in PCM and Delta Modulation: Block-Companded coding of speech signals," International Zurich Seminar, 1974.

[14] - Jayant N.S., "Adaptive quantization with a one-word memory," Bell Syst. Tech. J., Vol. 52, September 1973.

[15] - Mermelstein P., "Threshold of degradation for frequency-distributed band-limited noise in continuous speech," JASA, Vol. 72, 1982 (to appear).

[16] - Beranek L.L., "The design of communication systems," Proc. IRE, Vol. 35, September 1947.

[17] - Crochiere R.E., "An analysis of 16 Kb/s sub-band coder performance: Dynamic range, tandem connections and channel errors," Bell Syst. Tech. J., Vol. 57, October 1978.

[18] - Segall A., "Bit allocation and encoding for vector sources," IEEE Trans. Info. Theory, Vol. IT-22, March 1976.

[19] - Paez M.D. and T.H. Glisson, "Minimum mean-squared-error quantization in speech PCM and DPCM systems," IEEE Trans. Commun., Vol. COM-20, April 1972.

[20] - Tribolet J.M. and R.E. Crochiere, "Frequency domain coding of speech," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-27, October 1979.

[21] - Crochiere R.E., "On the design of sub-band coders for low bit-rate speech communication," Bell Syst. Tech. J., Vol. 56, May-June 1977.

[22] - Johnston J.D., "A filter family designed for use in Quadrature Mirror Filter banks," in Int. Conf. ASSP, April 1980.

[23] - Crochiere R.E., "A mid-rise/mid-tread quantizer switch for improved idle-channel performance in adaptive coders", Bell Syst. Tech. J., Vol. 57, October 1978.