ANALYTICAL AND SUBJECTIVE ANALYSIS OF DIFFERENTIAL PULSE CODE MODULATION VOICE COMMUNICATION SYSTEMS

by

DONALD CHAN

B.A.Sc., University of British Columbia, 1964

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

in the Department of
Electrical Engineering

We accept this thesis as conforming to the
standards required from candidates for the
degree of Master of Applied Science

Members of the Department
of Electrical Engineering

THE UNIVERSITY OF BRITISH COLUMBIA

November, 1967

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of ELECTRICAL ENGINEERING

The University of British Columbia
Vancouver 8, Canada

Date Nov 29, 1967

ABSTRACT

Differential pulse code modulation (DPCM) is a practical encoding scheme for speech, television, and telemetry signals. In this thesis, the mean square error $\overline{\varepsilon^2}$ is derived in terms of the quantizer characteristic, the spectra of the message, quantization noise, and channel noise, the sampling frequency f_s, the bandwidth W of the low-pass pre- and postfilters, and the coefficients α_i of the feedback filter whose impulse response $h(t) = \sum_{i=1}^{N} \alpha_i \, \delta(t-i/f_s)$, where δ is the unit impulse. The minimization of $\overline{\varepsilon^2}$ is discussed, and analytical solutions are obtained for some special cases.

A method is then developed for measuring the subjective quality of voice communication systems as a function of an arbitrary number of system parameters. The method was used to measure the quality of PCM speech and DPCM speech as a function of speech bandwidth W and number of quantization bits L. The quantization was logarithmic, f_s was constrained to equal 2.2W, channel noise was assumed negligible, and only previous-sample feedback DPCM was considered. An optimum W and L was found for every bit rate R = 2.2WL. For both PCM and DPCM, the equations relating the optimum W and L are of the form $W = a2^{bL}$, where a and b are constants independent of R. The maximum speech quality obtainable for PCM and DPCM was found for each R. Optimum DPCM was found to be better than optimum PCM for every R, and the superiority of DPCM over PCM increased with R. The reduction in bit rate which results

when optimum DPCM, rather than optimum PCM is used is determined as a function of the PCM bit rate and as a function of the maximum obtainable speech quality.

Included in the thesis is a description of the real-time DPCM system used in the research.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

ACKNOWLEDGEMENT

Grateful acknowledgement is given to the Defence Research Board of Canada and the National Research Council of Canada for financial support received under Grants DRB 2801-26 (UG) and NRC A-3308, respectively.

I would like to thank Dr. R.W. Donaldson, the supervisor of this project, not only for his valuable suggestions but also for his generous counsel and constant encouragement.

I would like to thank Dr. E.V. Bohn for reading the manuscript and for his helpful suggestions. I am grateful to Dr. J.S. MacDonald and Mr. G.A. Austin for their many helpful suggestions pertaining to the design of the real-time DPCM system. I am also grateful to the graduate students and staff members of the Electrical Engineering Dept. of U.B.C. who participated in the listening tests, and to Lenkurt Electric Company, Burnaby, Canada, for providing the low-pass filters used in the DPCM system.

I would also like to thank my wife and my parents for their encouragement, Mrs. R.G. Wein for typing the manuscript, Mr. J.P. Steger for preparing some of the illustrations, and Messrs. K. Peal, M. Ito, and P.J.R. Douville for proofreading the manuscript.

I. INTRODUCTION

1.1 Transmission of Messages over Digital Communication Channels

Processing a message x(t) for transmission over a noisy digital communication channel can be considered to be a concatenation of three operations: source encoding, channel encoding, and modulation (see Fig. 1.1). The source encoder

Fig. 1.1 General digital communication system.

transforms the message into a sequence of discrete symbols. Each symbol is chosen from a finite alphabet having A members. Since an ideal source encoder removes all redundancy from the message, each symbol in the output sequence occurs with probability 1/A, and is statistically independent of all other symbols in the sequence. In practice, only part of the message redundancy is removed by the source encoder. Removal of all redundancy is usually impractical, if not impossible.

The channel encoder adds, in a way that is optimum for the particular channel and modulation system used, enough redundancy to keep the probability of a transmission error below some specified limit. The modulator transforms the channel encoder output symbol sequence into a signal suitable for

transmission over the prescribed channel. The received signal is demodulated and decoded to yield a delayed replica $\hat{x}(t)$ of the original message.

1.2 Review of Previous Research on Source Encoders

Many source encoders which improve the redundancy reduction efficiency over that which results from using ordinary pulse code modulation (PCM) have been proposed. A practical, widely-used source encoder has the general configuration of Fig. 1.2. The system in Fig. 1.2 is a differential pulse code modulation (DPCM) system, of which PCM and delta modulation (ΔM) are special cases. PCM is DPCM with no feedback, while ΔM is identical to previous-sample feedback DPCM with one bit of quantization.

Numerous optimization studies on PCM systems have been carried out. Max[1], Williams [2], and Bruce [3,4] have considered the optimization of the quantizing process. Max determined the optimum nonuniform quantizer for a Gaussian input and a mean square distortion measure. Williams optimizes the quantizer when the message has the exponential amplitude probability-density function of speech. Bruce gives a computer algorithm for the determination of the optimum nonuniform quantizer for an

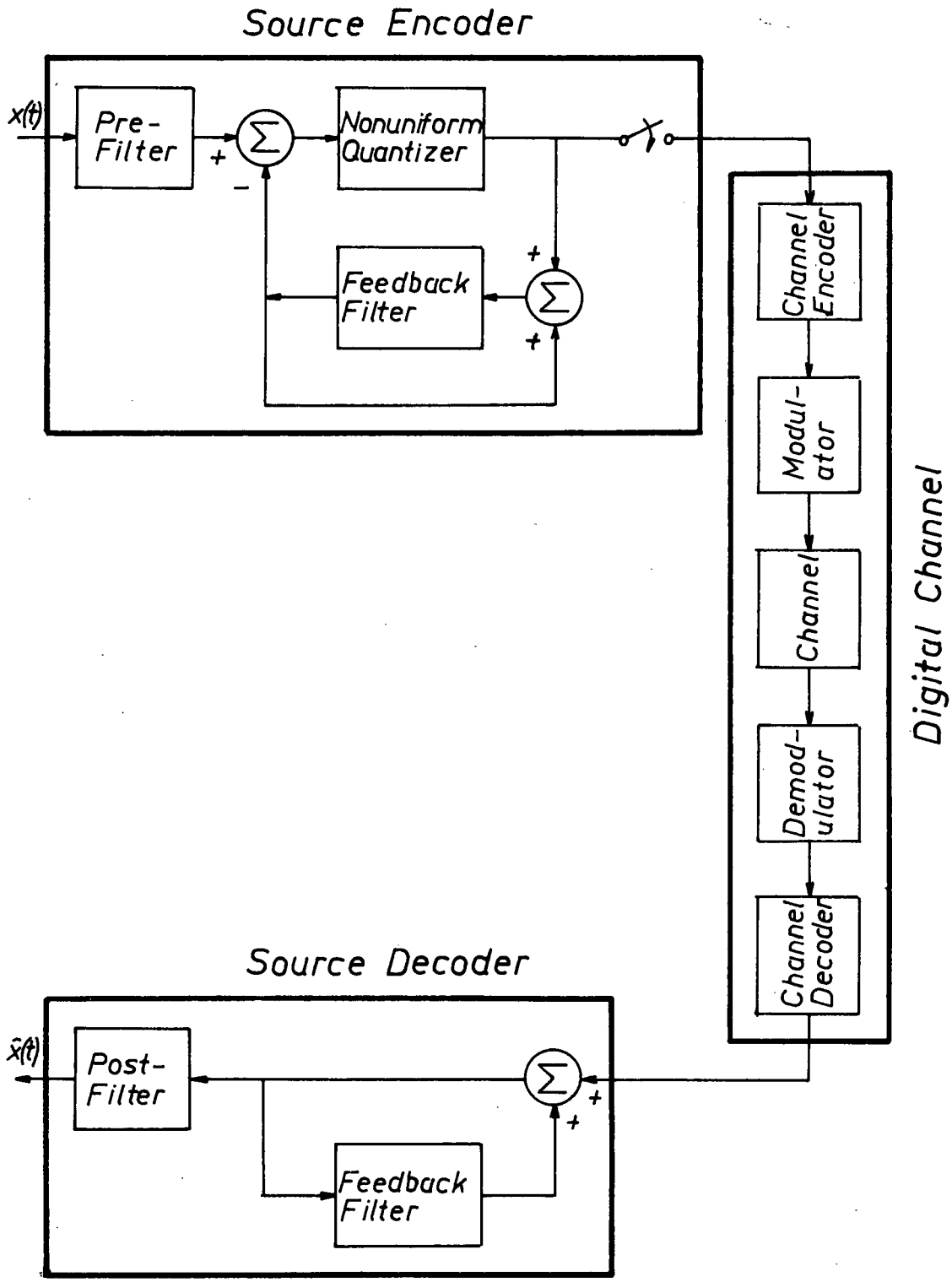


Fig. 1.2 General differential pulse code modulation (DPCM) system

arbitrary distortion measure and an arbitrary message contaminated by noise. Panther and Dite [5], Roe [6], and Algazi [7] have presented approximate practical techniques for obtaining optimum quantizers.

Other investigators have examined the problem of optimizing the combined process of quantizing, sampling, and reconstruction (source decoding). Goodman [8] derived a general equation for the sampling rate and quantizing fineness which minimizes the mean square error when ideal low-pass pre- and postfilters and Gaussian input signals are used. Liff [9], Katzenelson [10], and Steiglitz [11], consider the mean square distortion for arbitrary linear, time-invariant postfilters and arbitrary random inputs. Liff gives expressions for the mean square error caused by quantizing, sampling, and reconstruction. Katzenelson determines the optimum postfilter for reconstructing a uniformly quantized sampled signal. Steiglitz investigates for several postfilters, the optimum tradeoff between the error due to quantizing and the error due to the process of sampling and reconstruction.

Several investigators have analysed source encoders which use feedback around the quantizer in order to reduce the redundancy in the encoder output. Van De Weg [12] calculates the sampling rate required to obtain a certain signal-to-noise ratio as a function of the number of bits of uniform quantization for a system with an integrating network in the feedback path. Nitadori [13] obtains the quantizer characteristic which minimizes the quantization noise for speech signals when the feedback filter is an ideal integrator. O'Neal [14,15] and McDonald [16],

have derived approximate formulas for the signal-to-noise ratio of DPCM systems which do not use postfilters. O'Neal applied his analysis to video signals, while McDonald considered speech signals. Their approximations become invalid when the quantizer has fewer than approximately eight levels.

Although the previous discussion concerns non-adaptive systems, adaptive sampling, quantizing and feedback have also been studied [17,18]. At present, most adaptive systems are impractical.

1.3 Scope of the Thesis

In this thesis, the DPCM system shown in Fig. 2.2 is considered. The input to the system is first low-pass filtered, then compared with a linear estimate based on the input's past history. The estimate is obtained by delaying the feedback signal integer multiples of the sampling period and weighting each delayed signal by an appropriate value. The difference between the actual signal and its predicted or estimated value is then quantized and sampled before being presented to the channel encoder for transmission over the digital channel. The source decoder uses this transmitted quantized difference to reconstruct the bandlimited input signal.

No one has yet obtained the mean square error for the DPCM system in Fig. 2.2 as a function of W, α_1, α_2, ...α_N, f_s, the quantizer characteristic, and the statistics of the

message and digital channel noise. Such an expression is derived in Chapter II. Although the mean square error of a transmission system is indicative of its goodness, ultimate evaluation requires subjective measurements on real-time signals. Several investigators [19,20,21] have made subjective measurements on DPCM video systems; but few, if any, subjective measurements of DPCM voice systems have been published. Chapter IV describes a method for measuring the subjective goodness of a voice communication system as a function of an arbitrary number of system parameters. The method is then used to measure the subjective quality of PCM and previous-sample DPCM speech systems as a function of the sampling rate f_s and number of quantization bits L. The sampling rate is fixed at 2.2 times the 3 db cutoff of the low-pass filter W; the quantizer characteristic is logarithmic with $\mu = 100$ [22]; and the channel noise is zero. It was found that for both PCM and DPCM, an optimum choice of sampling rate and number of quantization levels exist for every bit rate $R = f_s L$, and that the optimum W and L are related by an expression of the form $W=a2^{bL}$, where constants a and b are different for PCM and DPCM. Furthermore, it was found that optimum DPCM is subjectively better than optimum PCM, and the superiority of DPCM over PCM increased with R.

Chapter III describes the design and construction of the DPCM system used to make the measurements described in Chapter IV.

II. MEAN SQUARE ERROR IN DPCM

2.1 Some Useful Relations in Communication Theory

Before proceeding to the analysis of the DPCM system shown in Fig. 2.2, some relations used in the subsequent derivation will be presented without proof.

2.1.1 Convolution

Given two functions $x_1(\alpha)$ and $x_2(\alpha)$, the integral

$$x(\alpha) = \int_{-\infty}^{\infty} x_1(\lambda)x_2(\alpha-\lambda)\,d\lambda = \int_{-\infty}^{\infty} x_2(\lambda)x_1(\alpha-\lambda)\,d\lambda$$

is known as the convolution of $x_1(\alpha)$ and $x_2(\alpha)$, and is often written

$$x(\alpha) = x_1(\alpha) \textcircled{x} x_2(\alpha) \quad .$$

Convolution has the property of being

(i) reflexive $\quad x_1 \textcircled{x} x_2 = x_2 \textcircled{x} x_1$,

(ii) distributive $\quad x_1 \textcircled{x} (x_2 + x_3) = x_1 \textcircled{x} x_2 + x_1 \textcircled{x} x_3$,

and (iii) associative $\quad x_1 \textcircled{x} (x_2 \textcircled{x} x_3) = (x_1 \textcircled{x} x_2) \textcircled{x} x_3$.

If $\delta(\alpha)$ is the unit impulse function, then

$$x(\alpha) \textcircled{x} \delta(\alpha) = x(\alpha) \quad .$$

2.1.2 Correlation and Power Spectrum

In the following discussion, only stationary processes are considered. The autocorrelation for a stationary random process $x_1(t)$ is defined by

$$R_{x_1x_1}(\tau) = R_{x_1}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_1(t)x_1(t-\tau)dt$$

and is usually abbreviated to

$$R_{x_1}(\tau) = \overline{x_1(t)\ x_1\ (t-\tau)} \ .$$

One also defines the cross-correlation for two random processes $x_1(t)$ and $x_2(t)$

$$R_{x_1x_2}(\tau) = \overline{x_1(t)x_2(t-\tau)} \ .$$

Clearly,

$$R_{x_1}(\tau) = R_{x_1}(-\tau)$$

and

$$R_{x_1x_2}(\tau) = R_{x_2x_1}(-\tau) \ .$$

The power spectrum (or spectral density) $S_{x_1x_1}(f)$ of a process $x_1(t)$ is the Fourier transform of its auto-correlation

$$S_{x_1x_1}(f) = S_{x_1}(f) = \int_{-\infty}^{\infty} R_{x_1}(\tau)\ e^{-j2\pi f\tau}\ d\tau \ .$$

Similarly, the cross-power spectrum $S_{x_1x_2}(f)$ of two processes $x_1(t)$ and $x_2(t)$ is the Fourier transform of their cross-correlation

$$S_{x_1x_2}(f) = \int_{-\infty}^{\infty} R_{x_1x_2}(\tau)e^{-j2\pi f\tau_d\tau} .$$

If * denotes complex conjugation, then

$$S_{x_1}(f) = S_{x_1}{}^*(f)$$

and $$S_{x_1x_2}(f) = S_{x_2x_1}{}^*(f) .$$

If $x_1(t)$ is a periodic function of period T, then the autocorrelation is defined by

$$R_{x_1x_1}(\tau) = R_{x_1}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} x_1(t)x_1(t-\tau)dt$$

and the power density by the Fourier transform of $R_{x_1}(\tau)$,

$$S_{x_1x_1}(f) = S_{x_1}(f) = \int_{-\infty}^{\infty} R_{x_1}(\tau)\, e^{-j2\pi f\tau}{}_d\tau .$$

If $x_1(t)$ is expressed as a Fourier series,

$$x_1(t) = \sum_{k=-\infty}^{\infty} c_k e^{j2\pi kt/T}$$

where $$c_k = \frac{1}{T}\int_{-T/2}^{T/2} x_1(t)e^{-j2\pi kt/T}\, dt$$

then $$R_{x_1}(\tau) = \sum_{k=-\infty}^{\infty} |c_k|^2 \, e^{j2\pi k \tau/T}$$

and $$S_{x_1}(f) = \sum_{k=-\infty}^{\infty} |c_k|^2 \, \delta(f-k/T).$$

2.1.3 Stationary Processes Under Linear Transformation

Consider the two linear systems shown in Fig. 2.1 with impulse responses $h_1(t)$, $h_2(t)$ and frequency responses $H_1(f)$, $H_2(f)$. With $x_1(t), x_2(t)$ as inputs and $y_1(t)$, $y_2(t)$ the

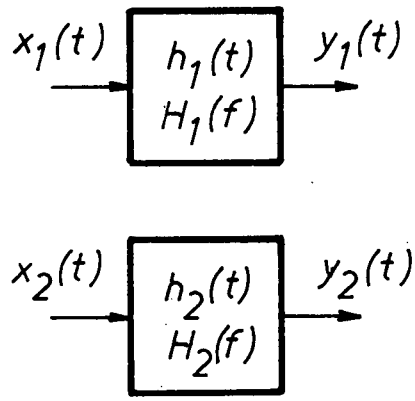


Fig. 2.1 Stationary processes under linear transformation.

corresponding outputs, the cross-correlation and cross-power spectrum of the outputs are given by

$$R_{y_1y_2}(\tau) = \int_{-\infty}^{\infty} h_1(\lambda)\,d\lambda \int_{-\infty}^{\infty} h_2(\alpha)\,d\alpha \quad R_{x_1x_2}(\tau - \lambda + \alpha)$$

and $$S_{y_1y_2}(f) = H_1(f)\, H_2^*(f) S_{x_1x_2}(f).$$

For the case when $h_1(t) = \delta(t)$,

$$R_{y_1y_2}(\tau) = R_{x_1y_2}(\tau) = R_{x_1x_2}(\tau) \textcircled{x} h_2(-\tau)$$

and $S_{y_1y_2}(f) = S_{x_1y_2}(f) = S_{x_1x_2}(f)\, H_2^*(f)$.

Another case of interest occurs when $x_1(t) = x_2(t)$ and $h_1(t) = h_2(t)$. For this case,

$$S_{y_1y_2}(f) = S_{y_1}(f) = S_{x_1}(f)\left|H_1(f)\right|^2.$$

2.2 Mean Square Error in DPCM Systems[+]

In the DPCM system of Fig. 2.2, $\hat{x}_1 = h \textcircled{x} (x_1+q)$ and $e = x_1 - \hat{x}_1 = x_1 \textcircled{x} (\delta - h) - q \textcircled{x} h$. If the digital channel noise $n(t)$ is an infinite sequence of periodic rectangular pulses of width Δ and random amplitude, and if the sampler is represented by a product modulator in which the sampler input signal is multiplied with an infinite sequence of periodic rectangular pulses $p(t)$ of width Δ and unit amplitude, then $r = p.(e+q)+n$ and $\hat{x} = r \textcircled{x} f \textcircled{x} g_o$, where $f(t)$ is the impulse response of the system having transfer function $1/(1-H(f))$. Therefore, after substituting for e and r,

---

[+]Part of the analysis of this section is based on an unpublished memorandum by R.W. Donaldson.

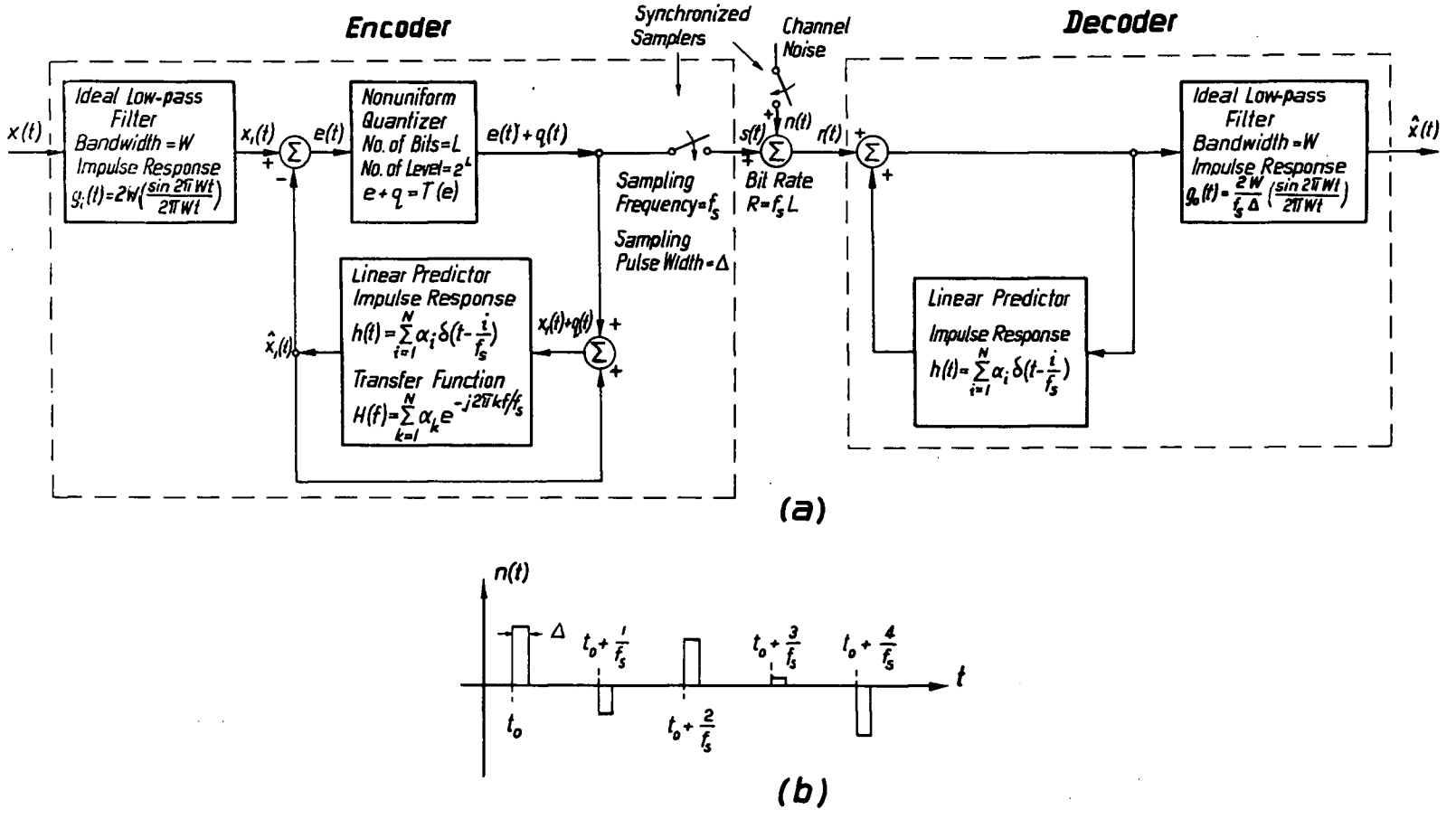


Fig. 2.2 (a) A practical DPCM system (b) Noise caused by transmission error in the digital channel

$$\hat{x} = \left[p.(x_1 \textcircled{x} (\delta - h))\right] \textcircled{x} f \textcircled{x} g_o + \left[p.(q \textcircled{x} (\delta - h))\right] \textcircled{x} f \textcircled{x} g_o + n \textcircled{x} (f \textcircled{x} g_o). \tag{2.1}$$

Periodic function $p(t) = \sum_{k=-\infty}^{\infty} c_k e^{j2\pi k f_s t}$, where

$c_k = (1/\pi k)\sin \pi k f_s \Delta$ for $k \neq 0$, and $c_o = f_s \Delta$. Since $x_1(t)$ is bandlimited by an ideal low-pass filter and $f_s \geq 2W$, $\left[p.(x_1 \textcircled{x} (\delta - h))\right] \textcircled{x} g_o = x_1 \textcircled{x} (\delta - h)$. Since $(\delta - h) \textcircled{x} f = \delta$,

$$\hat{x} = x_1 + \left[p.(q \textcircled{x} (\delta - h))\right] \textcircled{x} f \textcircled{x} g_o + n \textcircled{x} (f \textcircled{x} g_o). \tag{2.2}$$

If $x_2(t) = x(t) - x_1(t)$, then $\overline{\varepsilon^2} = \overline{(x - \hat{x})^2} = \overline{(x_1 - \hat{x})^2} + \overline{x_2^2} + \overline{2(x_1 - \hat{x})x_2}$.

The term $\overline{(x_1 - \hat{x})x_2} = 0$, since $(x_1 - \hat{x})$ and x_2 have no common spectral components. If $q(t)$ and $n(t)$ are uncorrelated random processes[+], then

$$\overline{\varepsilon^2} = \overline{x_2^2} + \overline{\left[p.(q \textcircled{x} (\delta - h)) \textcircled{x} f \textcircled{x} g_o\right]^2} + \overline{\left[n \textcircled{x} (f \textcircled{x} g_o)\right]^2}. \tag{2.3}$$

If $x(t)$ has power spectrum $X(f)$, then $\overline{x_2^2} = 2\int_W^{\infty} X(f)df$.

[+]Necessary and sufficient conditions for $q(t)$ and $n(t)$ to be uncorrelated are derived by R.E. Totty and G.C. Clark[32].

Let $N(f)$ be the power spectrum of $n(t)$. Then $\overline{\left[n \textcircled{x} (f \textcircled{x} g_o)\right]^2}$ $= (1/f_s\Delta)^2 \int_{-W}^{W} \left[N(f)/\left|1-H(f)\right|^2\right] df$. If the noise pulse are statistically independent, and if σ^2 is the variance of noise pulse amplitudes, then $N(f) = \sigma^2 f_s \Delta^2 (\sin \pi f\Delta / \pi f\Delta)^2$. If $\Delta \ll 1/2W$, then for $-W \leq f \leq W$, $N(f) = \sigma^2 f_s \Delta^2$ and $\overline{\left[n \textcircled{x} (f \textcircled{x} g_o)\right]^2}$ $= (\sigma^2/f_s) \int_{-W}^{W} (1/\left|1-H(f)\right|^2)\, df$.

Let $Q(f)$ be the power spectrum of the quantization noise $q(t)$. The spectrum of $q \textcircled{x} (\delta - h)$ is $Q(f)\left|1-H(f)\right|^2$, and the spectrum of $p.(q \textcircled{x} (\delta - h)))$ is $\sum_{k=-\infty}^{\infty} |c_k|^2 Q(f-kf_s)\left|1-H(f-kf_s)\right|^2$.

It follows that the power spectrum of $\left[p.\,(q \textcircled{x} (\delta - h))\right] \textcircled{x} f \textcircled{x} g_o$ equals $\sum_{k=-\infty}^{\infty} (|c_k|^2/c_o^2) \int_{-W}^{W} \left[Q(f-kf_s)\left|1-H(f-kf_s)\right|^2/\left|1-H(f)\right|^2\right] df$.

In Fig. 2.2, $H(f) = \sum_{m=1}^{N} \alpha_m e^{-j2\pi mf/f_s}$. Therefore, $H(f-kf_s) = \sum_{m=1}^{N} \alpha_m e^{-j2\pi m(f/f_s - k)} = H(f)$, and $\left|1-H(f-kf_s)\right| / \left|1-H(f)\right| = 1$.

If $1/\Delta$ is very much larger than the bandwidth of $q(t)$, then $c_k = c_o$ for all k for which $Q(f-kf_s)$ has any significant amplitude in the frequency range $-W \leq f \leq W$. In this case,

$$\overline{\varepsilon^2} = 2\int_{W}^{\infty} X(f)df + \sum_{k=-\infty}^{\infty} \int_{-W}^{W} Q(f-kf_s)df + (1/f_s\Delta)^2 \int_{-W}^{W} \left[N(f)/\left|1-H(f)\right|^2\right] df. \tag{2.4}$$

Note that if $f_s = 2W$, then $\sum_{k=-\infty}^{\infty} \int_{-W}^{W} Q(f-kf_s)df = \int_{-\infty}^{\infty} Q(f)df = \overline{q^2}$.

Let f_Q equal $\sum_{k=-\infty}^{\infty} \int_{-W}^{W} Q(f-kf_s)df$ when $\overline{e^2} = 1$. The rate distortion characteristic f_Q depends on the quantizer characteristic T(e) in Fig. 2.2, the amplitude probability density function $p_e(\beta)$ of e(t), the bandwidth W, and the sampling frequency f_s. If e(t) is multiplied by a constant K, and if the quantizer characteristic is scaled in both the horizontal and vertical direction such that $e + q = KT(e/K)$, then q(t) is multiplied by K also, and $\sum_{k=-\infty}^{\infty} \int_{-W}^{W} Q(f-kf_s)df = \overline{e^2} f_Q$. It will be assumed that the quantizer in Fig. 2.2 is scaled in proportion to $\sqrt{\overline{e^2}}$.

In Fig. 2.2, $\overline{e^2} = \overline{\left[x_1 \textcircled{x} (\delta - h)\right]^2} + \overline{(q \textcircled{x} h)^2} - 2\overline{(q \textcircled{x} h)\left[x_1 \textcircled{x} (\delta - h)\right]}$. Since $h(t) = \sum_{m=1}^{N} \alpha_m \delta(t-m/f_s)$,

$$\overline{e^2} = \overline{\left[x_1 - \sum_{m=1}^{N} \alpha_m x_1(t-m/f_s)\right]^2} + \overline{\left[\sum_{m=1}^{N} \alpha_m q(t-m/f_s)\right]^2} - 2\overline{\left[\sum_{m=1}^{N} \alpha_m q(t-m/f_s)\right]\left[x_1 - \sum_{m=1}^{N} \alpha_m x(t-m/f_s)\right]} \qquad (2.5)$$

Define $R_{x_1}(\tau)$ and $R_q(\tau)$ as the autocorrelation functions of $x_1(t)$ and q(t) respectively, $R_{x_1 q}(\tau) = \overline{x_1(t)q(t-\tau)}$ as the cross-correlation of $x_1(t)$ and q(t), and $R_{xq}(\tau)$ as the cross-correlation of x(t) and q(t). Thus,

$$\overline{e^2} = \sum_{i=0}^{N} \sum_{j=0}^{N} \alpha_i \alpha_j R_{x_1}((i-j)/f_s)$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j R_q((i-j)/f_s)$$

$$+ 2 \sum_{i=0}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j R_{x_1 q}((i-j)/f_s) \qquad (2.6)$$

where $\alpha_o = -1$, $R_{x_1}(\tau) = \int_{-W}^{W} X(f) e^{j2\pi f \tau} df$, and $R_{x_1 q}(\tau) = R_{xq}(\tau) \textcircled{x} g_i(\tau)$.

By replacing $\sum_{k=-\infty}^{\infty} \int_{-W}^{W} Q(f-kf_s)df$ in (2.4) by $\overline{e^2}.f_Q$, and by subsequently replacing $\overline{e^2}$ by (2.6), one obtains

$$\overline{\varepsilon^2} = 2\int_{W}^{\infty} X(f)df + \Big[\sum_{i=0}^{N} \sum_{j=0}^{N} \alpha_i \alpha_j R_{x_1}((i-j)/f_s) + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j R_q((i-j)/f_s)$$

$$+2 \sum_{i=0}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j R_{x_1 q}((i-j)/f_s)\Big] f_Q + (1/f_s\Delta)^2 \int_{-W}^{W} \Big[N(f)/\Big|1-\sum_{m=1}^{N} \alpha_m e^{-j2\pi mf/f_s}\Big|^2\Big] d$$

(2.7)

Note that if $h(t) = \sum_{i=1}^{N} \alpha_i \delta(t-i/f_s) = 0$, then the DPCM system shown in Fig. 2.2 reduces to a PCM system. Therefore, by letting $\alpha_i = 0$ $(i = 1,2,..,N)$ the previous results apply equally

as well to PCM systems.

2.3 Mean Square Error for Special Cases of the DPCM System

Given X(f), N, the allowable bit rate R, N(f), and the amplitude probability density $p_x(\beta)$ of x(t), $\overline{\varepsilon^2}$ in (2.7) is minimized by making an appropriate choice for f_s, W, α_i (i=1,2,...,N) and function T. In general, an exact analytical solution is impossible. Functions $R_{xq}(\tau)$ and $R_q(\tau)$ depend on f_s, W, α_i, T, and $p_x(\beta)$. Exact expressions for $R_{x_1q}(\tau)$ and $R_q(\tau)$ are impossible to obtain even when f_s, W, α_i, and T are known.

When $R/f_s \geq 3$, $|q(t)| << |e(t)|$. In this case, $|R_q(\tau)|$ and $|R_{x_1q}(\tau)|$ are much smaller than $|R_{x_1}(\tau)|$, hence $\overline{e^2} = \overline{\left[\sum_{i=0}^{N} \alpha_i x_1(t-i/f_s)\right]^2}$. If the channel noise N(f) is negligible[+] and if the rate distortion function f_Q is only weakly dependent on the α_i's, then $\overline{\varepsilon^2}$ can be minimized with respect to α_i(i=1,2,...N) by choosing α_i to be a solution of the system of linear equations [14,16,23]

$$R_{x_1}(i/f_s) = \sum_{j=1}^{N} \alpha_j R_{x_1}((i-j)/f_s) \qquad i=1,2,\ldots N. \tag{2.8}$$

For this choice of the α_i's,

[+] Channel noise can always be made negligible by efficient channel encoding provided the information rate R is less than the channel capacity C [24,25].

$$\overline{\varepsilon^2} = 2\int_W^\infty X(f)df + \left[2\int_0^W X(f)df - \sum_{i=1}^{N} \alpha_i R_{x_1}(i/f_s)\right] f_Q \qquad (2.9)$$

If f_Q for the optimum quantizer can be expressed in terms of f_s, W, and bit rate R, then (2.9) can be minimized with respect to W and f_s. When e(t) is a Gaussian process, and $f_s = 2W$, f_Q can be determined from the rate distortion curves obtained by Max [1]. Max shows that $f_Q = (2.2)2^{-2(R/f_s)}$ when $R/f_s \succ 3$. Panther and Dite [5] have developed an approximate method for finding the optimum quantizer and the corresponding f_Q for arbitrary amplitude probability density functions $p_e(\beta)$. Their method requires that $f_s = 2W$ and $R/f_s \succ 3$.

When the quantizer has fewer than approximately eight levels, the assumptions on which (2.8) is based do not apply. In any case, minimization of the mean square error does not necessarily maximize the subjective quality of the received signal. In order to determine the subjective effect of some of the system parameters on speech, the listening experiments in Chapter IV were conducted. The analytical results derived in this chapter were used as a guide in selecting the values of the constraining parameters.

III. A REAL-TIME DPCM SYSTEM

3.1 Introduction

The experiments described in Chapter IV were conducted under the assumption that the channel noise is negligible in comparison with the quantization noise. If the channel noise in Fig. 2.2(a) is neglected, the system shown in Fig. 3.1 gives an output $\tilde{x}(t)$ identical to the output $\hat{x}(t)$ in Fig. 2.2(a).

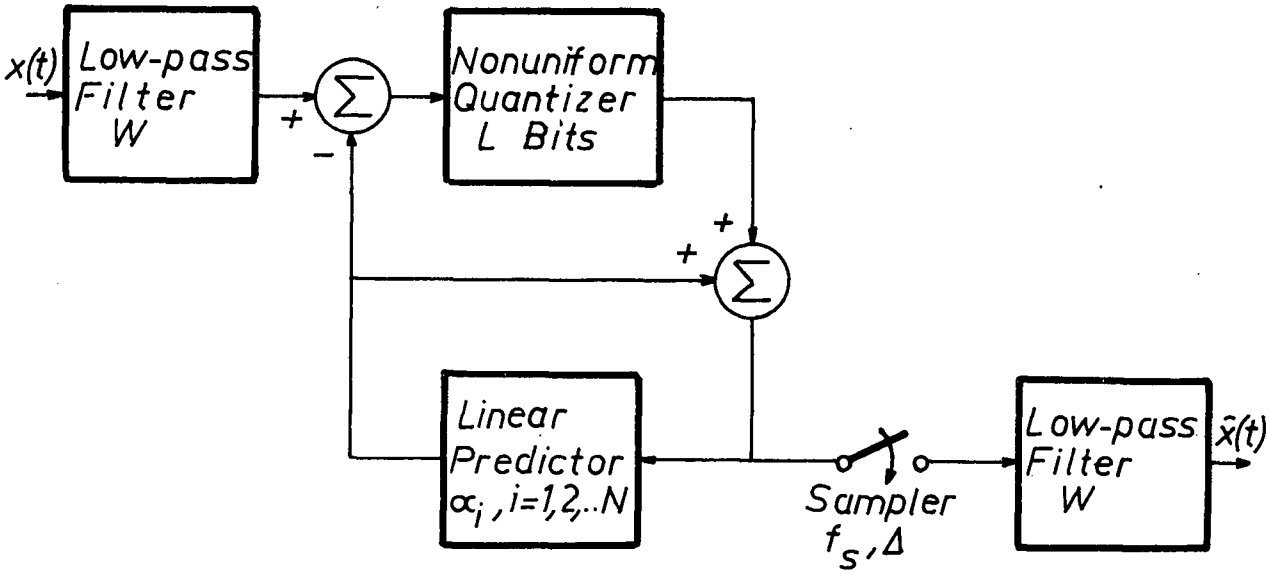


Fig. 3.1 A system equivalent to Fig. 2.2(a) when channel noise is neglected.

In this chapter a practical realization of the system

shown in Fig. 3.1 for N=1 is described. A discussion of the basic operation of the system is presented first, followed by a description of the constituent circuit blocks. It is shown that the implementation of the DPCM system in Fig. 2.2(a) requires only minor alterations to the system described in this chapter.

3.2 General Description of the System

A block diagram of the practical realization of Fig. 3.1 for N=1 is shown in Fig. 3.2. Consider initially that the input feedback sample and hold (S&H) of Fig. 3.2 has just stored

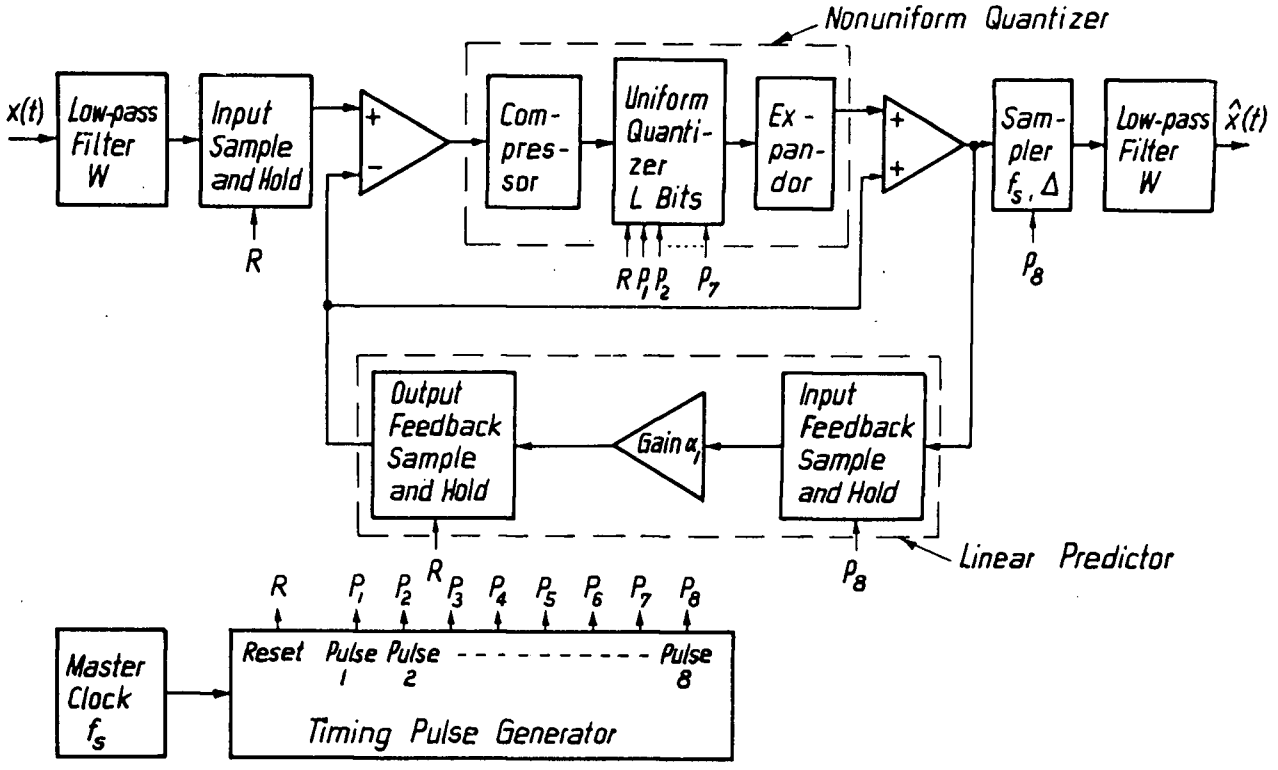


Fig. 3.2 Implementation of Fig. 3.1 when N=1 (previous-sample feedback DPCM).

the latest value quantized. The system remains in a quiescent

condition until the arrival of the next pulse from the master clock. The presence of a positive-going clock signal triggers a chain of monostables in the timing pulse generator, and initiates the periodic timing sequence; reset, pulse 1, pulse 2,.... pulse 8.

The reset pulse actuates the input S & H and the output feedback S & H. The input S & H samples the incoming low-pass filtered speech and stores the resulting value. At the same time, an estimate of this value based on the latest quantized value held in the input feedback S & H is sampled and stored in the output feedback S & H.

The actual and predicted values are subtracted, and the difference signal goes to the nonuniform quantizer for conversion to discrete format. Nonuniform quantization is achieved by compression, uniform quantization, and expansion under the control of the timing pulses 1 through 7.

After quantization, the difference signal is added to the predicted value of the input sample, and the resulting sum is sampled and low-pass filtered. The low-pass filter output $\hat{x}(t)$ is a replica of the original speech signal $x(t)$. The sum is also conveyed to the input feedback S & H where it is stored. The system now reverts to its quiescent state and awaits the arrival of the next positive-going master clock pulse before executing another cycle of operation.

System parameters W, f_s, L and α_i are variable, as are the compressor and expandor characteristics. To investigate the effects of channel noise, minor modifications and

additions are made to the system in Fig. 3.2, and the resulting system is used in conjunction with a general-purpose or special-purpose digital computer which simulates the digital channel (see Fig. 3.3). The remainder of the chapter contains a description of the major circuit blocks used to process the signal. Detailed circuit diagrams are given in the Appendix.

3.3 Timing Pulse Generator

A functional schematic of the timing circuitry is shown in Fig. 3.4(a). The timing pulse generator provides a pulse sequentially on each of nine outputs. It is triggered by a variable-rate pulse generator. The rate is determined by the sampling frequency f_s of the system. The output pulse durations are approximately 200 nsec except for the pulse 8 and the reset pulse. Pulse 8 and the reset pulse are approximately 2 μsec in duration to provide adequate acquisition time for the sample and hold circuits. The interval between timing pulses is determined by the propagation and settling times of the circuit blocks shown in Fig. 3.2.

3.4 Low-Pass Filter

The functional blocks that constitute the low-pass filter are shown in Fig. 3.5(a). The passive network is isolated by a pair of amplifiers in order to provide the terminating impedances required by the L-C network. The gain of the

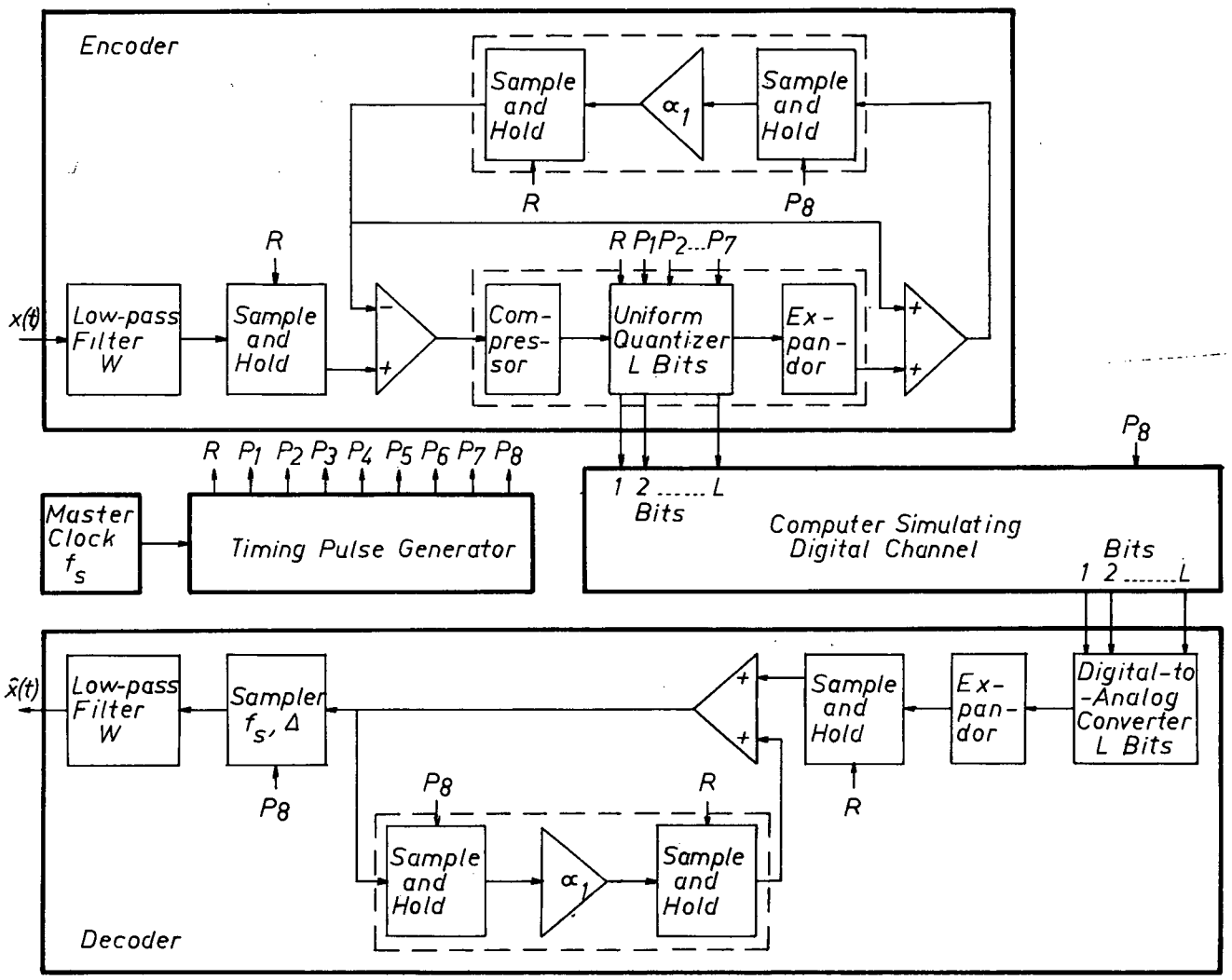
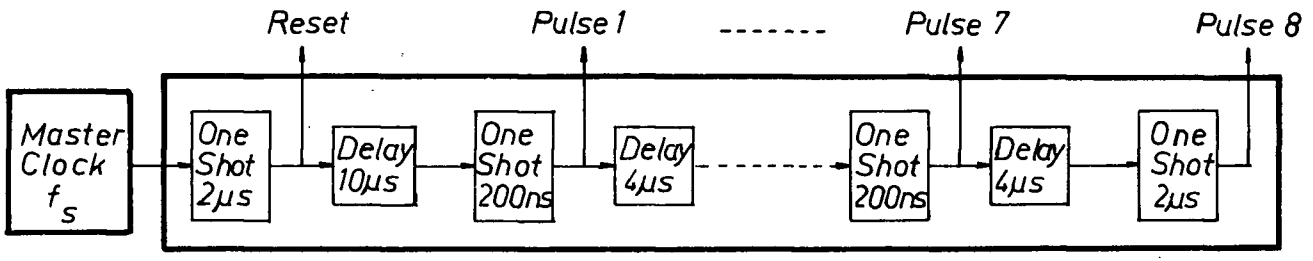
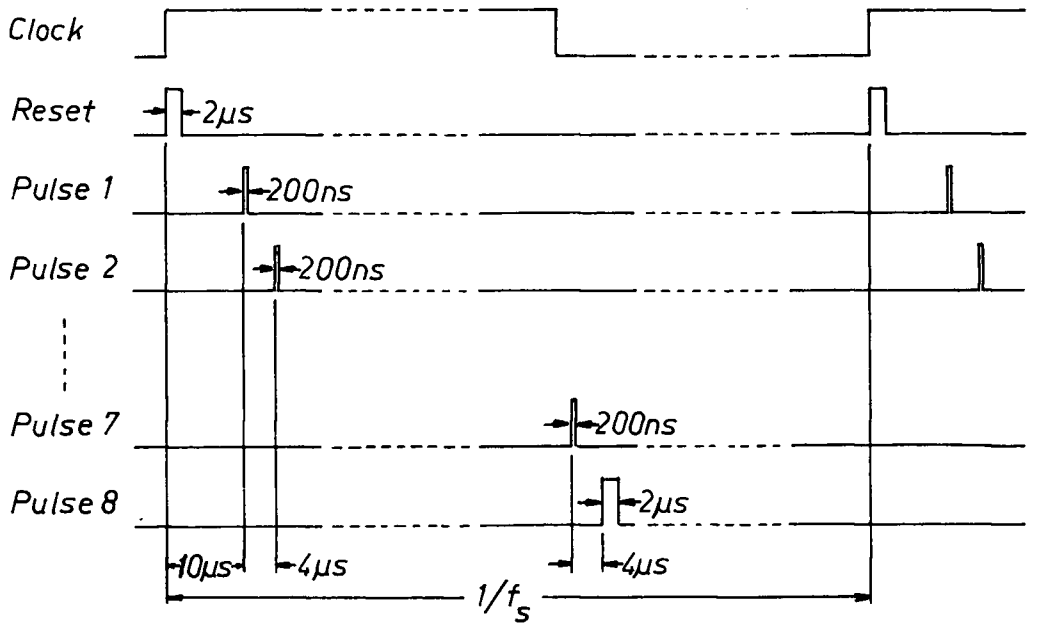


Fig. 3.3 Implementation of Fig. 2.2(a) when N=1 (previous-sample feedback DPCM).

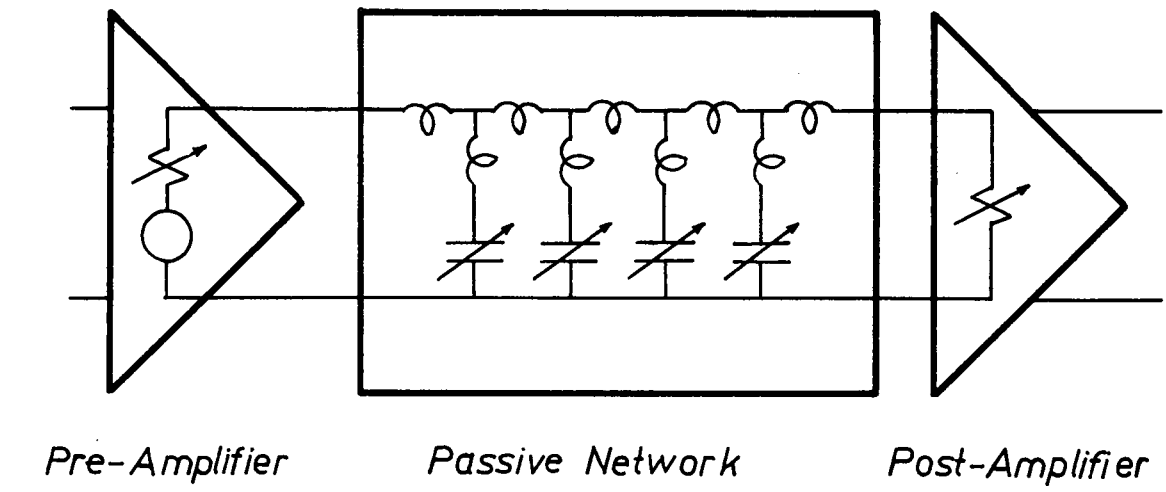


(a)

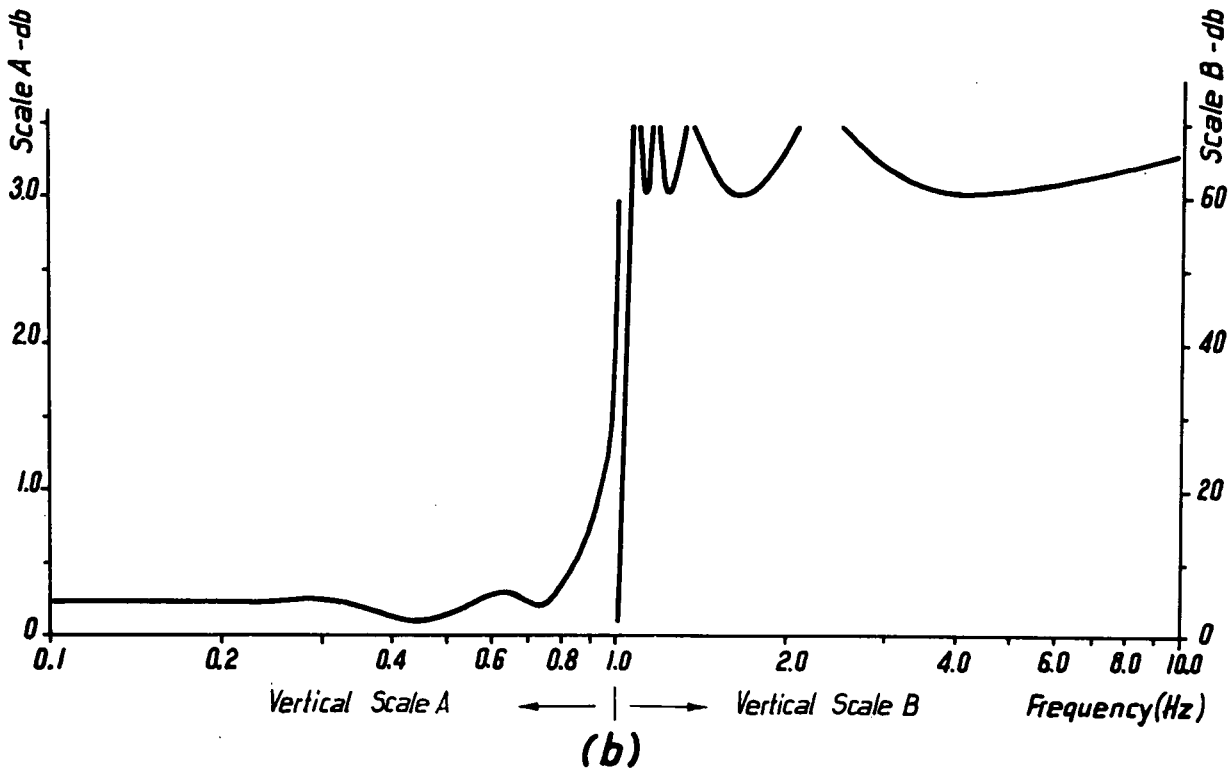


(b)

Fig. 3.4 (a) Timing pulse generator. (b) Timing pulse sequence.



(a)



(b)

Fig. 3.5 (a) Low-pass filter. (b) Normalized frequency response of low-pass filter.

amplifier is adjusted to compensate for filter insertion loss, as well as for signal loss encountered in the sampling process.

In order to vary the cut-off frequency of the filter, the capacitors were frequency and magnitude scaled in such a way that the inductances remained unaltered. The terminating impedance was then adjusted by varying the output and input impedance of the pre- and post-amplifiers. The normalized frequency response of the low-pass filter is shown in Fig. 3.5(b).

3.5 Sample and Hold

Fig. 3.6 illustrates the sample and hold process [26].

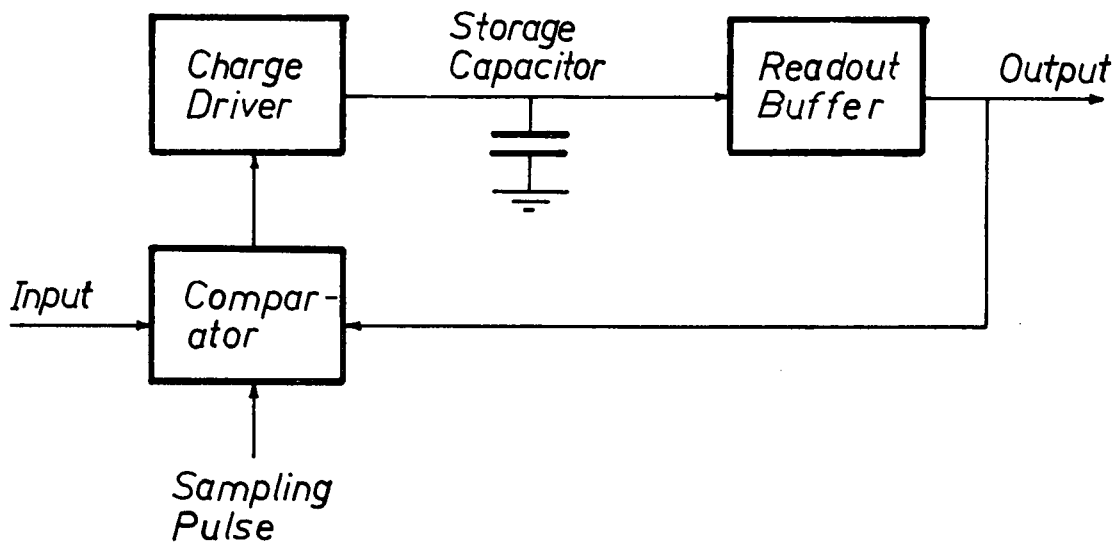


Fig. 3.6 Functional block diagram of sample and hold circuit.

The input voltage is applied to one input terminal of a difference amplifier comparator, and the output voltage is fed back to the other terminal. By gating the emitter current of the difference amplifier with the sampling pulse, the bipolar charge driver is connected to the storage capacitor. Current, controlled by the comparator error voltage, charges the storage capacitor during the sampling interval to make the readout voltage equal, both in magnitude and in polarity, to the input voltage. During the interval between sampling pulses, the charge driver is disconnected from the storage capacitor, and the capacitor holds the same voltage that it had at the instant the sampling pulse was removed.

The circuit diagram for the sample and hold appears in the Appendix. The circuit has a 2 μsec sampling time, a -5 to +5 volt dynamic range, a 10 mv absolute error for a hold time of ½ sec or less, and a DC accuracy of 10 mv.

3.6 Nonuniform Quantizer

Nonuniform quantization is achieved by inserting complementary nonlinear no-memory amplifiers before and after a uniform quantizer. For tapered quantization, preferential amplification of weak signals prior to uniform quantization (compression) and preferential attenuation after uniform quantization (expansion) is performed. The combined process is known as companding. A typical four-level tapered quantizer is shown in Fig. 3.7.

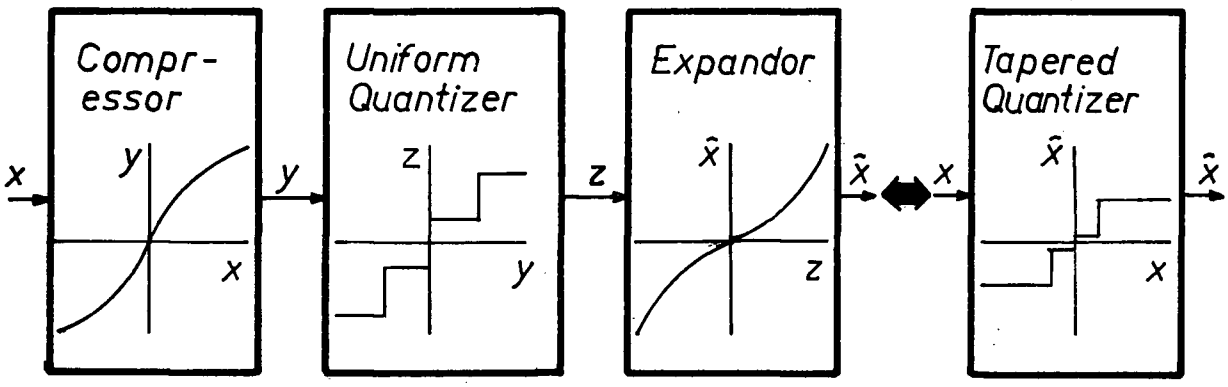


Fig. 3.7 Tapered quantizer.

3.7 Compressor and Expandor

The circuit blocks for the compressor and expandor are shown in Fig. 3.8. In either case, a nonlinear network follows an appropriate pre-amplifier. The network provides the desired nonlinear characteristic, and the pre-amplifier provides the required current and voltage gains.

For tapered quantization, the compressor or expandor characteristic is roughly similar in overall appearance to the inherent logarithmic voltage-current characteristic of semi-conductor diodes [27]. If these companding diodes are preceded by a piecewise-linear no-memory network, then any practical companding characteristic can be synthesized.

In Fig. 3.8 the response of the piecewise-linear no-memory network is controlled by the bridge circuits. The orientation of the diodes and the reference voltages determine the break points, while the potentiometer settings determine

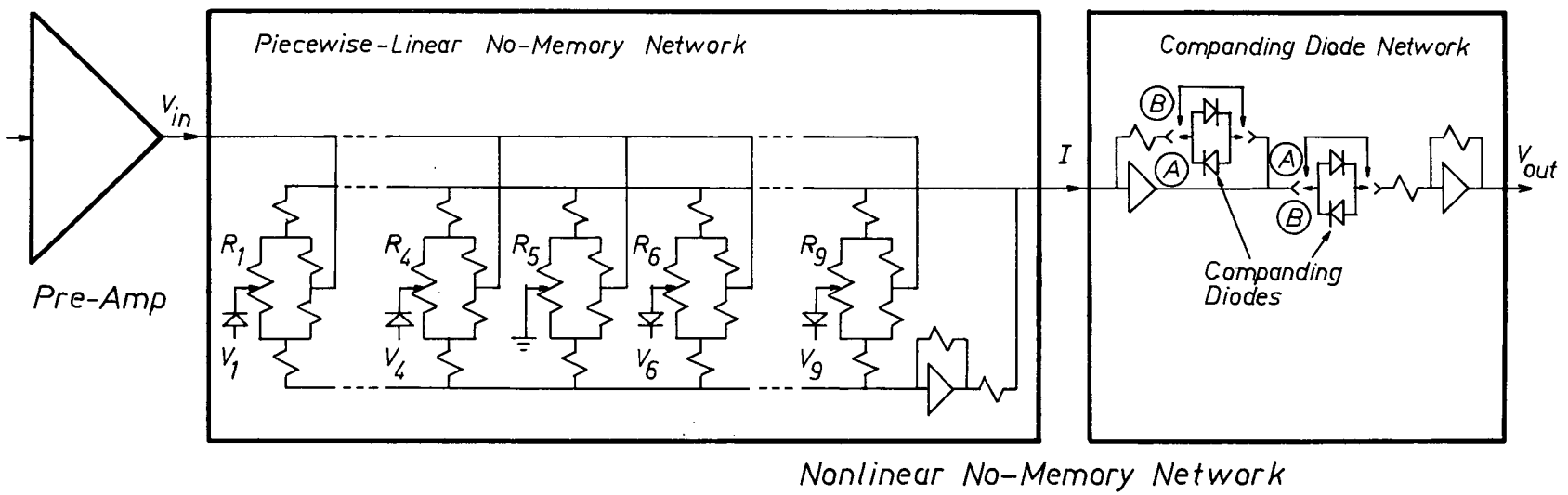


Fig. 3.8 Compressor or expandor.

the slopes. If the breakpoint diodes are ideal, then a typical piecewise-linear transfer characteristic is as shown in Fig. 3.9. Angles $\theta_1, \theta_2, \ldots \theta_9$ are adjusted by potentiometers

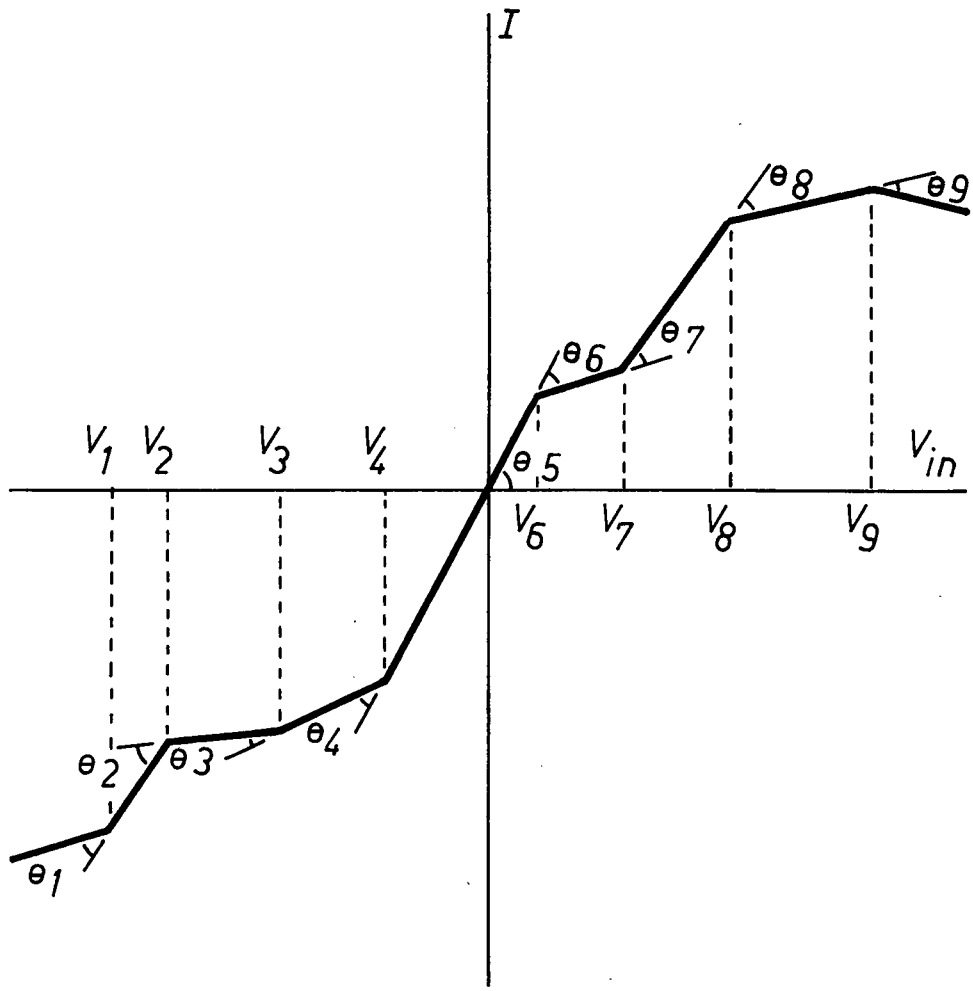


Fig. 3.9 Typical response of the piecewise-linear network in Fig. 3.8.

$R_1, R_2, \ldots R_9$, respectively.

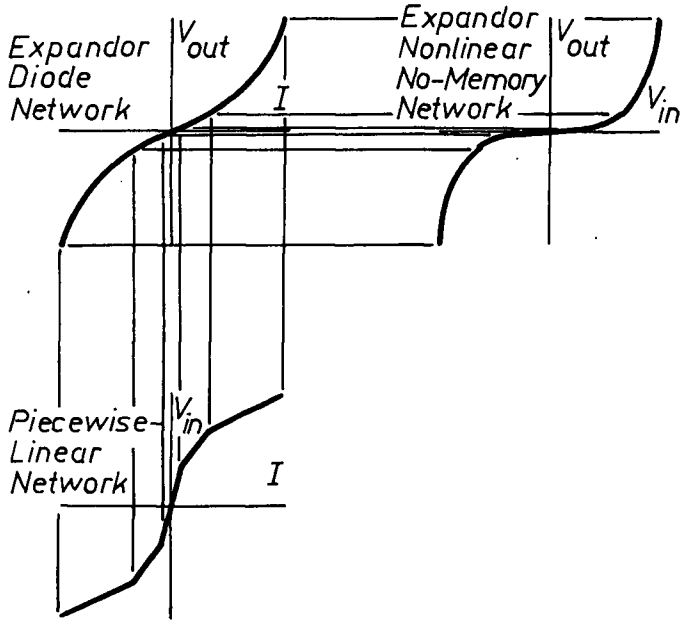
When the network of Fig. 3.8 is operated as a compressor,

option (A), the companding diodes are located in the feedback path of an operational amplifier in the companding network. The inverse or expansion network results when the companding diodes are located in the operational amplifier input path, option (B) . Typical curves for the piecewise-linear network and the companding diode network, as well as for the completed nonlinear no-memory network, are shown in Fig. 3.10.

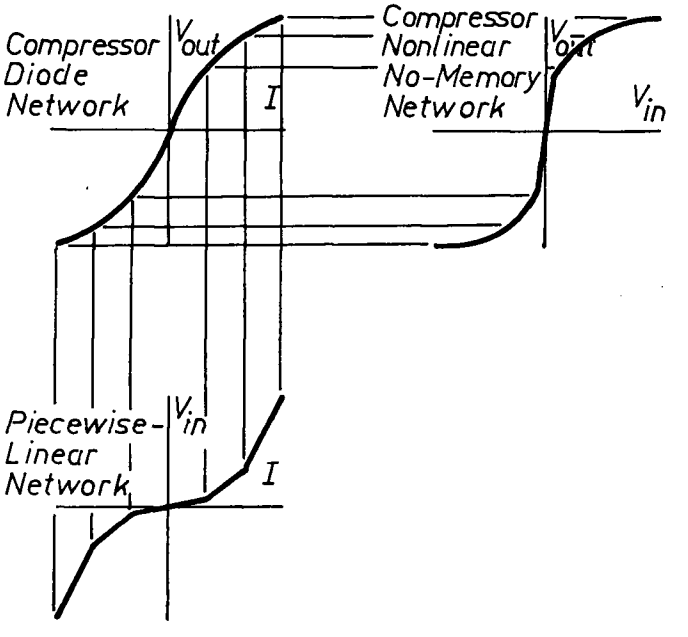
3.8 Uniform Quantizer

The uniform quantizer is a successive approximation analog-to-digital (A/D) converter which repeatedly divides the input voltage range in half [28] . The quantizer, shown in Fig. 3.11 consists basically of a digital-to-analog (D/A) converter, a comparator, a control logic circuit, and a delay. The number of bits used in the conversion is selected by switches S_2 through S_7. For example, if L bits are required, then switches $S_2, \ldots S_L$ are closed.

At the beginning of the conversion, the reset pulse sets a 1 in the flip flop (FF) representing the most significant bit, (FF1), and a 0 in all flip flops corresponding to bits of lesser significance, (FF2,FF3,...FF7). The comparator output now determines whether the next digital approximation should be larger or smaller than the first one. After the D/A converter and the comparator have settled, the next pulse, pulse 1, enters the control circuit. Pulse 1 sets FF2,which corresponds



(a)



(b)

Fig. 3.10 Typical transfer characteristics for the networks in Fig. 3.8.

(a) Expandor. (b) Compressor.

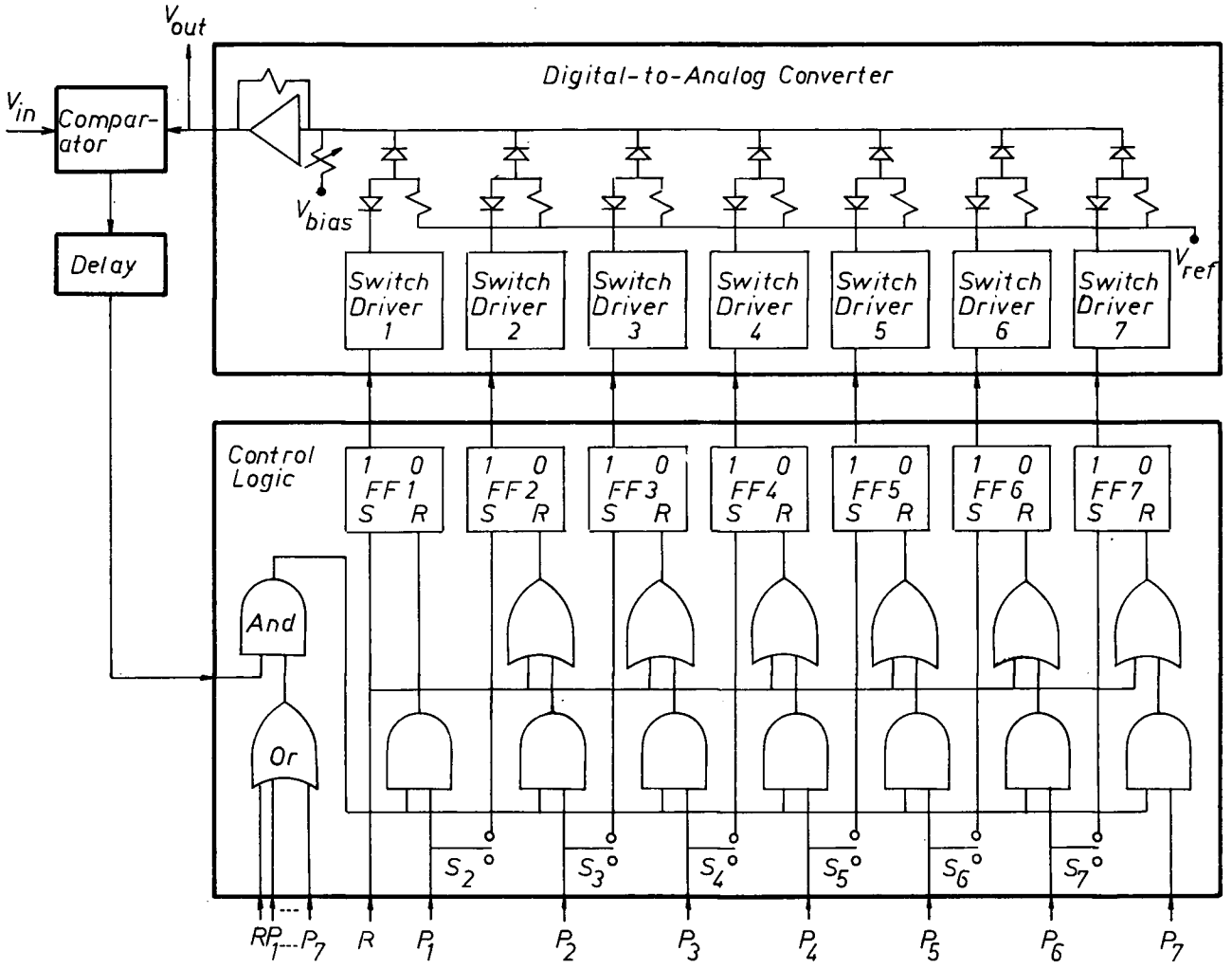


Fig. 3.11 Uniform quantizer.

to the next most significant bit, to a 1. FF1 either remains in the 1 state or is reset to 0, depending on the comparator output. The procedure is repeated until the final L-bit approximation of the input has been obtained.

3.9 Sampler

A schematic of the sampler is shown in Fig. 3.12. It

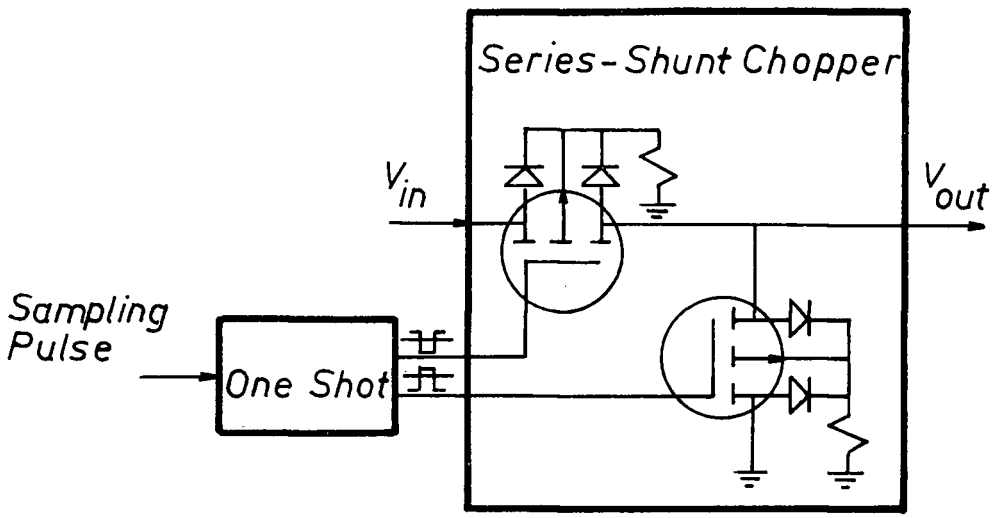


Fig. 3.12 Sampler.

consists of a one-shot multivibrator, and a series-shunt chopper in which P-channel insulated gate field-effect transistors (IGFET) are used as the series and the shunting devices. The gate drives for the IGFET pair are obtained from complementary outputs of a one-shot multivibrator. During the sampling interval, the series IGFET is turned on, and the shunting IGFET turned off.

IV. SUBJECTIVE TESTS AND RESULTS

4.1 Introduction

In an attempt to obtain quantitative subjective assessments of various sensory stimuli, numerous psychological scaling methods have been developed [29]. In this thesis, a modification of the paired comparison method was employed to determine equal preference (isopreference) contours. A version of the subjective-estimate method was then used to assign preference values to these contours.

Although either method could have been used to determine and to rate the isopreference contours, a more reliable evaluation method is attainable by using the two methods in combination. The paired comparison method based on simple A-B preference judgments is a reliable method [30] for obtaining isopreference contours, but unless variability in listeners' judgments is substantial, the rating scale derived from this method is questionable. Since it was found that the psychological distances separating the isopreference contours are large in comparison with the listeners' variability, the subjective-estimate method was used for rating the isopreference contours. In contrast to the paired comparison method, the reliability of the subjective-estimate method for obtaining isopreference contours is questionable, since this method requires that the listener be able not only to judge which

stimulus is better, but also to indicate how much better one stimulus is with respect to the other.

In this investigation, the following assumptions and restrictions were imposed.

(1) The digital channel shown in Fig. 1.2 provided error-free transmission.

(2) The nonuniform quantizer was logarithmic with μ=100. Smith [22] showed that such a quantizer was desirable when the message was speech. The characteristic of a 3-bit μ = 100 logarithmic quantizer is shown in Fig. 4.1.

(3) The sampling frequency was constrained to equal 2.2 times the 3 db cutoff of the low-pass filter in order to eliminate aliasing errors in PCM and DPCM, and idle channel oscillations in DPCM [16]. If the filter is an ideal low-pass, then sampling at the Nyquist rate is sufficient to eliminate these effects.

(4) Single-sample feedback was used in the DPCM system. For minimum mean square error the prediction coefficient α_1 was determined from equation (2.8) to be equal to the normalized autocorrelation of the bandlimited speech signal evaluated at the sampling period. However, since equation (2.8) is derived under the assumption that quantization is relatively fine, and since autocorrelation of filtered speech is not known exactly, an approximate relation for α_1 which depended only on the sampling period $1/f_s$ was used. The dependence of α_1 on $1/f_s$ is shown in Fig. 4.2.

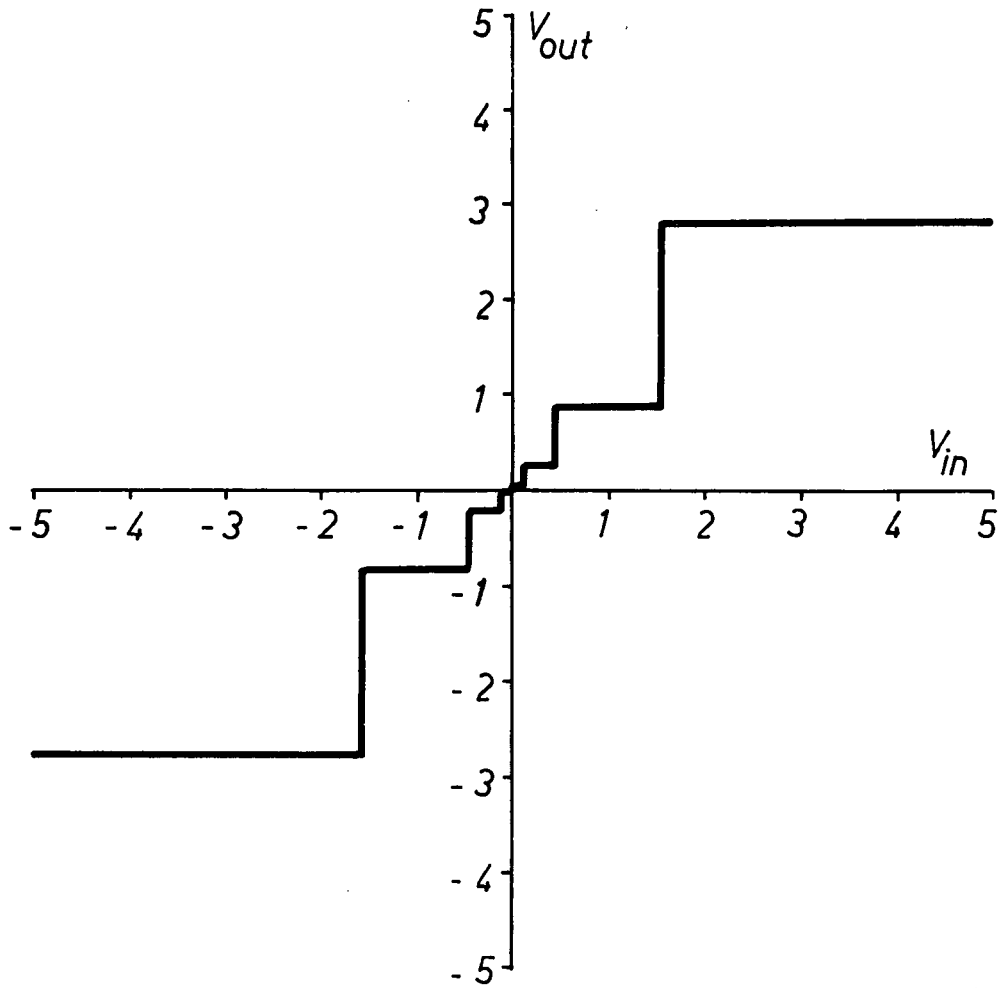


Fig. 4.1 Transfer characteristic of an eight-level, μ=100 logarithmic quantizer

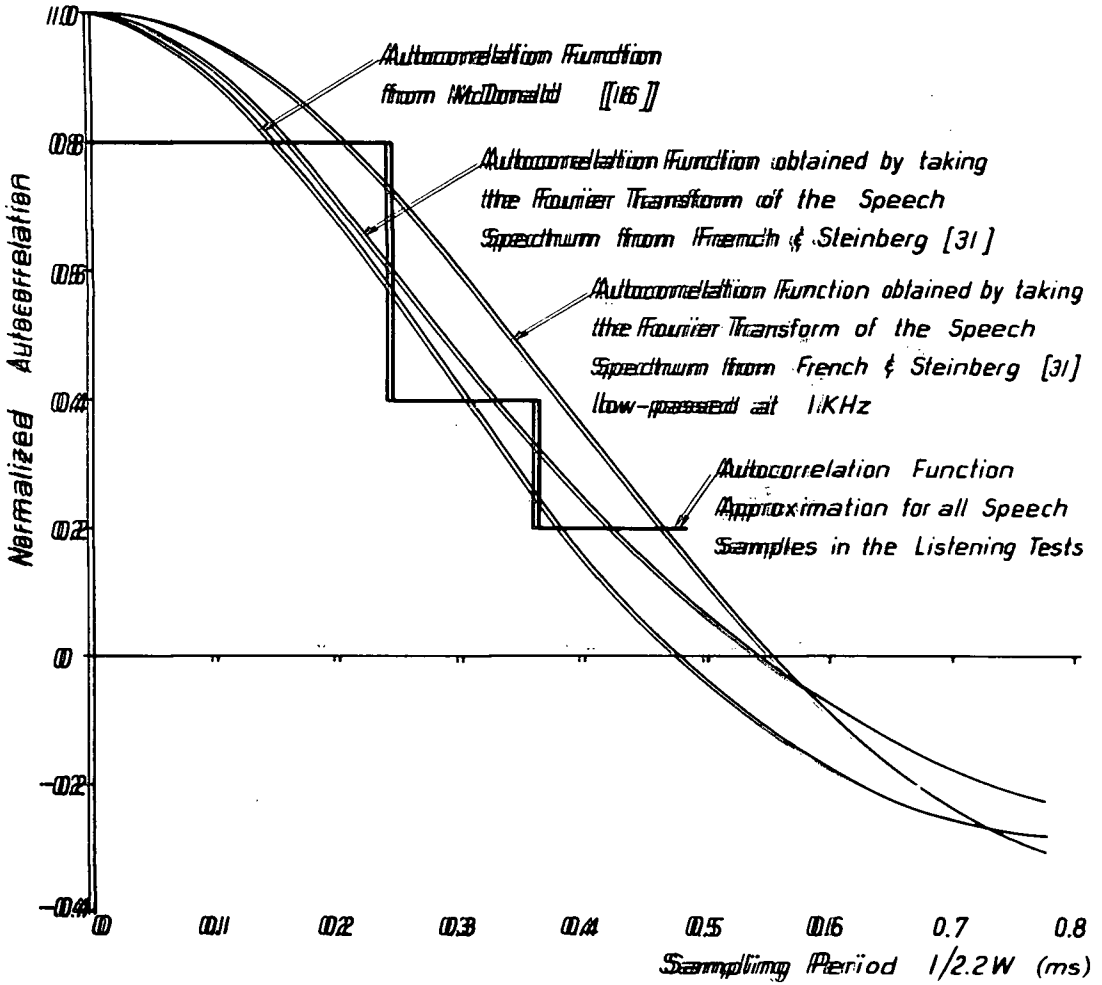


Fig. 4.2 Autocorrelation functions of low-pass filtered speech vs. 1/2.2W. W is the speech bandwidth.

(5) The input signal to the quantizer was constrained to occupy the full range of the quantizer input.

(6) Speech bandwidth W and number of bits of quantization L were variable in discrete steps. The values that W could have were 1.01, 1.23, 1.55, 2.12, 2.63, 3.17, 4.2 and 6.3 KHz. L took on integer values from 1 to 7, inclusive.

4.2 Preparation of the Speech Samples for the Listening Test

The first step in the listening test was to record a master sentence in an anechoic chamber, using a high quality recording system. The system consisted of a General Radio Type 1560-P3 PZT microphone and a Tandberg 64X tape-recorder. The sentence "Joe took father's shoe bench out" was chosen as the master sentence, because it contains most of the phonemes and has a frequency spectrum that is representative of conversational speech [31]. This sentence, spoken by a 28 year old male with a Western Canadian accent, was used in the preparation of all speech samples. An estimate of the speaker's spectrum obtained by measuring the spectral energy in eight adjacent frequency bands is shown in Fig. 4.3.

The speech samples were obtained by playing back the master sentence through either a PCM or a DPCM system and recording the system output on a second Tandberg 64X tape-recorder. The samples were then spliced together along with suitable lengths of non-magnetic tape to form a test sequence.

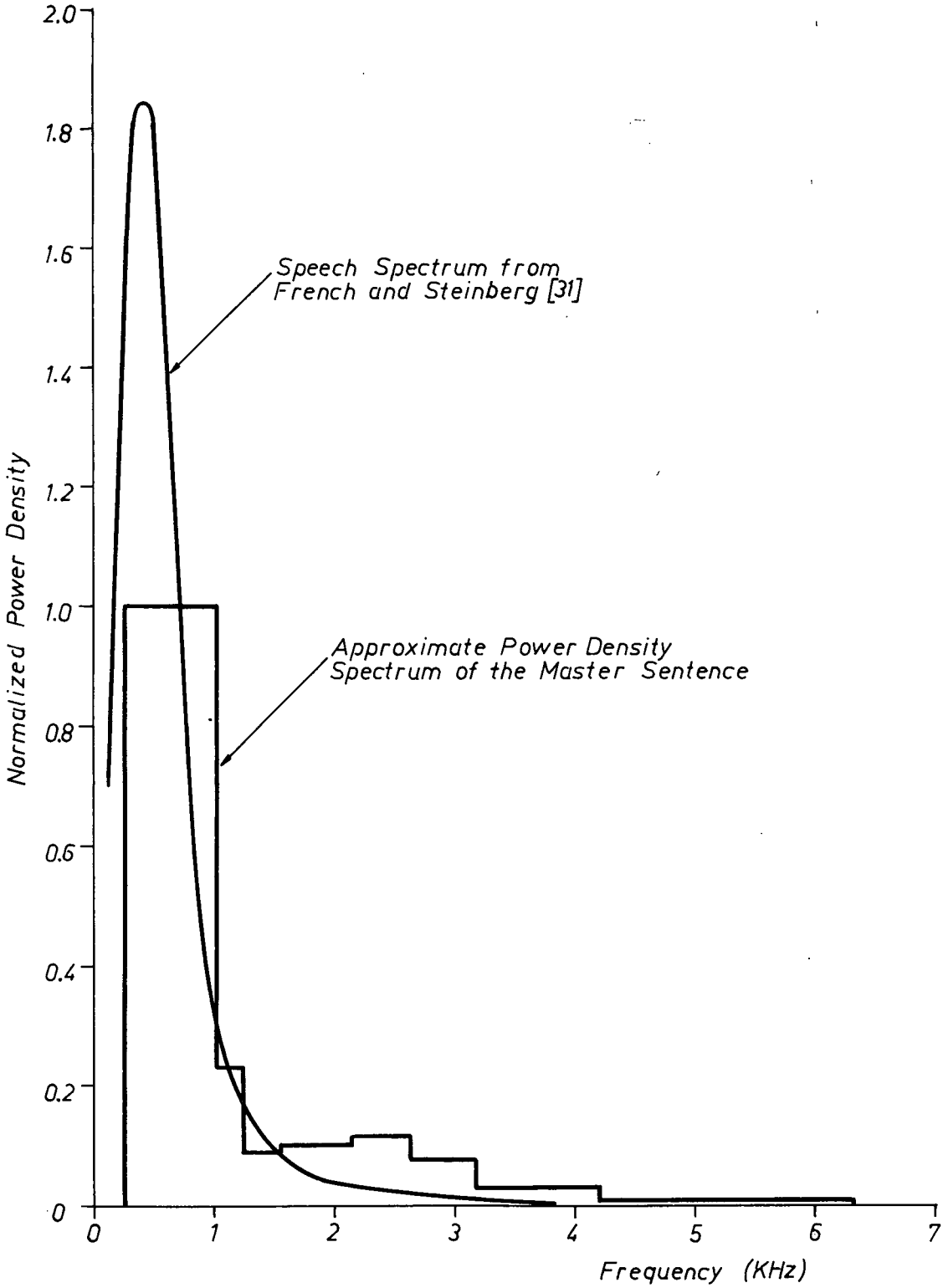


Fig. 4.3 Power density spectra of speech samples.

All tape playbacks were high-pass filtered at approximately 200 Hz to eliminate tape-recorder hum.

4.3 Paired Comparison Test for Determining Isopreference Contours

For the paired comparison test, three listening sessions were conducted: one in the morning and one in the afternoon of the same day, and one in the morning of the following day. Before each session began, the listeners received response forms with the following written instructions:

"In this test you will hear pairs of sentences; each pair is separated by a 5 second rest period. After listening to a pair, specify which sentence you would prefer to hear. If both sentences sound equally good, make an arbitrary choice. The first sentence of each pair is sentence A, and the second is sentence B."

Sentences A and B were separated by a one second silent interval. During each 25 minute session, sixty randomly ordered pairs (all different) were heard by each listener. After every twenty comparisons, the listeners were given a two to three minute rest. All listeners knew the speaker of the sentences.

The listening sessions were conducted in a quiet room. Binaural listening with Pioneer Model SE-1 stereo headphones was used in all sessions. The listeners were 12 graduate students and 2 Staff Members of the Electrical Engineering Department of U.B.C. Their ages ranged from 23 to 39

years, with a mean of approximately 27 years. All listeners showed no hearing abnormalities, and had little or no previous experience in listening tests. Ten listeners were present at each listening session. Although it was intended that the same group of ten would serve as listeners for all the comparisons, absence and prior commitments forced substitution.

4.4 Results of Paired Comparison Test

Isopreference contours connecting points of equal subjective quality on the W-L plane appear in Figs. 4.4 and 4.5.

As an example of how the isopreference points were obtained from the results of the paired comparison test, consider the points marked A and B on Fig. 4.4. PCM speech samples having W and L values of point A were compared with four other PCM samples at random times during the listening experiment. The four other samples all have the same bandwidth (3.17 KHz), but different number of quantization bits. The results of the four comparisons have been plotted in Fig. 4.6(a). The ordinate of Fig. 4.6(a) shows the percentage of judgments in which the listeners prefer system A to the system having a bandwidth of 3.17 KHz and number of quantization bits defined by the abscissa of the plot. The range in number of quantization bits was chosen to be large enough so that the

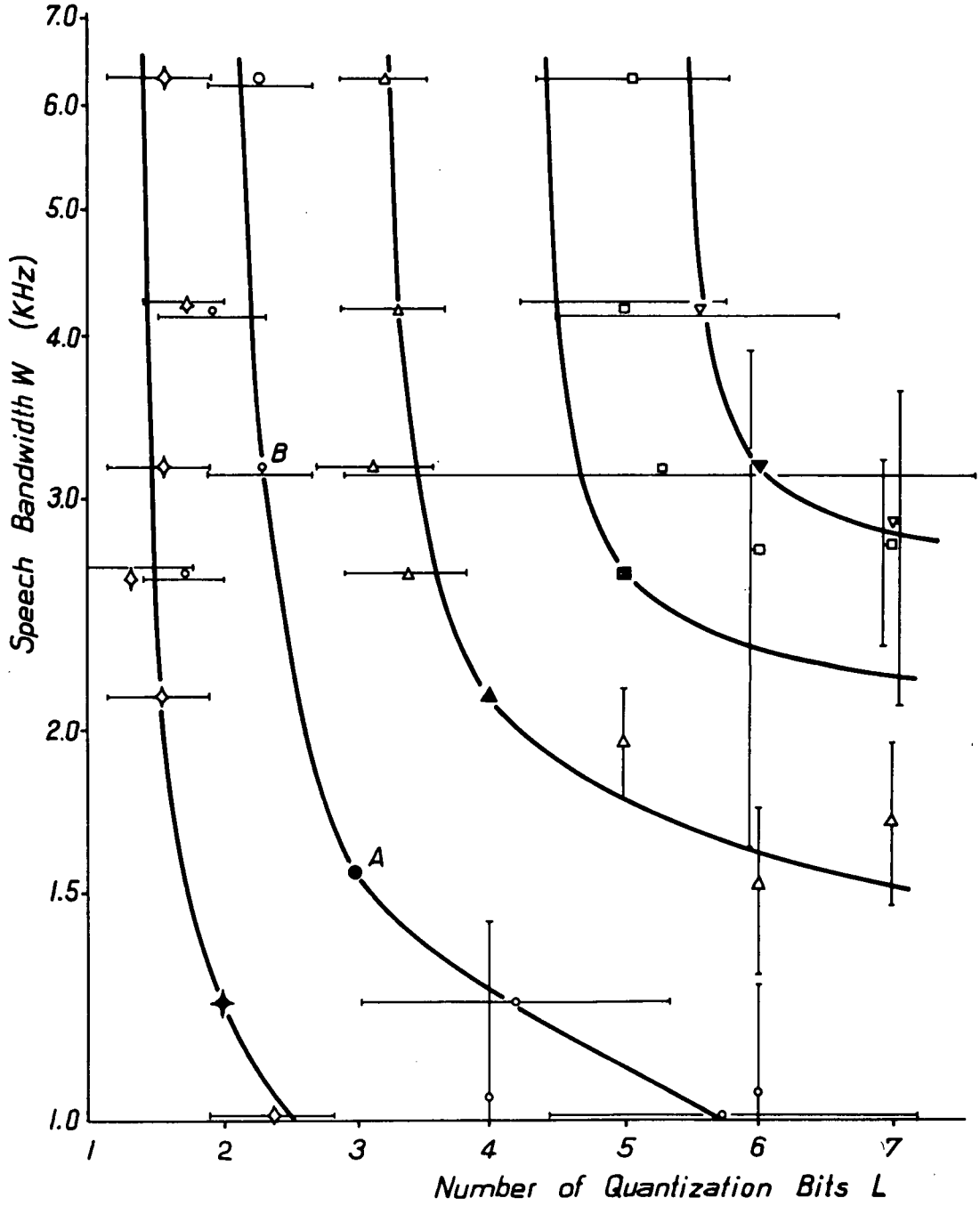


Fig. 4.4 Isopreference contours for PCM. The length of each bar equals the standard deviation of the experimental point. The length is measured with respect to the scale on the co-ordinate axis parallel to the bar. The reference points for each contour are drawn solid.

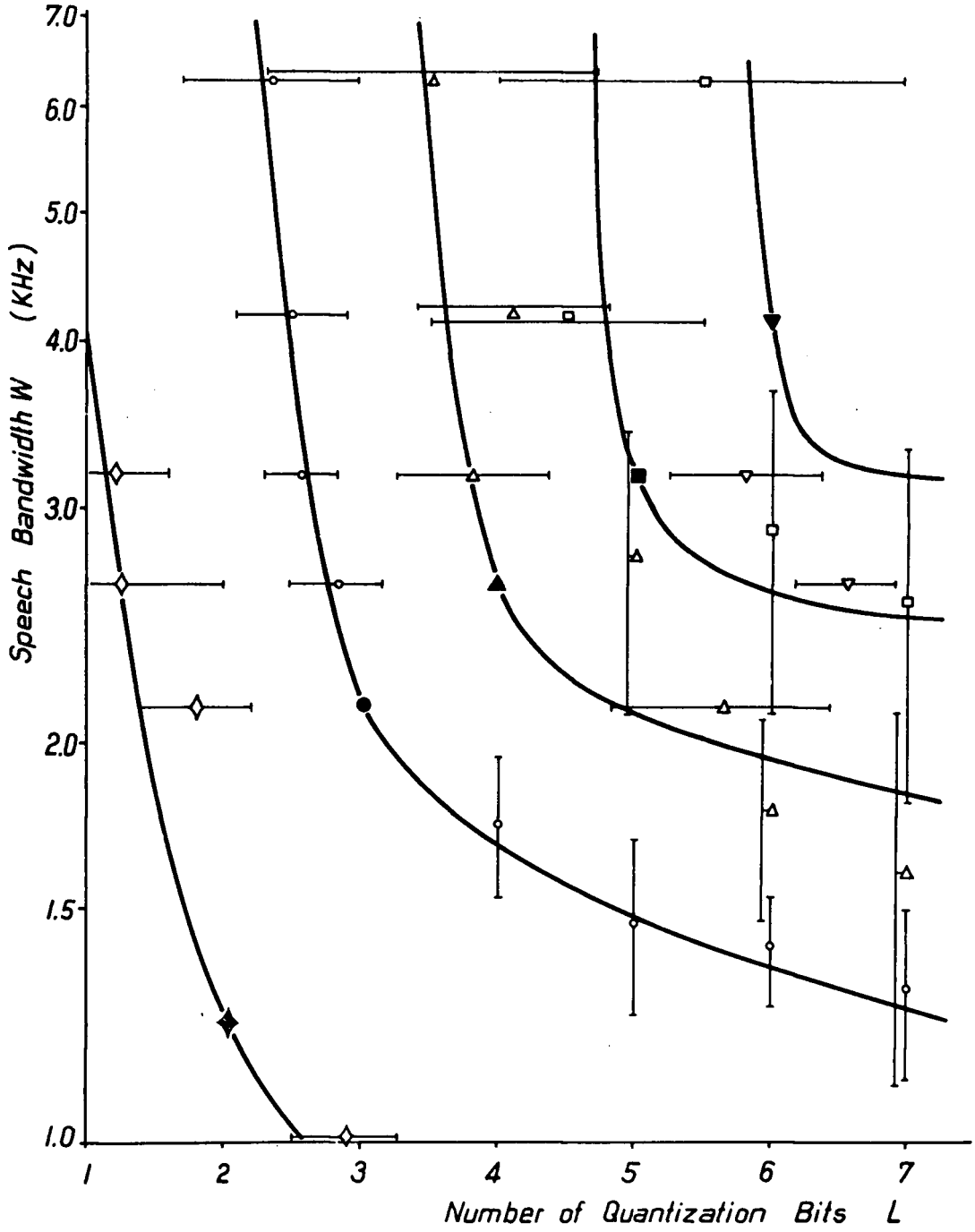


Fig. 4.5 Isopreference contours for DPCM. The length of each bar equals the standard deviation of the experimental point. The length is measured with respect to the scale on the co-ordinate axis parallel to the bar. The reference points for each contour are drawn solid.

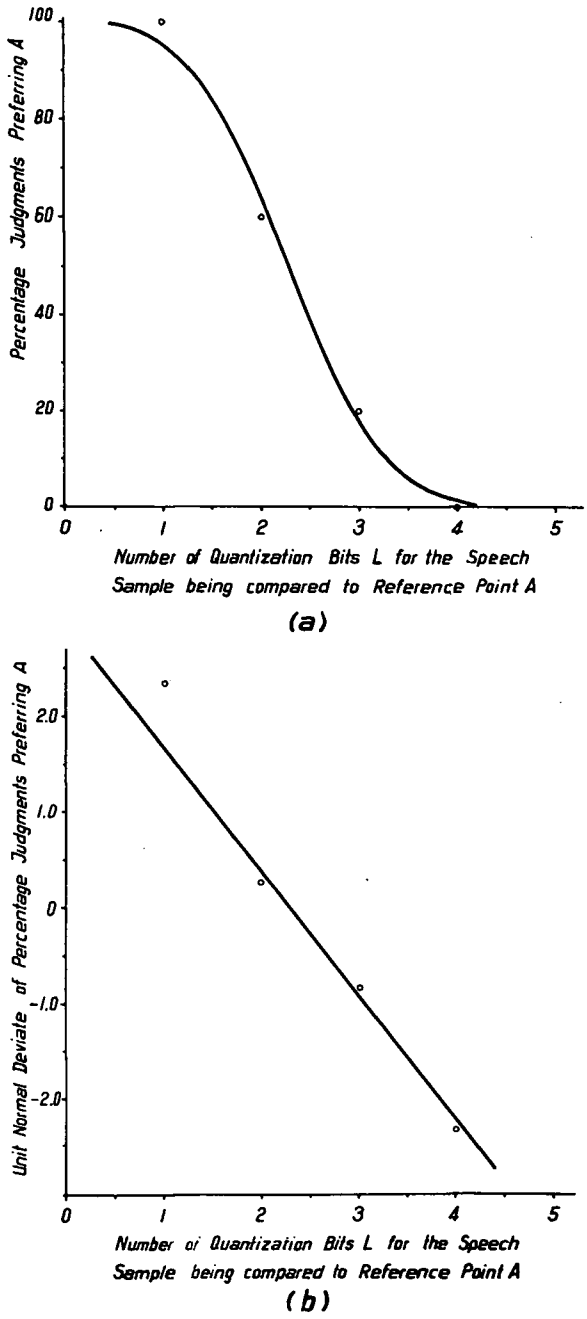


Fig. 4.6 Psychometric curve for obtaining point B in Fig. 4.4. The reference point is point A.

(a) Ordinate in linear preference units. (b) Ordinate in unit normal deviates.

preference judgments would vary from 0 to 100%. From the smooth psychometric curve drawn through the experimental points, the 50% or equal preference point was obtained, and the corresponding abscissa value was used to define point B on Fig. 4.4.

Points A and B are now assumed to have equal subjective preference, and speech samples corresponding to either point could be used for comparison with other speech samples to find additional isopreference points. However, number of bits of quantization is not a continuous parameter, but assumes integer values. For this reason, speech samples corresponding to point A were used as the reference samples in determining all the isopreference points of the contour. For other isopreference contours, the speech samples corresponding to the solid points shown in Figs. 4.4 and 4.5 were chosen as the reference samples.

The expected shape of the isopreference contours determined whether the independent variable was speech bandwidth or number of quantization bits. When it was expected on the basis of pilot tests that L was critical and W was relatively unimportant, as for point B, the comparison tests were made with L as the variable. In the lower right half of the W-L plane, W was the more critical variable.

In plotting all psychometric curves, it was found that a normal distribution curve fitted the data points. For this reason, the proportion of listeners preferring the reference speech sample was converted to unit normal deviates.

Since unit normal deviates of 1.00 and 0.00 are infinite, all proportions of 1.00 and 0.00 were changed to 0.99 and 0.01 respectively, before being converted. With the aid of a digital computer, a weighted least squares technique was used to fit a straight line to the data points. Fig. 4.6(b) shows the psychometric curve of Fig. 4.6(a) plotted in unit normal deviates. The abscissa value for zero unit normal deviates is the equal preference point. The reciprocal of the slope of the line is equal to the standard deviation σ of the points fitted by the line, and is a measure of the variability of the isopreference point. The standard deviation of each isopreference point shown in Figs. 4.4 and 4.5 is represented by the length of the line drawn through the point. The length is measured with respect to the scale on the coordinate axis parallel to the line. In drawing the isopreference contours, the variance associated with each point was considered. The isopreference contours were drawn close to points with small variance and were constrained to have the same general shape as their neighbouring contours.

4.5 Experimental Procedure for Scaling Isopreference Contours

After the isopreference contours had been determined, a test based on the subjective-estimate method was devised to provide a scale value for each contour. In this method,

a subjective scale is constructed directly from the listener's own quantitative estimates.

Before part I of the two-part test began, the listeners received the following written instructions:

"In this test you will hear pairs of sentences; each pair is separated by a 5 second rest period. If zero denotes a sentence which is just unintelligible, and ten denotes the first sentence (sentence A) of a pair, how would you rate the second sentence (sentence B) on a (0 to 10) equal half interval scale?"

A one second silent interval separated the sentences in each pair. Unquantized speech samples, bandlimited to 6.3 KHz, were chosen as sentence A on the basis of pilot experiments which indicated that these samples were subjectively better than any of the sentence B samples to be rated. To prevent listener bias, the presentation order of the pair was reversed for part II. Sentences A were the samples to be rated and sentences B were the unquantized 6.3 KHz samples. Before part II of the rating test began, the listeners received instructions similar to those of part I.

In each part of the rating test, 21 different, randomly ordered speech samples were rated. After every tenth rating, the listeners were given a two to three minute rest. The entire test required approximately 20 minutes. Samples corresponding to the ten reference points shown in Figs. 4.4 and 4.5 were rated, along with five additional points in Figs. 4.4 and 4.5, and six unquantized, bandlimited speech samples.

To familiarize the listeners with the range in quality covered by the speech samples of the rating test, a practice session was conducted just before the test began. In the practice session, the listeners were required to rate five speech samples which were approximately equally spaced and extended over the full range of the subjective scale.

The test was conducted in a quiet room. Binaural listening with stereo headphones was used. The listeners were 10 members of the group of 14 male listeners that participated in the paired comparison test. Five of the listeners were given part I first, followed by part II. For the other five listeners the order was reversed.

4.6 Results of Rating Test

The scale value S assigned to each rated speech sample was taken as the mean of the listeners' ratings for the sample. The standard deviation σ of the listeners' ratings for the sample was used as the measure of variability of the scale value. S and σ for various points in the W-L plane are shown in Figs. 4.7 and 4.8. Also shown in Figs. 4.7 and 4.8 are lines of constant bit rate R and the isopreference contours of Figs. 4.4 and 4.5. Each isopreference contour was assigned the scale value of the reference point through which it passed. The rated points between the contours served as a check on the consistency of the rating method. The consistency was extremely good.

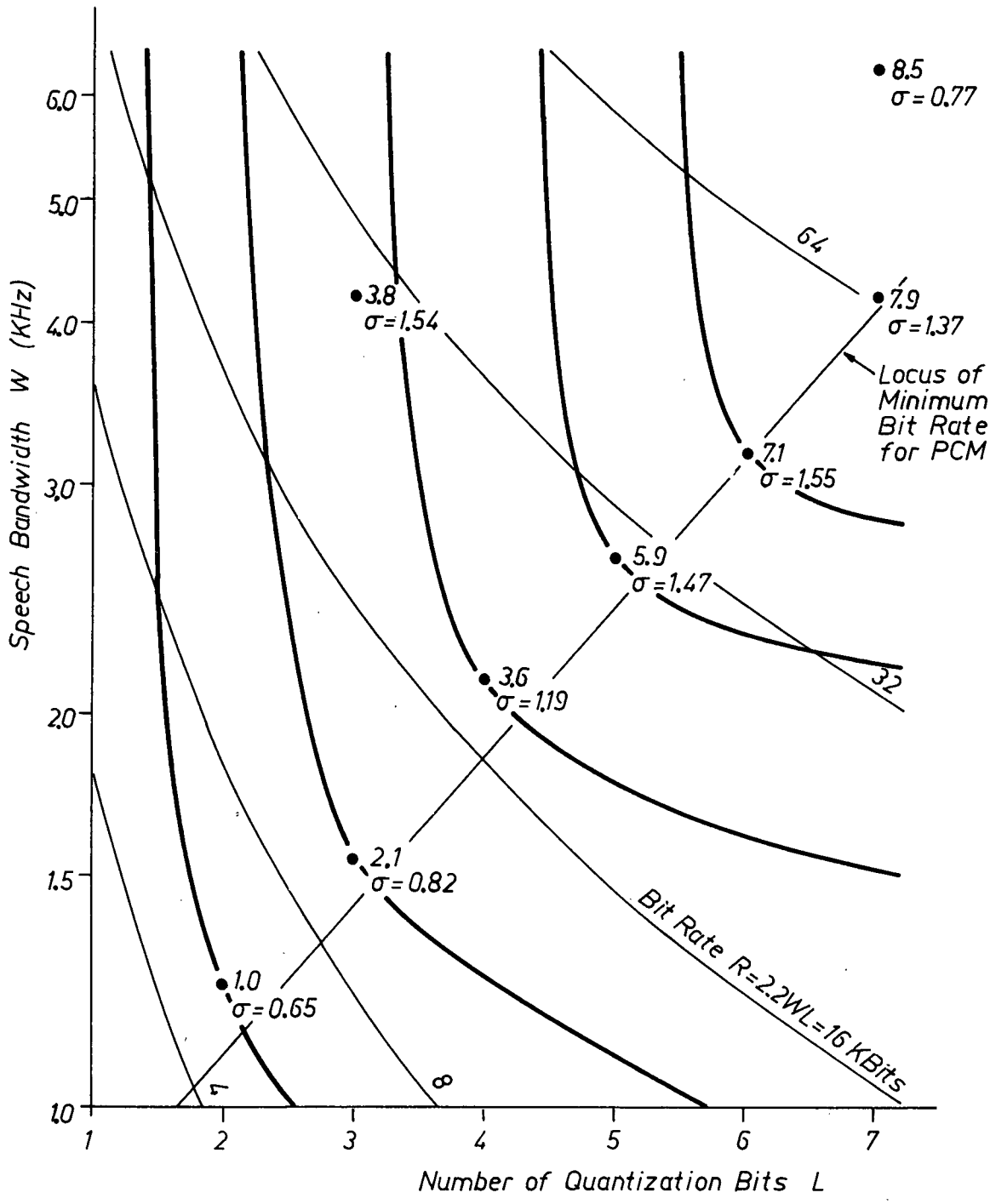


Fig. 4.7 Scaled isopreference contours for PCM, and curves of constant bit rate R=2.2WL. The standard deviation σ of each experimentally scaled point is shown.

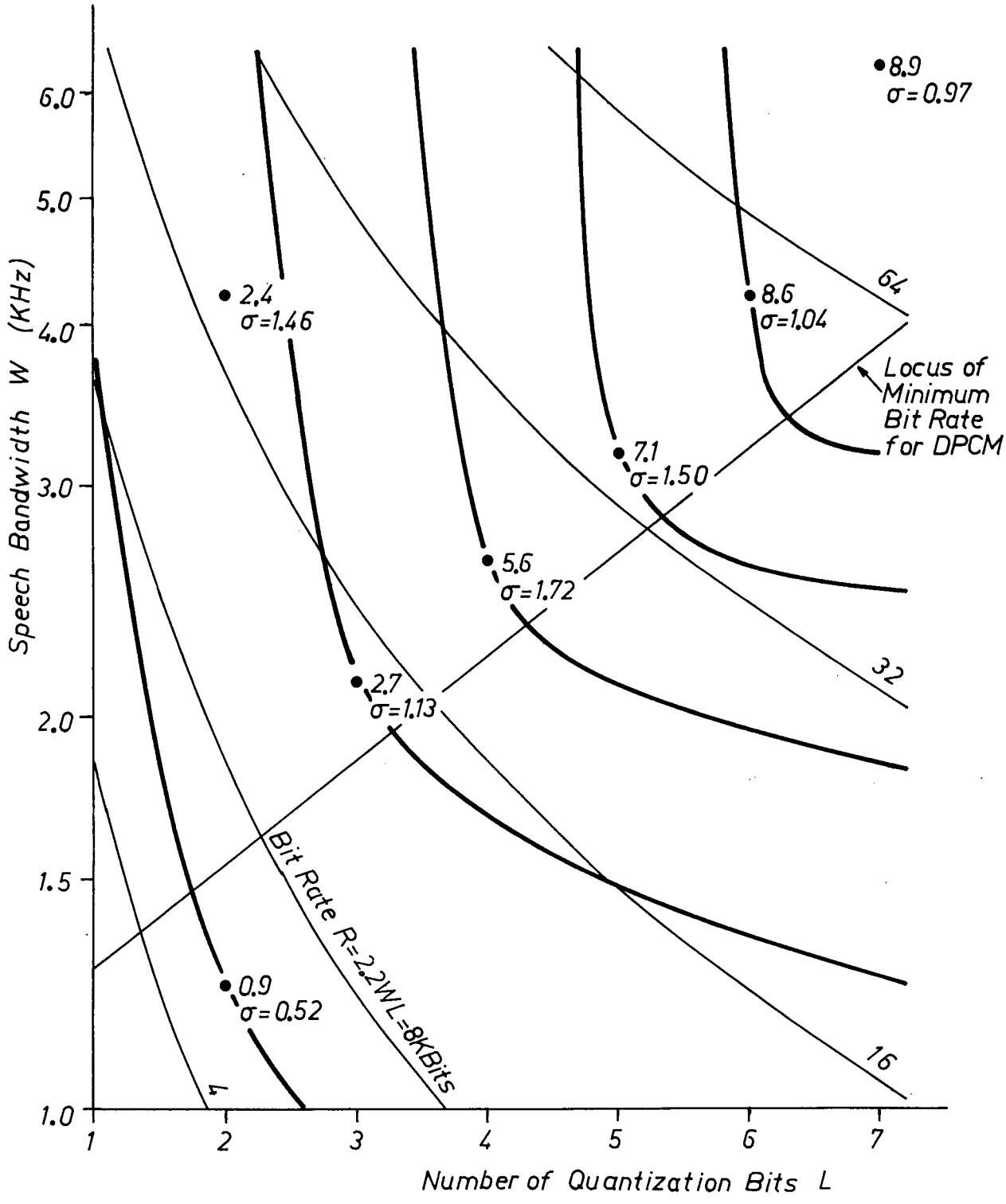


Fig. 4.8 Scaled isopreference contours for DPCM, and curves of constant bit rate R=2.2WL. The standard deviation σ of each experimentally scaled point is shown.

In Fig. 4.9, scale values are plotted versus speech bandwidths of the unquantized bandlimited samples used in the rating test. The equation of the normal distribution curve fitted to the points is

$$S = 10\ \Phi\ (3.5\ \log_{10} W - 0.66) \qquad (4.1)$$

where S is the scale value, W is the speech bandwidth in KHz and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy\ . \qquad (4.2)$$

Using Fig. 4.9, the scale values of the isopreference contours in Figs. 4.7 and 4.8 can be related to equivalent unquantized speech bandwidths. The equivalence is shown in Fig. 4.12.

4.7 Further Results and Conclusions

The isopreference contours in Figs. 4.7 and 4.8 show that as L is increased along a line of constant W, a region is reached in which further increase in L does not yield any significant improvement in quality. In this region, loss of naturalness caused by low-pass filtering limits the quality of speech. Similarly, as W is increased along a line of constant L, a region is reached in which further increase in W results in little speech quality improvement. In this region, quality depends mainly on the number of quantization levels. Improvement in speech intelligibility and naturalness afforded by increased bandwidth is

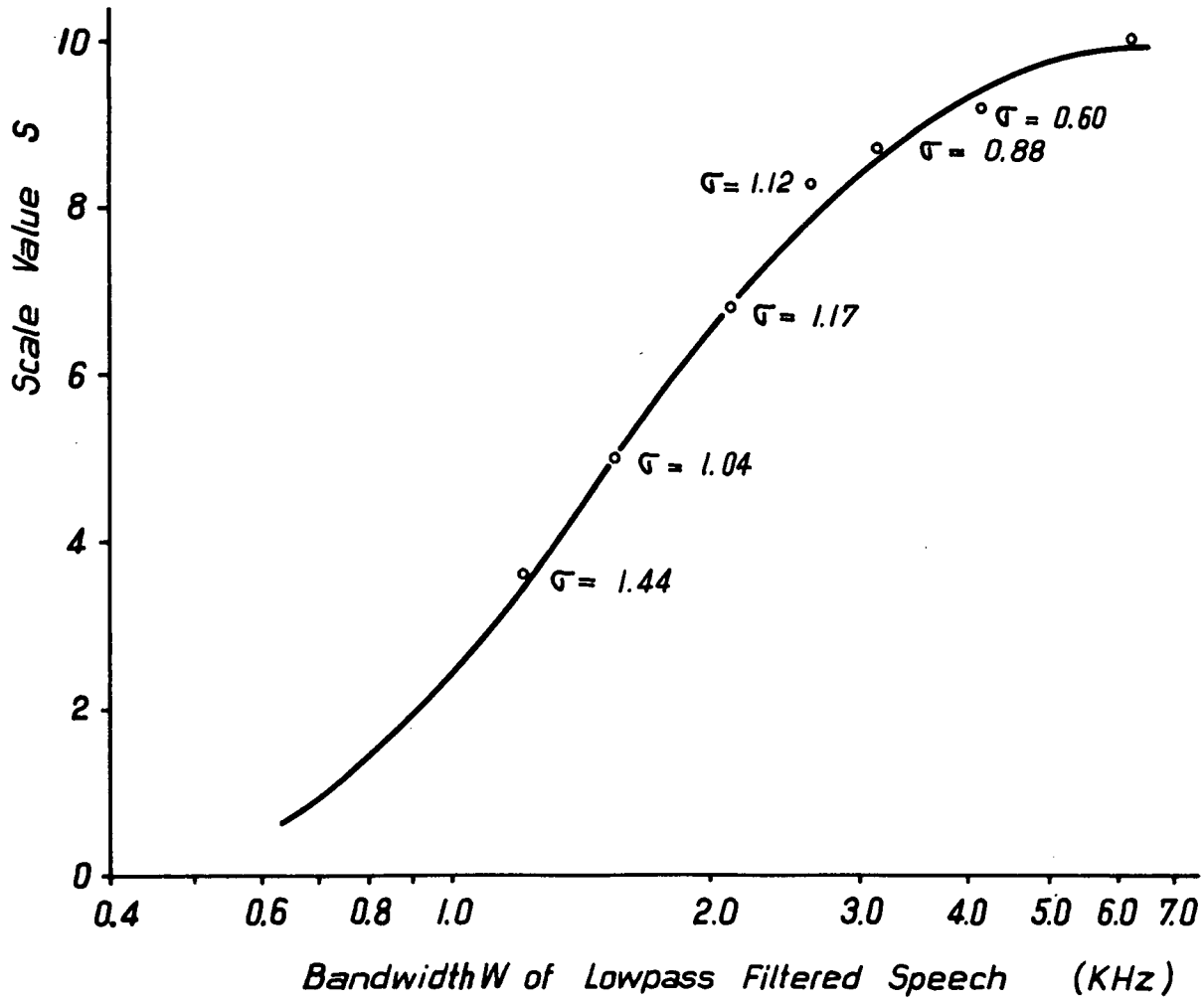


Fig. 4.9 Scale value S of unquantized, bandlimited speech vs. speech bandwidth W. The standard deviation σ of each scaled point is shown.

masked by quantization noise. Also, dependence of quality on L in this region is not as great for DPCM as for PCM.

In DPCM the difference signal e(t) occupies the entire range of the quantizer. If e(t) is reduced by more accurate prediction, the amount of quantization noise in the reconstructed speech signal is also reduced, in accordance with equation (2.9). The accuracy of prediction depends on the correlation between adjacent samples. Fig. 4.2 shows that as the sampling rate f_s is increased, the correlation between adjacent samples also increases. Since $f_s = 2.2W$, increasing the bandwidth W in DPCM not only improves intelligibility and naturalness in speech, but also reduces quantization noise. This reduction, not found in PCM as W is increased, accounts for the fact that the dependence of speech quality on L is less for DPCM than for PCM in the upper left region of the W-L plane.

Also shown in Figs. 4.7 and 4.8 are points on the isopreference contours where the bit rate R=2.2 WL is a minimum. For PCM, these points lie along the minimum bit rate locus

$$W = (0.646)2^{0.386L} \qquad (4.3)$$

where W is expressed in KHz. For DPCM, the locus of minimum bit rate is

$$W = (1.15)2^{0.245L} . \qquad (4.4)$$

Using the loci of minimum bit rate and the scaled isopreference

contours in Figs. 4.7 and 4.8, scale values as a function of L for minimum R are obtained. In Fig. 4.10, scale values versus L for minimum R are plotted. The equations of the normal distribution curves fitted to these points are

$$S_P = 10\,\Phi\,(0.45L - 2.2) \tag{4.5}$$

and

$$S_D = 10\,\Phi\,(0.53L - 2.2) \tag{4.6}$$

where S_P and S_D are the scale values for minimum R of PCM and DPCM respectively, and $\Phi\,(x)$ is defined in (4.2).

With the aid of Fig. 4.10, isopreference contours having integer scale values were obtained by interpolating and extrapolating the experimentally determined contours of Figs. 4.7 and 4.8. The integer-valued contours are shown in Fig. 4.11. The intersection points of these contours with the loci of minimum bit rate were then used to obtain the curves of Fig. 4.12, which give maximum scale values as a function of allowable bit rate R.

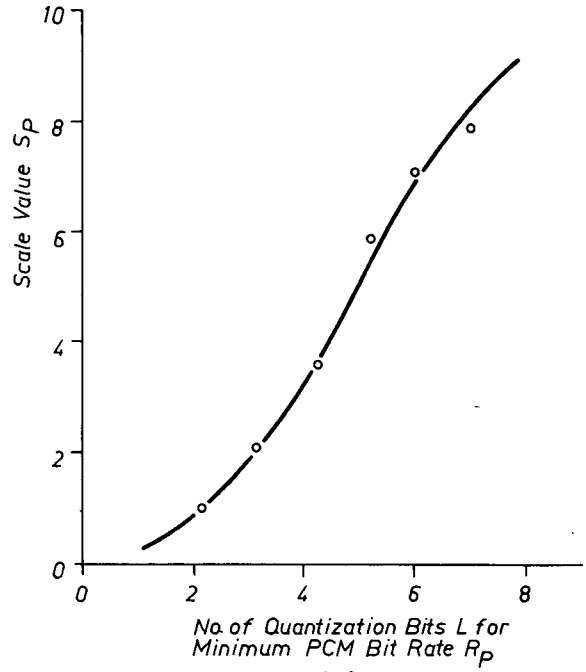
An expression for the PCM curve in Fig. 4.12 is obtained by substituting W from (4.3) into R = 2.2WL to give

$$R_P = 1.4L\ 2^{0.39L} \tag{4.7}$$

and then substituting L from (4.5) into (4.7) to give

$$R_P = 3.1\left[\Phi^{-1}(0.1S_P)+2.2\right]\ 2^{0.87\left[\Phi^{-1}(0.1S_P)+2.2\right]}. \tag{4.8}$$

Similarly, an expression for the DPCM curve in Fig. 4.12 is



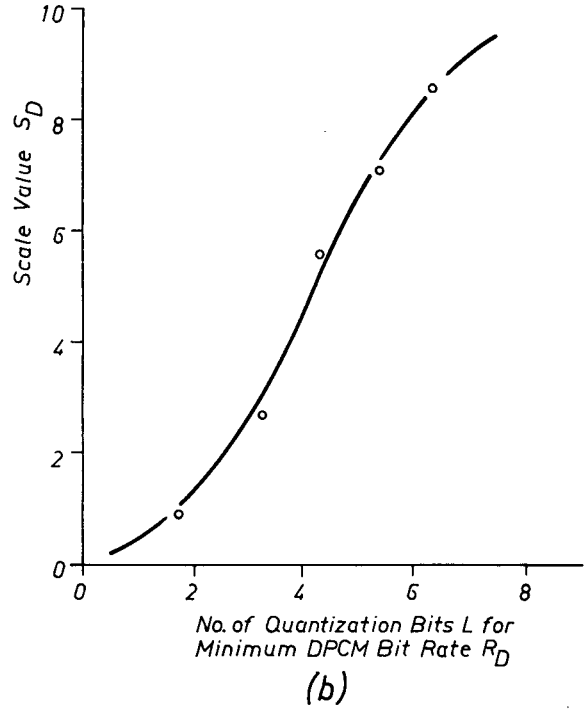


Fig. 4.10 Scale value vs. number of quantization bits for minimum bit rate.

(a) PCM. (b) DPCM.

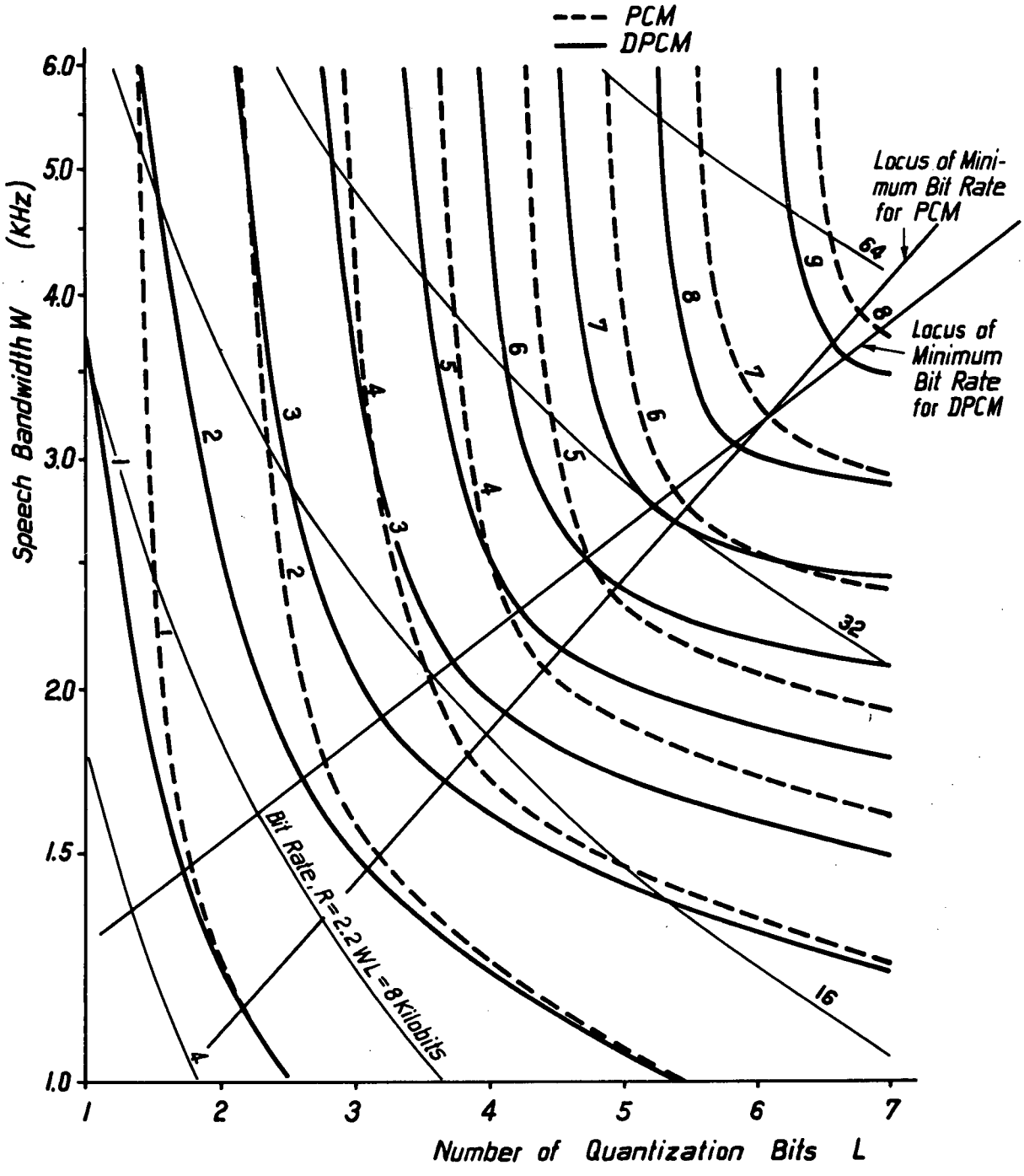


Fig. 4.11 PCM and DPCM isopreference contours having integer scale values.

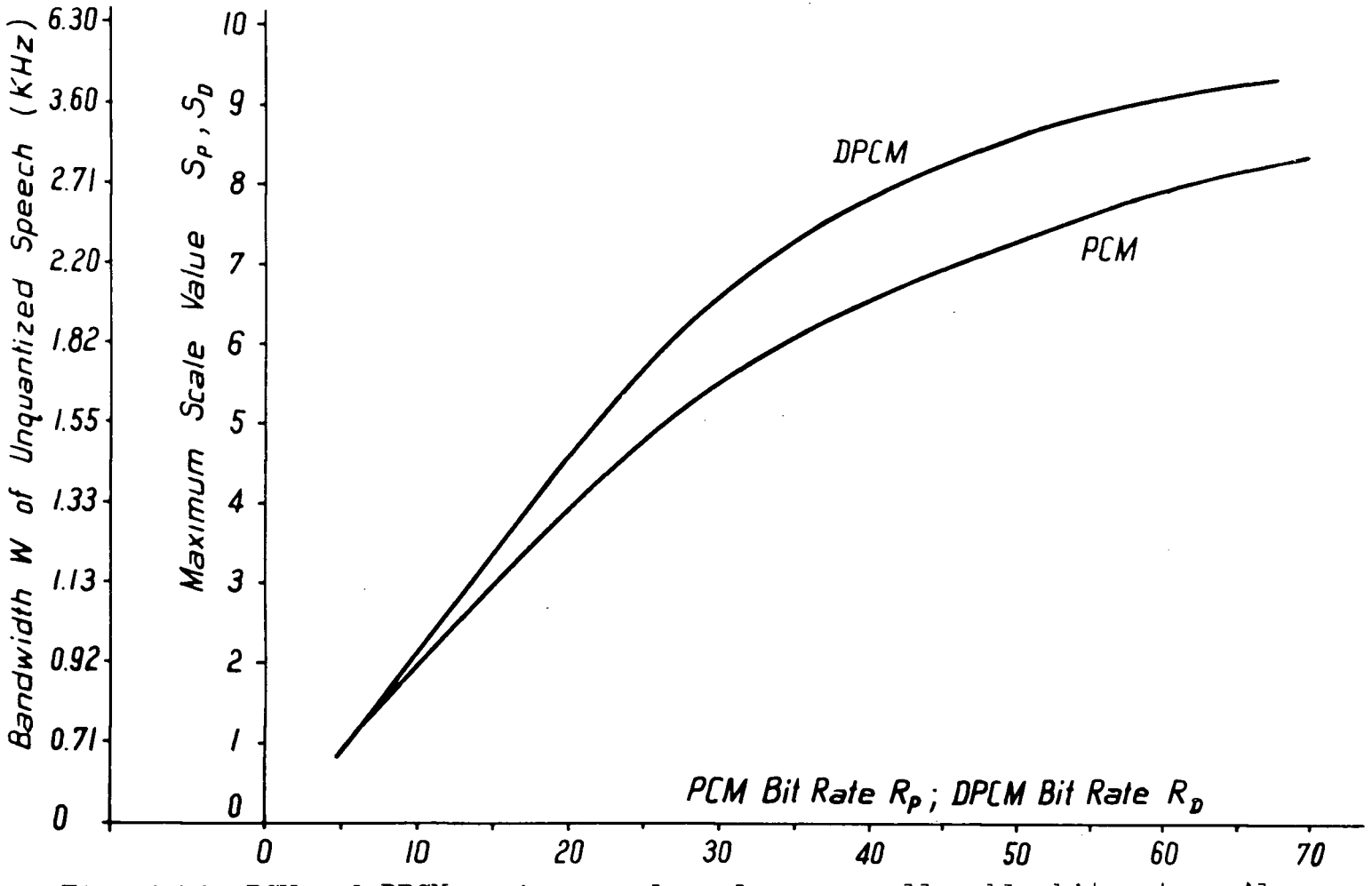


Fig. 4.12 PCM and DPCM maximum scale values vs. allowable bit rate. Also shown is the unquantized speech bandwidth equivalent of the scale values.

obtained. Using (4.4) and (4.6),

$$R_D = 2.5L\ 2^{0.25L} \tag{4.9}$$

and

$$R_D = 4.7\left[\Phi^{-1}(0.1S_D)+2.2\right]2^{0.47\left[\Phi^{-1}(0.1S_D)+2.2\right]}. \tag{4.10}$$

R_P and R_D are the minimum bit rates required to obtain the scaled values S_P and S_D for PCM and DPCM respectively.

For the same scale value $S(=S_P= S_D)$, the bit rate reduction ΔR of DCPM over PCM is obtained by subtracting (4.10) from (4.8),

$$\Delta R = 3.1\left[\Phi^{-1}(0.1S)+ 2.2\right]2^{0.87\left[\Phi^{-1}(0.1S)+2.2\right]} - 4.7\left[\Phi^{-1}(0.1S)+2.2\right]2^{0.47\left[\Phi^{-1}(0.1S)+ 2.2\right]}. \tag{4.11}$$

Bit rate reduction ΔR can also be expressed as a function of R_P, the minimum PCM bit rate required to obtain a scale value S. Hence, from (4.8) and (4.11),

$$\Delta R = R_P\left[1.0 - (1.5)2^{-0.40\lambda}\right] \tag{4.12}$$

where $(3.1\lambda)2^{0.87\lambda} = R_P$.

Equations (4.11) and (4.12) are plotted in Fig. 4.13.

Fig. 4.13 shows that DPCM is subjectively superior to PCM as an encoding system for speech redundancy reduction. For a given scale value S, the minimum DPCM bit rate required

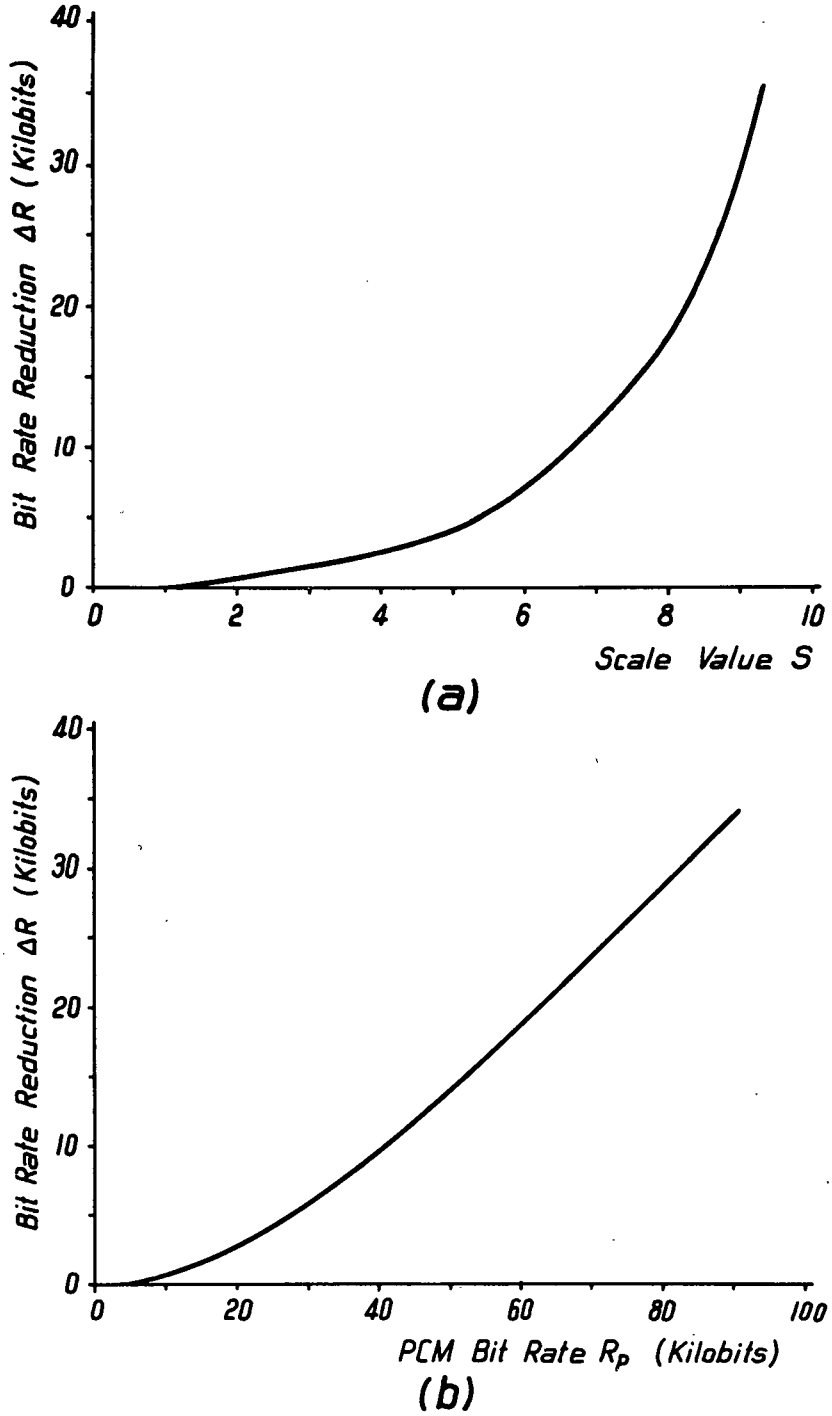
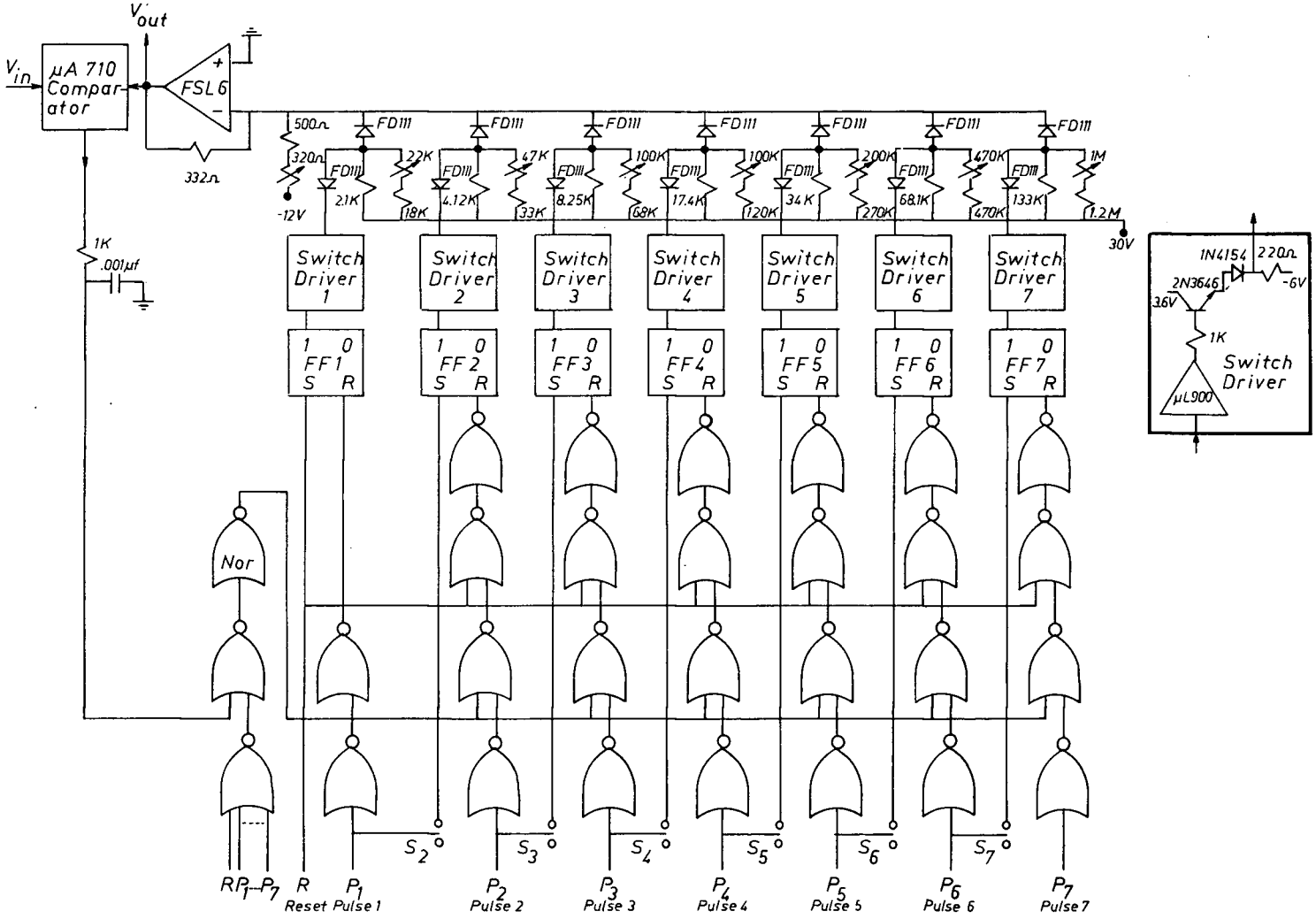


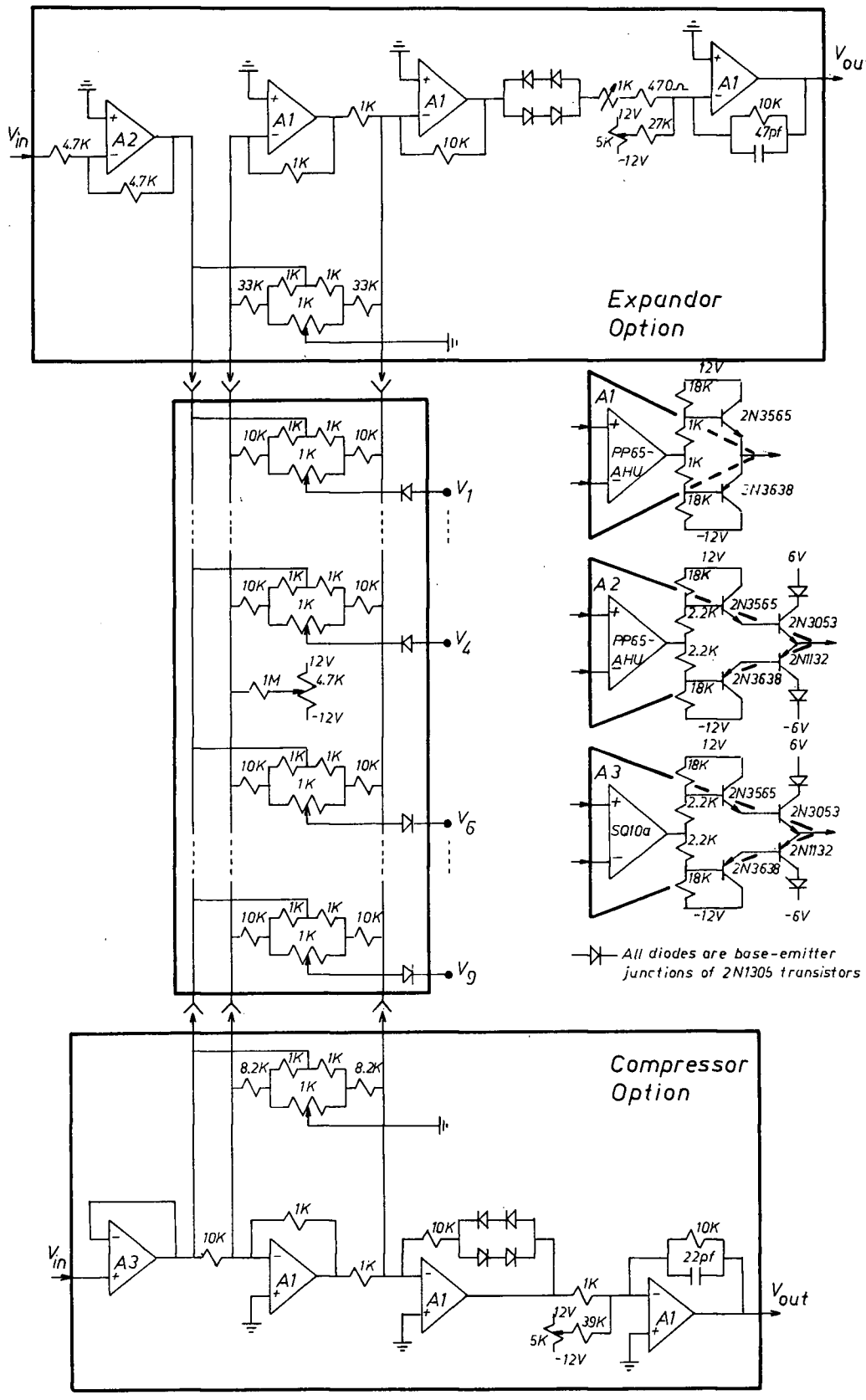
Fig. 4.13 (a) Bit rate reduction of DPCM over PCM vs. scale value.

(b) Bit rate reduction of DPCM over PCM vs. PCM bit rate.

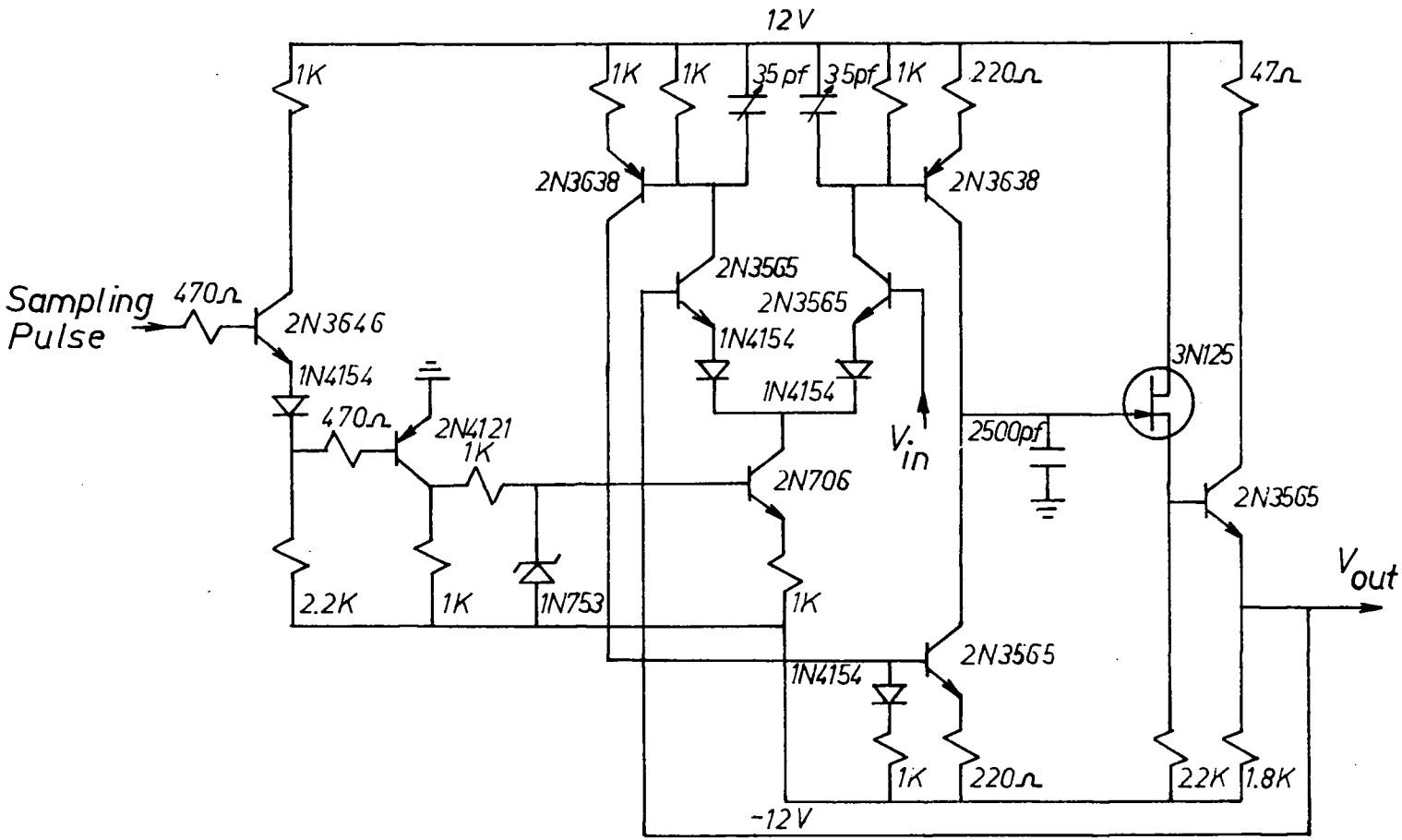
to obtain S is less than the minimum PCM bit rate required to obtain S. The difference is the bit rate reduction ΔR, and increases with S. Alternatively, for a given PCM bit rate R_p, the minimum bit rate required by DPCM to obtain the same scale value S attainable by R_p is less for DPCM by ΔR. The bit rate reduction ΔR increases with R_p.



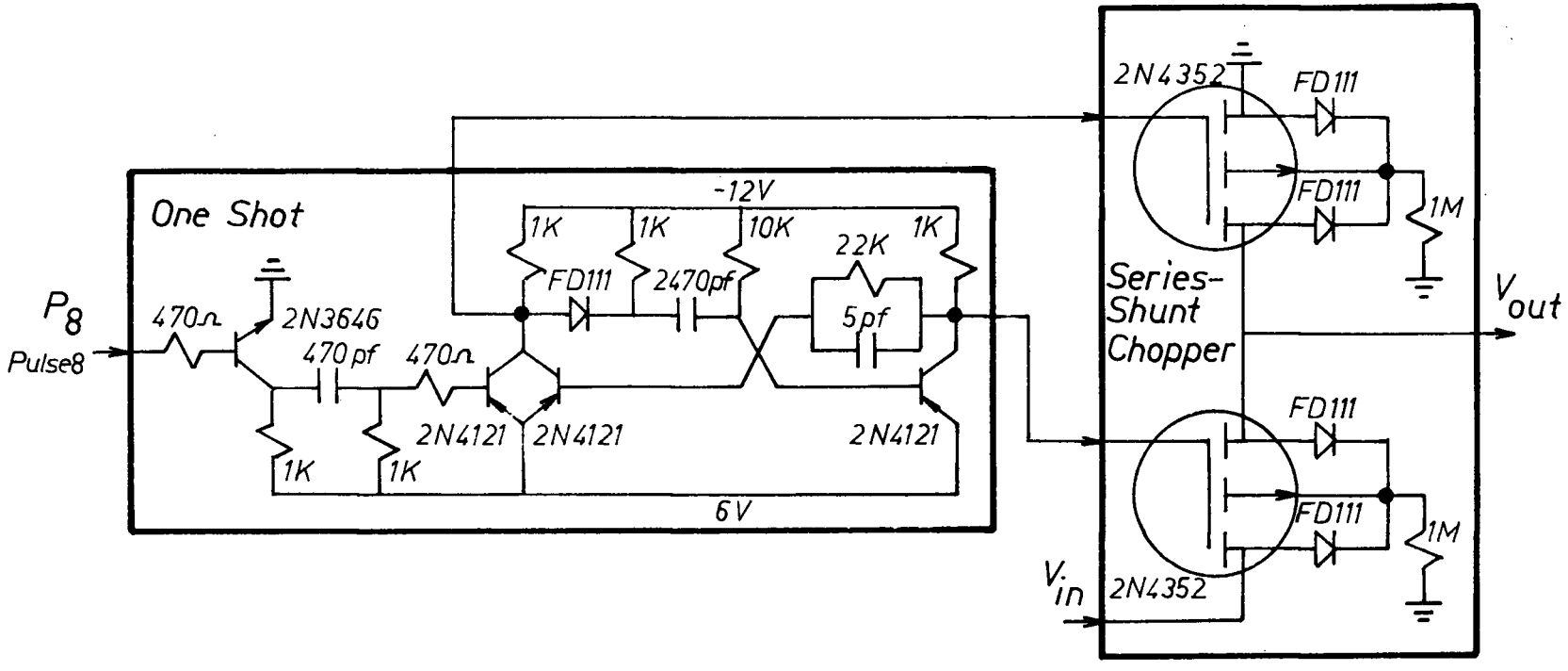
APPENDIX I Uniform Quantizer.



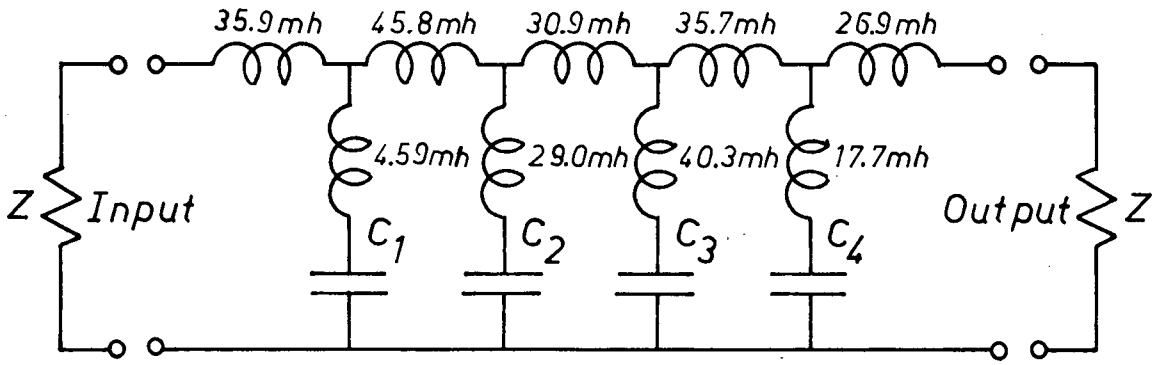
APPENDIX II Compressor and Expandor.



APPENDIX III Sample and Hold.



APPENDIX IV Sampler.



Cut off Frequency KHz	C_1 μf	C_2 μf	C_3 μf	C_4 μf	Z Ω
6	.0255	.0154	.0124	.0178	1200
4	.0576	.0348	.0280	.0402	800
3	.102	.0619	.0499	.0715	600
2.5	.147	.0890	.0718	.103	500
2	.230	.139	.112	.161	400
1.5	.408	.248	.200	.286	300
1.2	.637	.387	.312	.447	240
1	.918	.557	.449	.643	200

APPENDIX V Low-Pass Filter.

REFERENCES

1. J. Max, "Quantizing for Minimum Distortion", IRE Trans. on Information Theory, vol. IT-6, pp. 7-12, March, 1960.

2. G. Williams, "Quantizing for Minimum Error with Particular Reference to Speech", Electronics Letters, vol. 3, pp. 134-135, April, 1967.

3. J.D. Bruce, "Optimum Quantization", Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Tech. Rept. 429, March, 1965.

4. J.D. Bruce, "On the Optimum Quantization of Stationary Signals", IEEE International Convention Record (USA), vol. 12, pp. 118-124, 1964.

5. P.F. Panter and W. Dite, "Quantization Distortion in Pulse-Count Modulation with Nonuniform Spacing of Levels", Proc. IRE, vol. 39, pp. 44-48, January, 1951.

6. G.M. Roe, "Quantizing for Minimum Distortion", IEEE Trans. on Information Theory (Correspondence), vol. IT-10, pp. 384-385, October, 1964.

7. V.R. Algazi, "Useful Approximations to Optimum Quantization", IEEE Trans. on Communication Technology, vol. Com-14, pp. 297-301, June, 1966.

8. L.M. Goodman, "Optimum Sampling and Quantizing Rates", Proc. IEEE (Correspondence), vol. 54, pp. 90-92, January, 1966.

9. A.I. Liff, "Mean-Square Reconstruction Error", IEEE Trans. on Automatic Control (Correspondence), Vol. AC-10, pp. 370-371, July, 1965.

10. J. Katzenelson, "On Errors Introduced by Combined Sampling and Quantization", IRE Trans. on Automatic Control, vol. AC-7, pp. 58-68, April, 1962.

11. K. Steiglitz, "Transmission of an Analog Signal over a Fixed Bit-Rate Channel", IEEE Trans. on Information Theory, vol. IT-12, pp. 469-474, October, 1966.

12. H. Van de Weg, "Quantizing Noise of a Single Integration Delta Modulation System with an N-Digit Code", Philips Research Rept., vol. 8, pp. 367-385, 1953.

13. K. Nitadori, "Statistical Analysis of ΔPCM", Electronics and Communications in Japan, vol. 48, pp. 17-26, February, 1965.

14. J.B. O'Neal, "Predictive Quantizing Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals", Bell Sys. Tech. J., vol. 45, pp. 689-721, May-June, 1966.

15. J.B. O'Neal, "A Bound on Signal-to-Quantizing Noise Ratios for Digital Encoding Systems", Proc. IEEE, vol. 55, pp. 287-292, March 1967.

16. R.A. McDonald, "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems-Particular Applications to Voice Signals", Bell Sys. Tech. J., vol. 45, pp. 1123-1151, September, 1966.

17. L.S. Golding and P.M. Schultheiss, "Study of an Adaptive Quantizer", Proc. IEEE, vol. 55, pp. 293-297, March, 1967.

18. J.E. Abate, "Linear and Adaptive Delta Modulation", Proc. IEEE, vol. 55, pp. 298-308, March, 1967.

19. T.S. Huang, "PCM Picture Transmission", IEEE Spectrum, vol. 2, pp. 57-63, December, 1965.

20. T.S. Huang and M. Chikhaoui, "The Effect of BSC on PCM Picture Quality", IEEE Trans. on Information Theory, vol. IT-13, pp. 270-273, April, 1967.

21. R.C. Brainard, "Subjective Evaluation of PCM Noise Feedback Coder for Television", Proc. IEEE, vol. 55, pp. 346-353, March, 1967.

22. B. Smith, "Instantaneous Companding of Quantized Signals", Bell Sys. Tech. J., vol. 36, pp. 653-709, March, 1957.

23. A. Papoulis, Probability, Random Variables, and Stochastic Processes, New York: McGraw-Hill, 1965, pp. 390-394.

24. J.M. Wozencraft, and I.M. Jacobs, Principles of Communication Engineering, New York: J. Wiley and Sons, 1965, pp. 320-323.

25. C.E. Shannon and W. Weaver, The Mathematical Theory of Communication, Urbana: The University of Illinois Press, 1962, pp. 67-68.

26. P.E. Harris and B.E. Simmons, "DC Accuracy in a Fast Boxcar Circuit via a Comparator", IEEE Trans. on Electronic Computers, vol. EC-13, pp. 285-288, June, 1964.

27. H. Mann, H.M. Straube, and C.P. Villars, "A Companded Coder for an Experimental PCM Terminal", Bell Sys. Tech. J., vol. 41, pp. 173-226, January, 1962.

28. The Digital Logic Handbook (1967), Digital Equipment Corporation, Maynard, Mass., pp. 273-293.

29. W.S. Torgerson, "Theory and Methods of Scaling, New York: J. Wiley and Sons, 1958.

30. W.A. Munson and J.E. Karlin, "Isopreference Method for Evaluating Speech-Transmission Circuits", J. Acoust. Soc. Am., vol. 34, pp. 762-774, June, 1962.

31. N.R. French and J.C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds", J. Acoust. Soc. Am., vol. 19, pp. 90-119, January, 1947.

32. R.E. Totty and G.C. Clark, "Reconstruction Error in Waveform Transmission", IEEE Trans. on Information Theory (Correspondence), vol. IT-13, pp. 336-338, April, 1967.