8014963

DHADESUGOOR, VAMAN RAO

DELTA MODULATION SYSTEMS FOR VOICE COMMUNICATION

City University of New York PH.D. 1980

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106 18 Bedford Row, London WC1R 4EJ, England

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs ______
2. Colored illustrations ______
3. Photographs with dark background ✓
4. Illustrations are poor copy ______
5. Print shows through as there is text on both sides of page ______
6. Indistinct, broken or small print on several pages ______ throughout ______
7. Tightly bound copy with print lost in spine ______
8. Computer printout pages with indistinct print ______
9. Page(s) ______ lacking when material received, and not available from school or author ______
10. Page(s) ______ seem to be missing in numbering only as text follows ______
11. Poor carbon copy ______
12. Not original copy, several pages with blurred type ______
13. Appendix pages are poor copy ______
14. Original copy with light type ______
15. Curling and wrinkled pages ______
16. Other ______

DELTA MODULATION SYSTEMS FOR VOICE COMMUNICATION

by

VAMAN RAO DHADESUGOOR

A dissertation submitted to the Graduate faculty in Electrical Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1980

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

1/28/80
Date

Chairman of Examining Committee

1/28/80
Date

F E Thau
Executive Officer

Prof. D.L. Schilling (mentor)

Prof. T. Apelewicz

Dr. J. Garodnick

Prof. H. Taub

Prof. C. Ziegler

Dean F. Thau

The City University of New York

ABSTRACT

In this dissertation, the performances of the Song Voice Adaptive Delta Modulator (SVADM) and the Continuously Variable Slope Delta Modulator (CVSD) in terms of dynamic range, sampling rate and channel errors are compared. The use of SVADM and CVSD as source encoders in packet voice networks, the digital silence detection algorithm and the performance of the network are presented. The performance of the network has been evaluated subjectively for both the fixed packet size scheme as well as the variable packet size scheme. The parameters employed for subjective evaluation are packet size, silence detection algorithm, bit rate and packet loss rate. In addition, a new concept of variable rate delta modulator has been presented.

To my wife and my parents

"Genious is one percent inspiration and
ninety-nine percent perspiration"

Thomas Edison

Acknowledgements

The author wishes to express his sincere gratitude to his mentor, Prof. D.L.Schilling, for the expert counsel and continued encouragement throughout this research. The author also expresses deep appreciation and thanks to Prof. D.L.Schilling for his help outside this dissertation without which the dissertation might not have been completed.

The author wishes to thank the members of his doctoral committee: Prof. T.Apelewicz, Dr. J.Garodnick, Prof. H.Taub, and Prof. C.Ziegler for the time and effort each has spent on the reading and evaluation of this dissertation. In addition, the author expresses special thanks to Prof. C.Ziegler for providing valuable help in this research.

The author wishes to thank all of his T502F colleagues including Prof. H.Tanaka, Dr. N.Scheinberg and Prof. C.V. Chakravarthy for their useful discussions.

The author also, wishes to thank his wife, Janaki, for her constant support. Finally, the author wishes to express his appreciation to his seven months old son, Sreesha, for sleeping eleven hours every night without disturbing the author.

This dissertation has been partially supported by ARPA under grant MDA-903-78-C-0182.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1

Introduction

Digital modulation of analog signals such as speech, video, etc., is taking over from the conventional analog means of transmission because of the following main advantages:

In a short haul baseband digital transmission system, almost error free transmission can be achieved by the use of repeaters at suitable intervals.

System reliability can be enhanced at unattended places by the ease of introducing redundancy in digital circuits.

The use of error correcting schemes permit the transmission of digital signals over noisy channels like satellite links.

The use of digital computers for reliable and quick signal handling like message switching, routing etc., required the signal to be in digital form.

The encryption techniques on signals processed digitally are also fast becoming very popular and effective.

The conversion of a signal into a digital format involves two processes - Sampling the signal and representing the samples in a suitable binary code.

1.1 The Pulse Code Modulation

The first digital conversion system, which is still the existing form of digital telephone communication is Pulse Code Modulation (PCM). Here the signal is sampled at the Nyquist rate and each sample is represented by a N-bit binary number. As a result of replacing every sample by N-bits, the bandwidth increases by a factor of N over the bandwidth required to pass the original signal. The use of PCM also produces noise. Representing each sample by a N-bit binary word limits the number of levels of the input signal to only 2^N rather than the infinite number of levels possible in an analog signal. The result is the "quantization" noise. To decrease the quantization noise requires an increase in N which is accompanied by a bandwidth increase.

When bandwidth is at a premium, one looks for means of reducing the bandwidth without simultaneously increasing the noise. Research has been conducted to evolve different coding schemes such as Differential Pulse Code Modulation and Delta Modulation to reduce the bandwidth without significantly increasing the noise

1.2 Differential Pulse Code Modulation

In a Differential Pulse Code Modulation (DPCM) scheme, we form the difference between successive samples and then encode the first difference. Figure 1.2.1 shows the block schematic diagram of the DPCM. From Fig. 1.2.1, the first difference, Y(k), at the k^{th} interval is given by

$$Y(k) = M(k) - M(k-1) \tag{1.2.1}$$

where,
M(k) is the input sample at the k^{th} interval,
M(k-1) is the input sample at $(k-1)^{th}$ interval.

In Fig. 1.2.1 Y(k) is the quantized difference signal and M(k) is the approximated k^{th} input sample. The varianceof the first difference Y(k) is given by

$$\sigma_Y^2 = \left[\{X(k) - X(k-1)\}^2\right] \tag{1.2.2}$$

where.
$\{\ \}$ denotes the expectation operation.

By rewriting the Eq. (1.2.2), we get

$$\sigma_Y^2 = \{X^2(k)\} + \{X^2(k-1)\} - 2\{X(k)\ X(k-1)\} \tag{1.2.3}$$

$$= 2\sigma_x^2 - 2\rho\sigma_x^2 \tag{1.2.4}$$

$$= 2\ (1-\rho)\sigma_x^2 \tag{1.2.5}$$

where,
σ_x^2 is the variance of X(k) and
ρ is the correlation between adjacent input samples.

If ρ is larger than 0.5, then σ_Y^2 would be smaller than σ_x^2. As a result, y(t) has a smaller deviation than x(t) and we have reduced the dynamic range of the quantizer input. Thus we can reduce the number of quantization levels. This is the basic advantage of the DPCM.

However, the simple DPCM scheme of Fig.1.2.1 has one serious drawback. At the receiver, the quantization noise cumulatively increases. The technique is readily improved by forming the signal $\hat{M}(k)$ at the transmitting end as in Fig. 1.2.2. Then we use $\hat{M}(k-1)$ to form the difference, so that the transmitter is aware of the quantizing noise at each sample. This technique will ensure that the quantizing noise does not build up at the receiving end.

If the sampling rate $R_s(=1/T_s)$ is sufficiently high, the value of $\rho \lessapprox 1$ especially for speech and TV signals and the quantizer levels can be reduced to just two levels resulting in the Delta Modulator (DM). Then the decoder becomes just an accumulator for a binary stream, i.e., an integrator.

1.3 Linear Delta Modulation

The Linear Delta Modulator (LDM) is shown in Fig. 1.3.1 and is seen to be a special case of DPCM system. In the LDM shown in Fig. 1.3.1, the difference between m(t) and the integrator output x(t) is converted into one of two levels '1' or '0' depending on whether $e(t)>0$ or $e(t)\leq 0$. At each clock period T, the integrator output increases or decreases by a fixed value S called the step-size depending upon whether $e(k)$ was '1' or a '0' respectively. The signal $m(t)$ is therefore approximated by increments (or decrements) of S at each clock period. This again produces quantizing noise.

Figure 1.3.2 illustrates two types of quantizing error in DM - slope overload noise and granular noise. The slope overload is said to occur when the step size S is too small to follow the steep segment of the input waveform. Granularity on the other hand, refers to a situation where the stair case function $x(t)$ hunts around a relatively flat segment of the input waveform, with a step size that is too large relative to the local slope characteristics of the input. To reduce the granular noise, we could reduce the step size S. However, this introduces the slope overload noise, since the maximum slope that can be represented by the system is $S.T$ where T is the sampling interval. The combination of the slope overload and granular noises result in a very low dynamic range. In order to overcome this problem, we should design a delta modulator which generates a small step size when the input slope is small and large step size when the input slope is large. This kind of design leads us to the "Adaptive" Delta Modulation (ADM).

1.4 Adaptive Delta Modulation

Many different step size adaptive algorithms have been proposed and implemented to increase the dynamic range of a delta modulator [4,6,]8,21]. In the present study, we have used a ADM system which is a variation of the Song Mode Robust ADM [21] and call it the Song Voice

Adaptive Delta Modulator (SVADM). The performance of this system is compared to the most efficient commercially available system - the Continuously Variable Slope Delta Modulator (CVSD) [18]. Both the SVADM and the CVSD have been described in later chapters.

1.5 Statement of the Problem

This dissertation is concerned with Voice Communication using various ADM techniques. We explore the SVADM algorithm which was originally designed and developed in the Communications Laboratory at The City College of New York and the CVSD, which has been made commercially available by the Motorola and the Harris Corporations. We perform various tests to compare the two systems. Since there are no standard test procedures specified for delta modulators, we selected certain tests required by the institutions who supported this research. The emphasis is, however, on the subjective evaluation of the processer voice.

We also develop methods of using delta modulation in packet voice networks. We examine techniques which detect, silent periods in the voice. A reduction in the transmission rate is achieved by eliminating these silent periods from transmission.

Finally, we design the Hybrid ADM (HADM), which utilizes a different type of adaptive technique. Presently, all ADMs have adaptive step size algorithms, where the bit rate remains constant. We develop a new algorithm which has an adaptive bit rate in addition to a step size algorithm.

1.6 Summary of Results

In this section we present a brief description of the contents of this dissertation and we qualitatively summerize the results obtained.

In Chapter 2, we begin with the description of the CVSD algorithm. We discuss briefly the implementation of the algorithm and also the way the algorithm uses the syllabic characteristic of the speech signal. The CVSD has been developed exclusively to encode the speech signals.

In Chapter 3, we describe the SVADM algorithm. Also, we discuss the implementation in detail. We present different error correction techniques studied and evaluation tests performed using sinusoidal signal as the input.

In Chapter 4, we describe the subjective tests performed to compare the SVADM with the CVSD. We found that for bandlimited speech signals (300Hz - 2500Hz), the SVADM has 10 to 15 dB higher dynamic range than the CVSD at all bit rates, f_s (the bit rate is same as the sampling rate for DM). For example, at f_s = 32Kb/s, the SVADM has a dynamic range of 40dB while the CVSD has a dynamic range of 30dB and at f_s = 16 Kb/s, the SVADM has a 30dB dynamic range while the CVSD has a 20dB dynamic range. These results were established subjectively by decreasing the input signal level in steps of 10dB and varying the bit rate from 32Kb/s to 9.6Kb/s. In addition, when we introduced channel errors, we found that at an input level of 0dB and an error rate of 10^{-1}, the performance of the CVSD was preferred to that of the SVADM. However, for all other conditions of operation, the SVADM is preferred to the CVSD. The subjective comparison of the SVADM and the CVSD constitutes the first part of this dissertation.

The second part of this dissertation comprises a study of Packet Voice networks using delta modulation as the source encoding technique. Chapter 5 gives a brief introduction of the packet networks. Chapters 6 and 7 describe the studies made in Packet Voice networks using delta modulators.

The enormous increase in voice traffic coupled with the availability of computer networks has resulted in a need to establish a virtual connection between two parties rather than a physical connection. The use of a virtual connection will result in a more efficient use of bandwidth. It has been a known fact that conversational speech consists of a lot of silent periods, i.e. periods when both parties do not speak [12]. Suitable algorithms have been developed to detect these silent periods and not transmit any packets during these periods. Thus an efficient packet communication is realized. In a packet network, the customers use the network only during the active period of speech and not during the silent period and thus they only pay for the packets transmitted. This is a tremendous advantage for a packet voice network. In addition, the packet network would enable the transmission of data and voice simultaneously.

A packet is defined as a group of binary digits including data and call control signals which are switched as a composite whole. The term packet switching is defined as the transmission of data by means of addressed packets whereby a transmission channel is occupied for the duration of transmission of the packet only. The channel is then available for use by packets being transferred between different data terminal equipment (CCITT definition). As mentioned earlier, the connection between two places will be established via different paths depending upon the traffic at that instant. Therefore, it is obvious that the rate of transmission of the packet be as small as possible.

Among the methods of digital conversion of the input voice signal, we have selected DM for the study of a packet voice network because of the following reasons:

The bit rate and hence the packet rate for PCM is higher than for ADM.

PCM is highly susceptible to channel noise while ADM is highly robust in the presence of bit errors.

The other methods of bandwidth reduction such as Linear Predictive coders and Adaptive predictive coders involve complex design and are highly susceptible to channel errors.

The use of ADM in packet transmission involved the detection of the active and silent periods in speech from the digital output of the encoder, and formation of suitable groups or packets. Only the active periods are transmitted. This dissertation presents new algorithms and studies the performance of these algorithms for detection of silence, for packet formation, algorithms to be used at the receiving end when a packet is lost and the effect of packet loss on the speech quality. Both fixed and variable packet sizes have been considered. The CVSD and the SVADM algorithms are compared. The study presents the following new results:

The performance of the packet voice network using the SVADM and the CVSD is similar. However, the dynamic range difference between the SVADM and the CVSD still remains

A packet loss rate up to 10^{-2} is not noticeable.

The algorithms developed for silence detection were able to eliminate nearly all of the silence from transmission, thereby achieving an efficient packet network using delta modulators as source encoders.

The third part of this dissertation presents a new algorithm for the ADM. The SVADM and the CVSD have algorithms for step size adaptation. The new algorithm uses a multiple frequency sampling technique, in addition to the step size adaptation. Chapter 8 describes this new delta modulator which we called the "Hybrid Adaptive Delta Modulator (HADM). As an example, the bit rate during the active period of speech is high and during the silent period of the speech, the bit rate is low. This kind of adaptation is particularly useful when we do not use the packet voice network. This is due to the fact that when a silent period is not transmitted, it is difficult to reinsert the silence at the receiver. Thus, HADM has both the step size adaptation as well as the bit rate adaptation. The HADM requires suitable buffers both at the transmitting end and at the receiving end. The transmission is accomplished at an average rate. The bit rate switching information is contained in the past history and as such, there is no need to transmit additional information.

1.7 Significance of the Results

Summarizing,this dissertation consists of the comparison of the SVADM and the CVSD, the use of ADM in packet voice networks and a different look at the adaptive technique in ADM.

By using subjective tests, we have been able to show that the SVADM is preferred to the CVSD. We were, also able to establish that ADM is an efficient and a viable alternative as a source encoder in packet voice networks. Significant reduction in the bit transmission rate has been achieved by using simple silence detection schemes and not transmitting packets during these silent periods. The HADM also achieves a bit rate reduction by transmitting the bits at an average rate. On the whole, we can develop an efficient delta modulation system without adding significantly to the hardware.

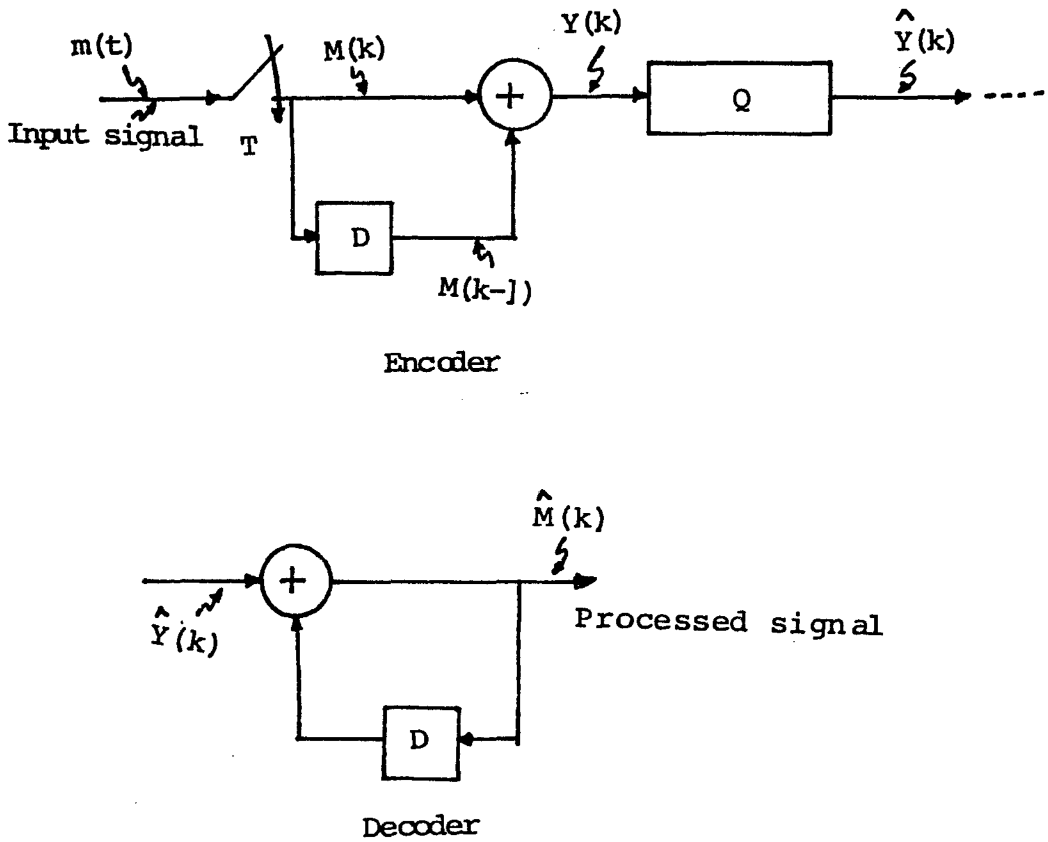
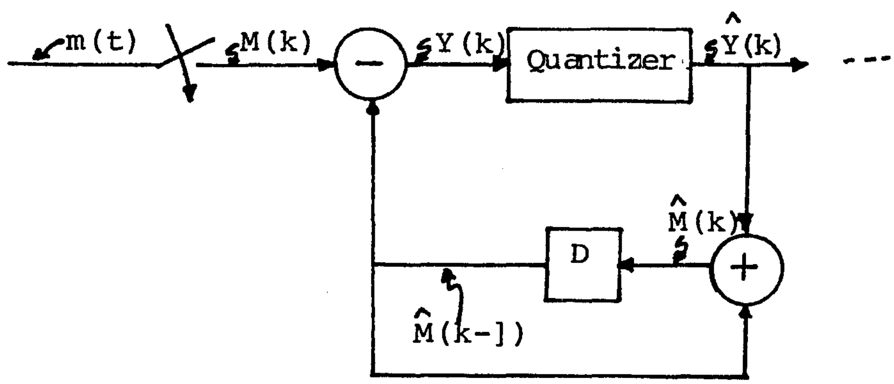
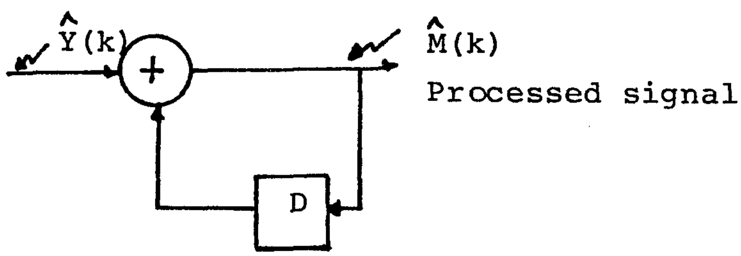


Fig. 1.2.1 Differential Pulse Code Modulation (DPCM)



Encoder



Decoder

Fig. 1.2.2 Modified DPCM

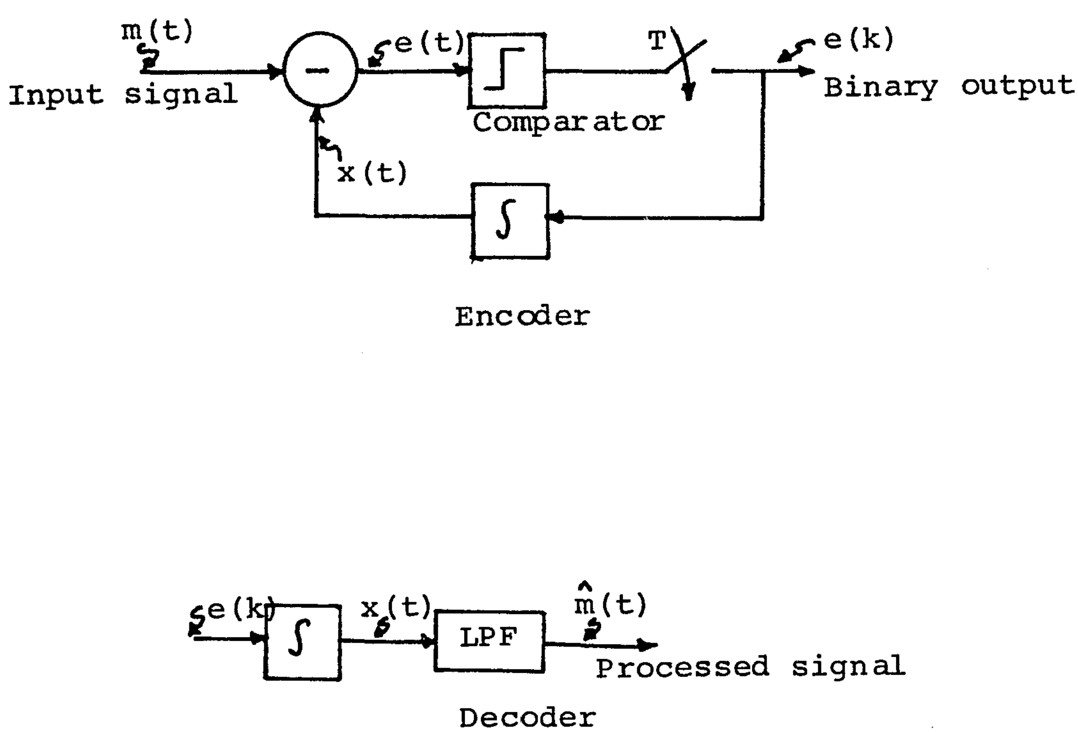


Fig.].3.] Linear Delta Modulation (LDM)

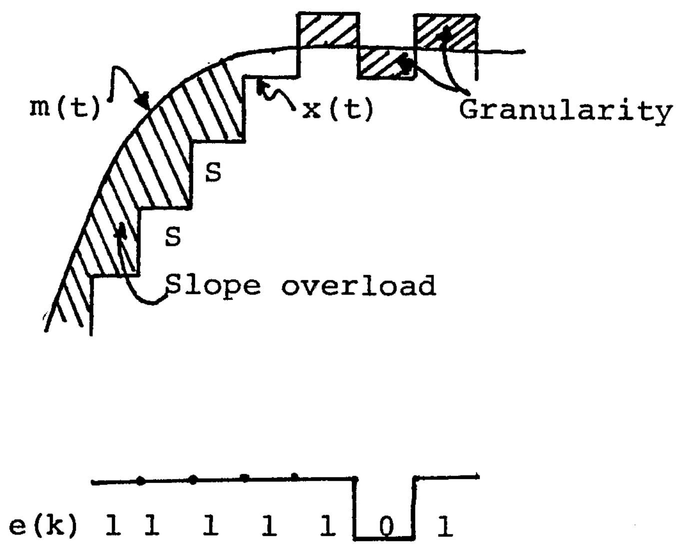


Fig. 1.3.2 Slope overload and granular noise

Chapter 2

Continuously Variable Slope Delta Modulator(CVSD)

The CVSD is an adaptive delta modulator specifically used as voice processor. The adaptive technique of CVSD exploits the syllabic characteristics of speech waveform to minimize the number of bits required in its digital description. We have been able to study the performance of CVSDs developed by both the Harris and the Motorola corporations.

Algorithm

There are several CVSD voice processors developed by different groups. However, the basic principle involving the design of the CVSD is the same. We basically limit our discussions to outline the principle of operation of the CVSD. Figure 2.1 shows the block diagram of the CVSD in general. The general algorithm is given by

$$X(k+1) = aX(k) + \left|(1-a)S(k)\right| e(k) \qquad (2.1)$$

Where,

$$e(k) = \mathrm{Sgn}\left[M(k) - X(k)\right] \qquad (2.2)$$

and

$X(k)$ is the estimate of the incoming analog signal

a is the leakage factor associated with the estimate integrator,

$S(k)$ is the k^{th} step size, and

$M(k)$ is the k^{th} input sample.

Furthermore, $S(k+1)$ is generated by syllabic companding and is given by

$$S(k+1) = bS(k) + (1-b)(V+V_1) \tag{2.3}$$

where,
V is a constant voltage when three consecutive outputs from the CVSD encoder are identical (sometimes this number could be two or four [6,18] ,
V_1 is just a constant voltage added to V to ensure that the minimum step size is non-zero, and
b is the leakage factor associated with the step integrator.

In Fig. 2.1 , the output of the overload detector is either 0 or V volts depending on the three consecutive digital outputs of the encoder. The feedback circuit of the encoder is the decoder.

In particular, the CVSD described in [7] has a time constant for the step size integrator, τ_1= 5.69 msec. and the time constant for the estimate integrator, τ_2 = 1 msec., which gives

$$b = \exp\ (-1/f_s)/(5.69\text{x}10^{-3}) \tag{2.4}$$
$$a = \exp\ (-1/f_s)/10^{-3} \tag{2.5}$$

When f_s = 16 Kb/s, b = 0.99 a = 0.94

It is of interest to note that the coefficients a, b have been adjusted differently in different CVSD processors.

Since the CVSD processor is designed for voice signals, the time constants τ_1 and τ_2 are chosen with reference to the actual wave forming the voice signal. A typical voice signal (speech) has most of its energy from 700 Hz to 1000 Hz and has an envelope of 60 to 100 Hz. The step size integrator of the CVSD generates the envelope of the speech signal and therefore the time constant τ_1 is adjusted to 5.69 msec. which corresponds to approximately 100 Hz and τ_2 is adjusted to 1 msec. which corresponds to 1000 Hz. Figure 2.2 describes the simple circuit implementation of CVSD encoder.

For comparison with the SVADM, we have used the CVSD developed by both the Harris and the Motorola corporations. However, we found that the Motorola CVSD offers a better performance compared to the Harris CVSD. Thus, we emphasized the use of the Motorola CVSD for comparison with the SVADM.

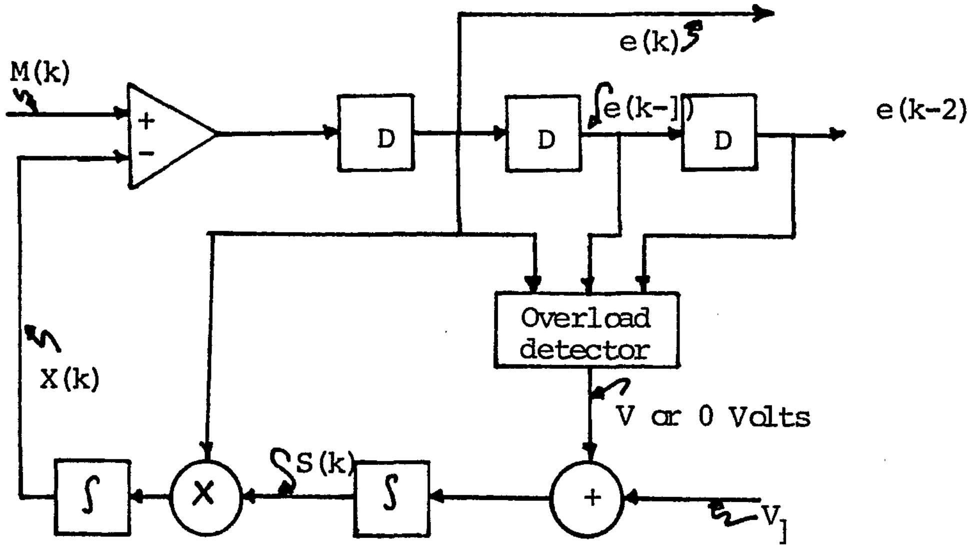


Fig. 2.1 Block diagram of the CVSD

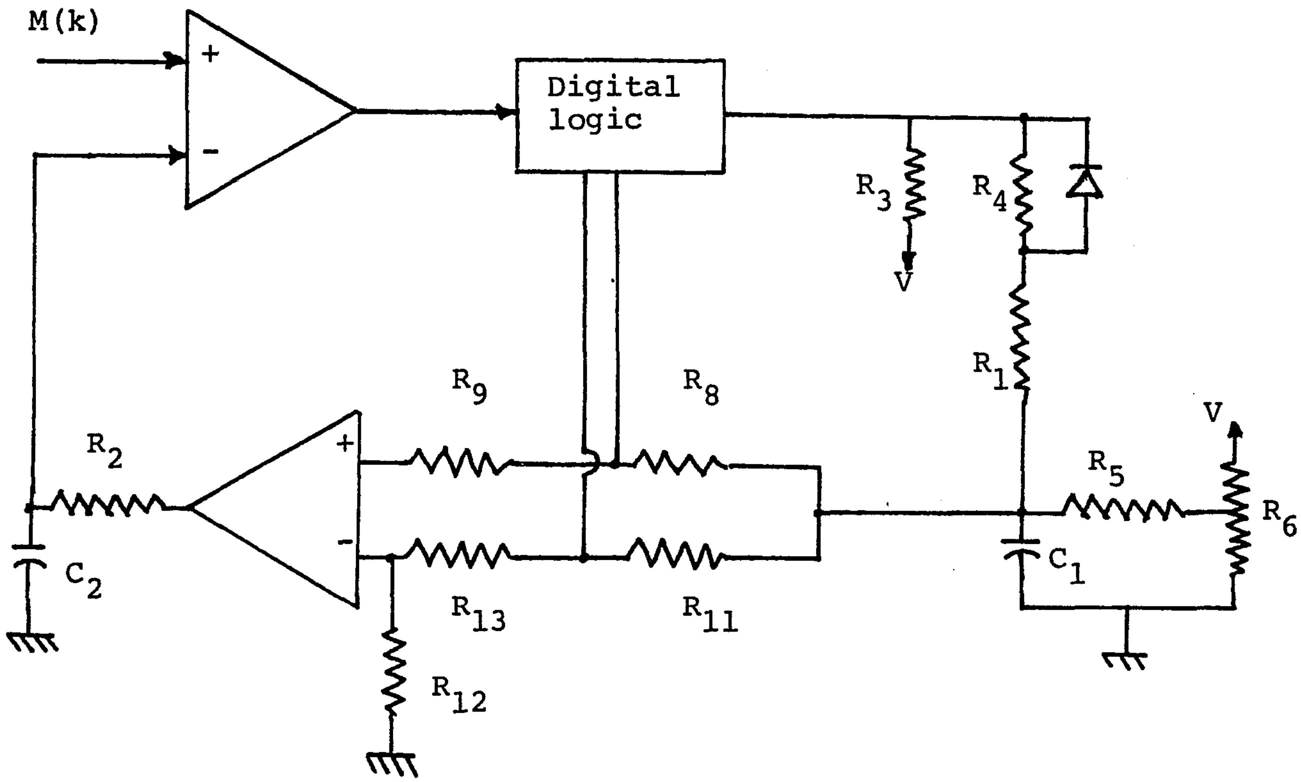


Fig. 2.2 Circuit implementation of the CVSD

Chapter 3

The Song Voice Adaptive Delta Modulator

The Song Voice Adaptive Delta Modulator (SVADM) encoder-decoder [21] has a word intelligibility of 99% at 16 kb/s bit rate and a word intelligibility of more than 90% at 9.6 kb/s bit rate. It also offers a 40 db dynamic range. In the implementation of the SVADM, ±5 V was chosen to be maximum signal level. Thus, a signal having an amplitude ranging from ±50mV to ±5V could be encoded with approximately the same Signal to Noise Ratio (SNR). The SVADM algorithm is easy to implement digitally.

3.1 Algorithm

The Algorithm of the SVADM is

$$X(k+1) = X(k) + S(k+1) \tag{3.1.1}$$

Where x(k) is the estimate of the incoming analog signal at the sample time k/f_s where f_s is the sampling rate and S(k+1), the new step size at time $(k+1)/f_s$ is given by

$$S(k+1) = |S(k)|\, e(k) + S_o\, e(k-1) \tag{3.1.2}$$

Where e(k)is the sign of the error which occurs at k/f_s and S_o is the voltage associated with the minimum step size. In the SVADM, 10-bit arithmetic was employed and therefore S_o 10 mV. If M(k) is the signal value at time k/f_s, then

$$e(k) = \mathrm{Sgn}\left[M(k) - X(k)\right] \tag{3.1.3}$$

The magnitude of the new step size S(k+l) will exceed that of the current step size S(k) by S_o so long as e(k) is of the same sign as e(k-l). However, if e(k) and e(k-l) differ in sign, the magnitude of S(k+l) will be less than that of S(k) by S_o.

Figure 3.1.1 shows the block diagram of the SVADM. It consists of a comparator which determines the sign of the difference between the input analog signal M(k) and the estimate X(k). The step generator generates S(k+l) using S(k), e(k), and e(k-l) as inputs. The new estimate X(k+l) is then realized by adding the new step size S(k+l) to the current estimate X(k). e(k) is the output of the encoder. The feedback circuit of the encoder is essentially the decoder. The factor βS_o added to generate X(k+l) is required to correct the estimate when channel errors are present.

Figure 3.1.2 shows the estimate generation using the SVADM algorithm and the bit pattern produced for a step input. Here, we see that so long as e(k) is of the same sign as e(k-l), the magnitude of the new step size S(k+l) will exceed the magnitude of the current step size S(k) by S_o, the minimum step size. On the otherhand, if e(k) and e(k-l) differ in sign, the magnitude of S(k+l) will be less than that of S(k) by S_o. Also, we notice that when the input is constant, the SVADM reaches a steady state condition. The response of the SVADM is an estimate signal which exhibits a periodic e(k) pattern of 11001100··· is generated. Thus for a bit rate of f_s, the estimate oscillates with a fundamental frequency of $f_s/4$. The amplitude of the oscillation depends on the step size at the time of oscillation . For a perfact D.C. input (constant), these oscillations might start with a large step size. It is generally found that the amplitude of a speech signal rises very fast, but decreases with a time constant of about 10 ms. This particular characteristic has been used in the design of the CVSD which was described in Chapter II. Therefore, when speech signals are encoded by the SVADM, the steady state oscillations have lower amplitudes. However, these oscillations may fall within the speech handwidth, when the SVADM is operated at lower bit rates. For example, when the speech signals are bandlimited from 300 Hz to 2500 Hz and if the bit rate, f_s=10 Kb/s. the steady state oscillations occur at 2.5 KHz, which fall into the speech band. These oscillations are heard at the decoder output. To eliminate these oscillations when the SVADM is operated at lower bit rates, we use a digital low pass filter at the output of the SVADM decoder.

3.2 The Digital Low Pass Filter

The Digital Low Pass Filter (DLPF) that was employed to eliminate the steady state oscillations is a four term non-recursive filter. It is shown in Fig. 3.2.1(a). The output of the DLPF is A(k), which is realized by

$$A(k) = (1/4)\left[X(k) + X(k-1) + X(k-2) + X(k-3)\right] \quad (3.2.1)$$

If the steady state output of the SVADM shown in Fig. 3.1.2 is the input to the DLPF, the output A(K) is

$$A(k) = M(k) \quad (3.2.2)$$

for all k, as long as X(k) has reached steady state condition.

Thus after a four term averaging, the output A(k) is a constant D.C. level in the steady state and thus we eliminated the oscillations.

To illustrate the function of the DLPF, we shall plot its frequency response. Assuming zero initial conditions and taking Z transform of equation (3.2.1) we obtain

$$H(z) = A(z)/X(z) = (1/4)(1 + z^{-1} + z^{-2} + z^{-3}) \qquad (3.2.3)$$

To know the frequency characteristics of this filter, we let $z = \exp(jwT_s)$ where $T_s = 1/f_s$. Then, from Appendix A

$$H(w) = \cos(wT_s/2)\cos(wT_s) \qquad (3.2.4)$$

Figure 3.2.2 shows the frequency response of the DLPF. The zeros of the transfer function occur at integer multiples of $f_s/4$, when the integer is divisible by four. It is exactly the first zero that eliminates the oscillations heard at the output of the SVADM decoder.

To actually realize this filter, we have modified the Eq. (3.2.1) by

$$A(k) = (1/2)\Big[(1/2)\{X(k) + X(k-1)\} + (1/2)\{X(k-2) + X(k-3)\}\Big] \qquad (3.2.5)$$

$$= (1/2)\{Z(k) + Z(k-2)\} \qquad (3.2.6)$$

where

$$Z(k) = (1/2)\{X(k) + X(k-1)\} \qquad (3.2.7)$$

The realization of Eqs. (3.2.5), (3.2.6) and (3.2.7) is shown in Fig. 3.2.1(b). When we compare Figs. 3.2.1(a) and (b), it is clear that we saved the hardware in (b) which uses

The use of DLPF to eliminate steady state oscillations is essential only when the SVADM is operated at twice the Nyquist rate. Earlier, we explained that these oscillations are inband when f_s<10 Kb/s for bandlimited speech of 300Hz - 2500Hz. However, when f_s>10 Kb/s the oscillations occur at a frequency 2.5 KHz. These oscillations are eliminated, since the processed speech is also bandlimited from 300Hz to 2500Hz.

Although, the DLPF eliminates the steady state oscillations, because of its low pass characteristics, it also attenuates certain baseband frequencies. Therefore, we have used a simple pre-emphasis filter at the output of the DLPF to boost the attenuated baseband frequencies.

3.3 Pre-emphasis Filter

The Pre-emphasis filter was used only when the DLPF was used. Figure 3.3.1 shows the implementation of the pre-emphasis filter and its frequency response. The frequency response is only plotted up to 4 KHz. The typical high pass characteristics has been achieved by varying the capacitance. It should be noted however, that the frequency response has to be adjusted depending on the input speech bandwidth and the bit rate. Since we need this filter only at bit rates ≤ 10kb/s, for designing the filter, we assumed f_s =10 kb/s. Also, the bandwidth of the input speech is 300 Hz - 2500 Hz. Since, it has been established that most of the speech energy occurs at frequencies below 1000Hz, we boosted the high frequency components from linearly from 600 Hz upto 2 KHz to nullify the effect of the DLPF on these frequencies. Subjective evaluation shows a significant improvement in the performance of the SVADM system operating at f_s = 10 Kb/s when using the DLPF and the pre-emphasis filter. It is emphasized again that both the DLPF and the Pre-emphasis filter are not required at f_s >10Kb/s.

3.4 Overflow

In the presence of channel noise, the decoder will be in a different state compared to that of the encoder. So, we modified the SVADM decoder by introducing an overflow detector to prevent the estimate to overflow and a channel correction logic to correct the state of the decoder.

Overflow Detection Logic (ODL) is needed in the decoder in conjunction with the estimate X(k). In the presence of channel noise, when X(k) and S(k+1) are of the same sign, X(k+1) may overflow. This situation might occur when the magnitudes of X(k) and S(k+1)are large. Figure 3.4.1 shows the overflow condition when an error occurs. Overflow-1 occurs because of an error in the received e(k). The difference between the encoder X(k) and the decoder X(k) is very large at the time of overflow. This large difference is not acceptable during speech. To prevent such a large difference in the estimate, ODL was implemented. The logic is very simple to implement digitally. In the SVADM, the new estimate is realized by

$$X(k+1) = X(k) + S(k+1) \tag{3.4.1}$$

Overflow occurs when

$$\left[\overline{\mathrm{Sgn}\{X(k)\} \oplus \mathrm{Sgn}\{S(k+1)\}}\right] \cdot \left[\mathrm{Sgn}\{X(k+1)\} \oplus \mathrm{Sgn}\{X(k)\}\right] = 1 \tag{3.4.2}$$

where

$\oplus$ denotes the exclusive OR function.

There are two types of Overflow, positive and negative (case A and case B respectively).

Case A: If $X(k)$ and $S(k+1)$ are positive and $X(k+1)$ is negative then positive overflow occurs. The logic detects this overflow and sets $X(k+1)$ to the most positive value (maximum).
Case B: If $X(k)$ and $S(k+1)$ are negative and $X(k+1)$ is positive, then negative overflow occurs. The logic detects this overflow and sets $X(k+1)$ to the most negative value (minimum).

3.5 Error Correction Logic:

In the presence of channel errors, the state of the decoder is different from that of the encoder. To allow the decoder to attain the state of the encoder, the error correction logic is implemented.

The delta modulator encoder output usually is transmitted using some form of channel encoding procedure such as PSK, FSK, DPSK etc. The state of the delta modulator decoder is affected if there is interference on the received signal, since it will cause an occasional error in the data bit stream by inverting a bit. We can consider this interference error as a random error.

The random error causes an inversion of the data bit, causing the state of the decoder to be different from that of the encoder. Both $X(k)$ and $S(k)$ of the decoder will usually be different from $X(k)$ and $S(k)$ of the encoder. In order to correct this error we install a "leaky integrator", so that the state of the decoder is corrected in a few sampling instants. In order to study the performance using a leaky integrator we rewrite the equations for describing SVADM encoder - decoder.

Encoder:

$$X(k+1) = X(k) + S(k+1) \quad (3.5.1)$$

$$S(k+1) = |S(k)|e(k) + S_o\, e(k-1) \quad (3.5.2)$$

$$e(k) = \mathrm{Sgn}[M(k) - X(k)] \quad (3.5.3)$$

Decoder:

$$X'(k+1) = X'(k) + S'(k+1) \tag{3.5.4}$$

$$S'(k+1) = |S'(k)|\, e'(k) + S_o\, e'(k-1) \tag{3.5.6}$$

The decoder equations use the symbols $e'(k)$, $X'(k)$ and $S'(k)$ to represent the quantities perturbed due to channel errors. We define the noise voltage, i.e. the difference between the transmitted and the receiver estimates as

$$N(k+1) = X'(k+1) - X(k+1) \tag{3.5.7}$$

$$= X'(k) - X(k) + S'(k+1) - S(k+1) \tag{3.5.8}$$

therefore

$$N(k+1) = N(k) + S'(k+1) - S(k+1) \tag{3.5.9}$$

Equation (3.5.9) shows that the noise voltage accumulates as the errors come through the system. If we can use a leaky integrator with a leak factor $0 \leq L < 1$, we can rewrite the equations (3.5.1), (3.5.4) and (3.5.4)

$$X(k+1) = L \cdot X(k) + S(k+1) \tag{3.5.10}$$

$$X'(k+1) = L \cdot X'(k) + S'(k+1) \tag{3.5.11}$$

$$N(k+1) = L \cdot N(k) + S'(k+1) - S(k+1) \tag{3.5.12}$$

The value of L usually depends on the sampling rate. It has been found experimentally that for input speech of 300Hz - 2500 Hz band, the leak factor L should be as shown in Table 3.5.1.

In Table 3.5.1 the values of L are conveniently described for digital implementation. For example, 1/128 corresponds to a shift of the estimate to the right by 7 bits. (1-1/128) X(k) can then be implementing by simply substracting the shifted estimate from the original estimate. The values of L listed in Table 3.3.1 at different sampling rates correspond to the minimum number of shifts required without degrading the signal to noise to ratio of the estimated speech and to enable the error correction at error rates of 10^{-4}, 10^{-3}, and 10^{-1}. The arithmetic varies as a function of the number of shifts used. However, for practical implementation of the delta modulator, a single leak factor of L=(1-1/64) has been chosen for all sampling rates. This leak factor needs an additional 6 bits of arithmetic to generate the shifted estimate.

In order to avoid additional arithmetic needed to use the true leaky integrator described above, we have implemented a non-linear leaky integrator which requires a minimum of additional logic. There are two types of non-linear leaky integrators studied. The types of leaking are similar, but the leak factors are different.

Non-Linear Leaky Integrator 1:

The new estimate X(k+1) is given by

$$X(k+1) = X(k) + S(k+1) + \beta S_o \qquad (3.5.13)$$

Let us represent X(k) and S(k+1) by N-bit words so that

$$X(k) = x_0 x_1 x_2 x_3 \ldots x_{N-1} \qquad (3.5.14)$$

and

$$S(k+1) = s_0 . s_1 s_2 s_3 \ldots s_{N-1} \qquad (3.5.15)$$

Where x_0 and s_0 are the sign bits, x_1 and s_1 the most significant bits and x_{N-1} and s_{N-1} are the least significant bits of X(k) and S(k+1) respectively. Then,

$$\beta = \begin{cases} +1 & \text{if } x_0 = s_0 = 1 \text{ and } x_{N-1} \oplus s_{N-1} = 0 \\ -1 & \text{if } x_0 = s_0 = 0 \text{ and } x_{N-1} \oplus s_{N-1} = 1 \\ 0 & \text{otherwise} \end{cases} \qquad (3.5.16)$$

This leak is performed on the average one out of every eight times and degrades the performance of the system only at input levels of -30 dB and below.

Non-Linear Leaky Integrator 2:

In order to improve the performance even at input level of - 30 dB and below a different non-linear integrator was developed. The new leaky integrator will leak only at larger estimates. For this case, the new estimate X(k+1) is still given by

$$X(k+1) = X(k) + S(k+1) + \beta S_o \qquad (3.5.17)$$

Where for X(k) negative

$$\beta = +1 \text{ when } x_0=s_0=1, \bar{x}_1+\bar{x}_2+\bar{x}_3=1 \text{ and } x_{N-1} \oplus s_{N-1}=0 \qquad (3.5.18)$$

for X(k) positive

$$\beta = -1 \text{ when } x_0=s_0=0, x_1+x_2+x_3=1 \text{ and } x_{N-1} \oplus s_{N-1}=1 \qquad (3.5.19)$$

and

$$\beta = 0 \text{ otherwise} \qquad (3.5.20)$$

This type of leak causes the leak to occur only during larger estimates while smaller estimates are maintained the same. Experiments have shown that this system produces better SNR at the input levels through -40dB and thus has a larger dynamic range over the non-linear leaky integrator 1. For comparing the SVADM with the CVSD, we have used the non-linear Integrator 2.

3.6 Performance of SVADM Encoder-Decoder:

In order to study the performance of the SVADM, we have used the test set up shown in Fig. 3.6.1. All three leaky integrator algorithms have been tested and measurements were made. In general, there are no rigid specifications set to define the performance of delta modulators. The users of delta modulators give their own specifications, which are closely related to the PCM specifications laid out by the C.C.I.R. The measurements made on the SVADM were required by different users. Specifically, the parameters to be estimated using a standard test signal are:

(a) Bandwidth
(b) Idle channel noise
(c) Dynamic range
(d) SNR vs Fs
(e) Linearity

The following tests were conducted:

(1) Bandwidth Test:

For this test, we set the band pass filters from 300 Hz to 3400 Hz and the bit rate, fs = 37.5 Kb/s. The input frequency was varied from 300 Hz to 3400 Hz and the output level was measured by the RMS meter. Figure 3.6.2 shows the output level as a function of the input frequency. For this test, input levels of

Odb and -10dB were chosen. We infer from Fig.3.6.2 that the output level varied within $\pm$1dB over an input frequency range of 600 Hz - 2400 Hz and within $\pm$ 3dB over a frequency range of 300 Hz - 3400 Hz. This result shows that the SVADM offers a good bandwidth for speech inputs.

(2) Idle Channel Noise Test:

In this test, instead of the band pass filter at the output of the decoder, we used a C message weighted filter, whose frequency characteristics is shown in Fig. 3.6.3. To measure the idle channel noise, the input signal to the encoder was removed and the encoder was terminated. The noise level was measured at the output using the RMS meter. We found that the idle channel noise is 65 dB below the maximum input signal when no channel errors are introduced. Also, the idle channel noise is 55dB below the maximum input signal when channel errors are introduced at the rate of 10^{-3}.

(3) Dynamic Range Test:

There are several definitions for dynamic range. For our measurements, we use a bit rate of 37.5 Kb/s and bandpass filters are set from 300 Hz to 3400 Hz. This bit rate was desired by the Army. We have varied however, the bit rate when we performed the subjective test. We define the dynamic range as the input range for which the output signal to noise ratio (SNR) is greater than or equal to 25dB. For this test, a 1 KHz signal was used. All three leaky integrators were considered. The input signal was varied from 0dB (maximum input level) down to -40dB in steps of 10dB and the SNR at the output of the decoder was measured. The SNR as a function of the input level was plotted in Fig 3.6.4 for SVADMs using three leaky integrators. We notice that the SVADM using the True Leaky Integrator (TLS) offers higher SNR at lower input levels compared to the other two. However, the SVADM using the Non-linear Leaky Integrator 2 (NLI 2) has a higher dynamic range over the SVADM using the Non-linear Leaky Integrator 1(NLI 1). We also infer from the definition of the dynamic range and the Fig.3.6.4.that the dynamic range of the SVADM is 30dB.

Even though, we have measured the dynamic range by using a sinusoid, the real importance of this parameter comes in assessing the subjective quality of the processed speech, which will be presented in the next chapter. For speech signals, the ratio of maximum to minimum signal amplitude is about 30dB. Therefore, the SVADM whose dynamic range is 30dB is a good encoder of speech signals.

(4) SNR vs bit rate:

In this test, a 1 KHz tone was used as the input to SVADM. The band pass filters were set from 300 Hz to 2500 Hz The bit rate was varied and each time, the SNR was measured using a distortion analyzer. Fig.3.6.5 shows the graph of SNR vs fs of the SVADM (NLI2), which will eventually be used for subjective comparison with the CVSD.

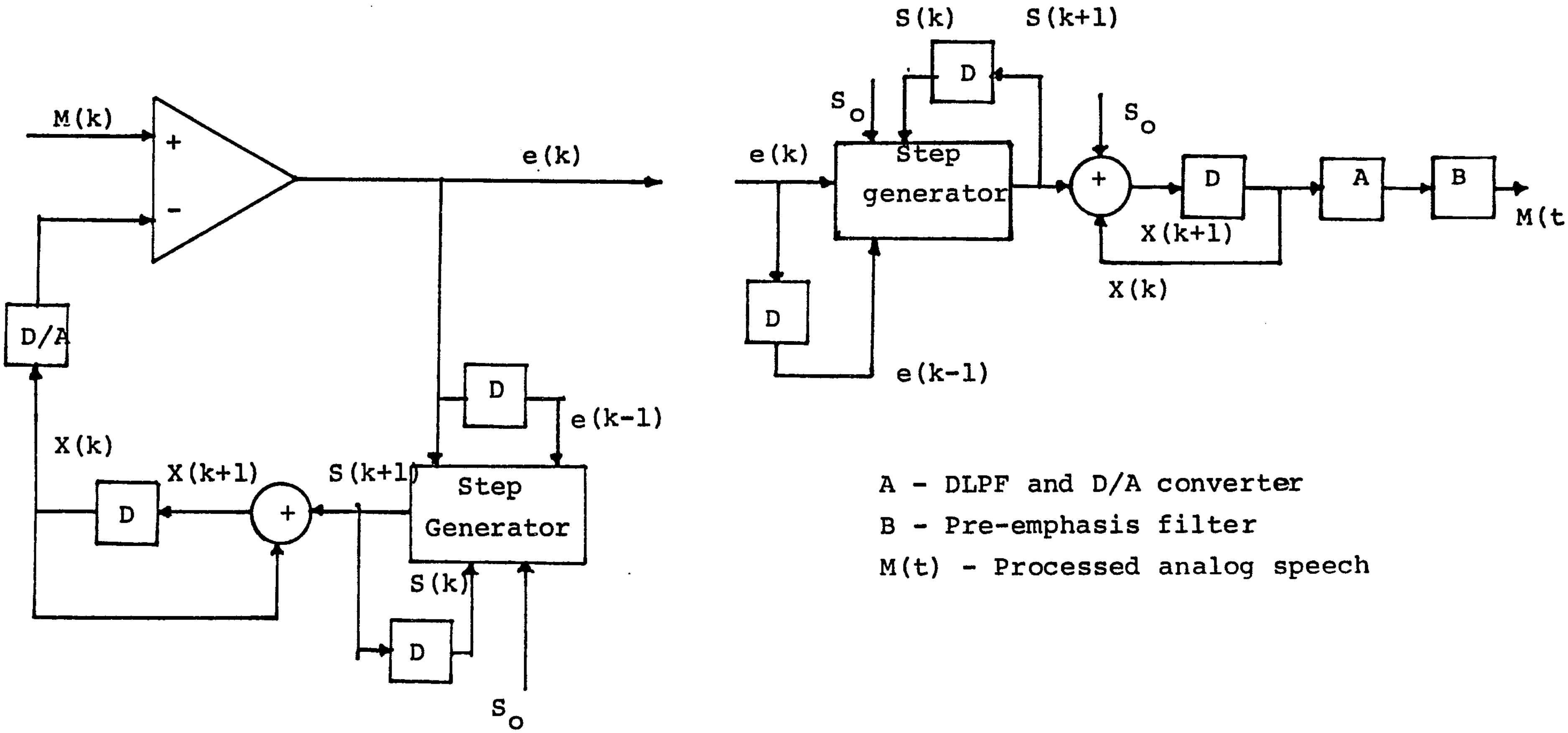
The SVADM appears to degrade almost linearly with the bit rate, f_s down to 8 Kb/s. However, when the SVADM was operated at bit rates lower than 8 Kb/s, the degradation increases significantly. It has been shown that the SNR reduces by 9dB when the bit rate is reduced by a factor of 2 to 1 [14]. This has been true in Fig.3.6.4 as long as the bit rate is above 8 Kb/s. Below 8 Kb/s, the

SNR reduces more rapidly.

(5) Linearity:

The system is said to be linear, if the gain from input to output is constant for all input levels. The linearity test was performed to determine the degree that the output to input amplitude ratio of the SVADM is constant for different input levels. A test signal of 1KHz was used. When the input level is varied from Odb (full load) to 35 dB below maximum the output level was within $\pm$ 0.5 dB.

The ultimate test to evaluate the performance of the SVADM is the subjective quality of the processed speech. In the next chapter, we compare the performance of the SVADM and the commercially available CVSD.



Encoder   Decoder

A - DLPF and D/A converter
B - Pre-emphasis filter
M(t) - Processed analog speech

Fig. 3.1.1   Block diagram of the SVADM

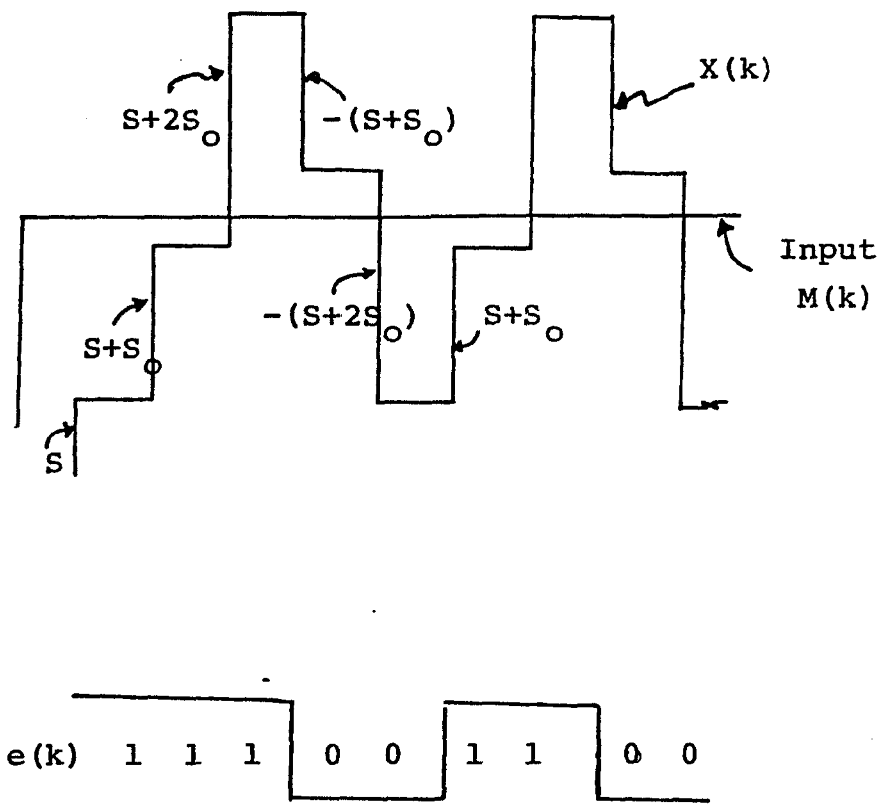
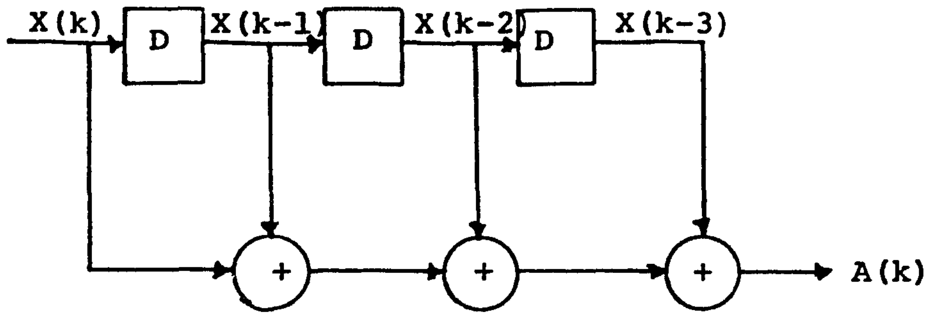
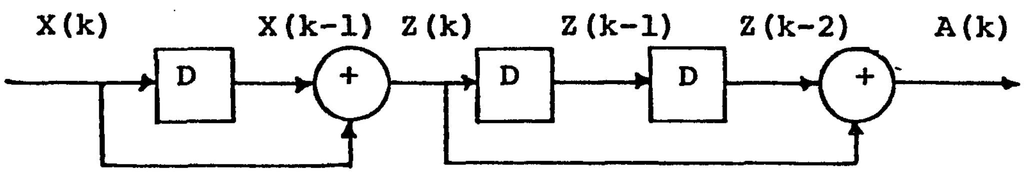


Fig. 3.1.2 Response of the SVADM to a step input



(a)



(b)

Fig. 3.2.1 Implementation of the DLPF

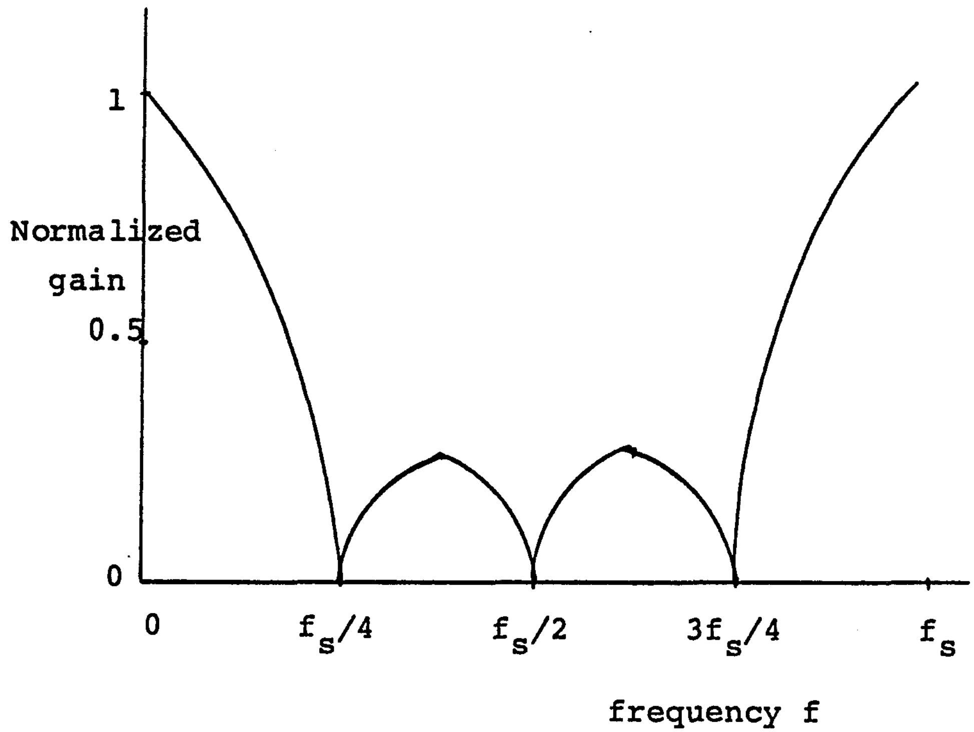


Fig. 3.2.2 Frequency response of the DLPF

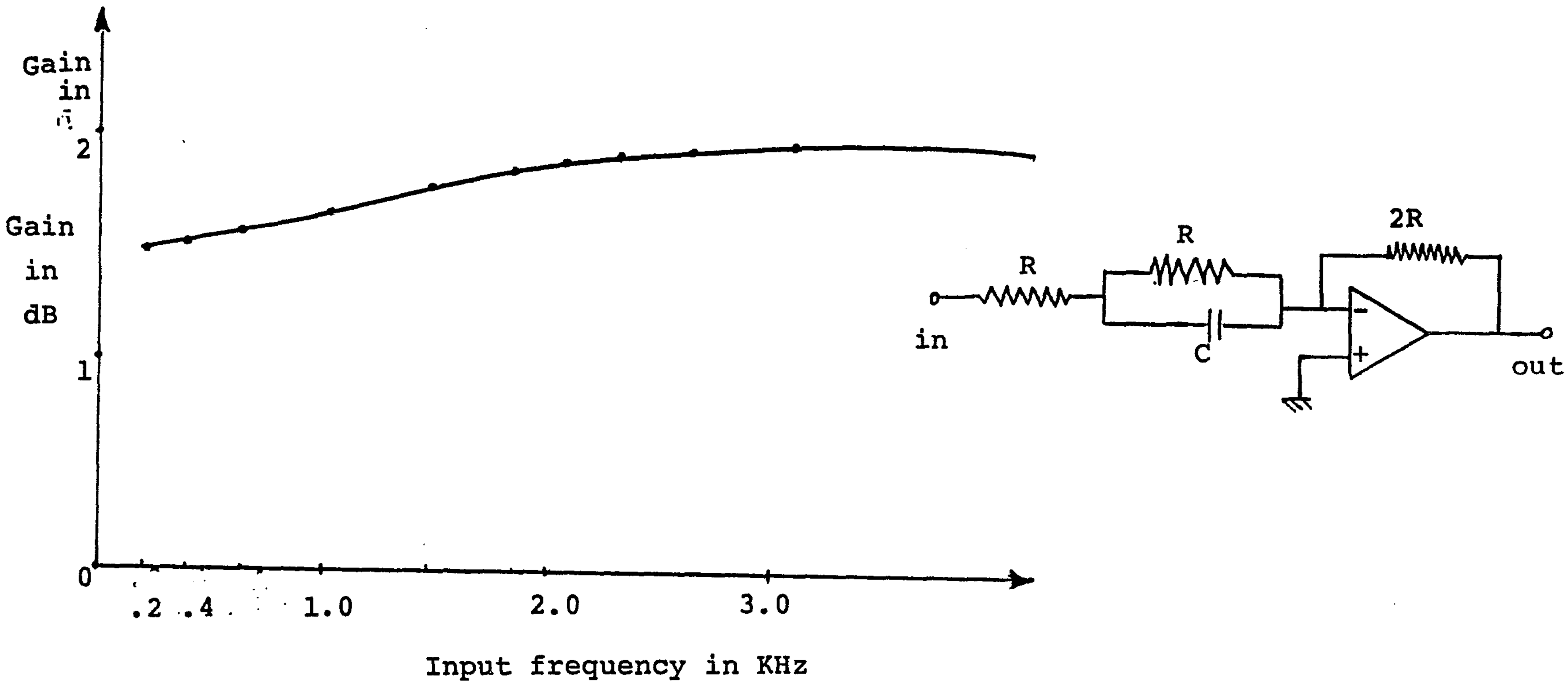


Fig. 3.3.1 Pre-emphasis filter and its frequency response

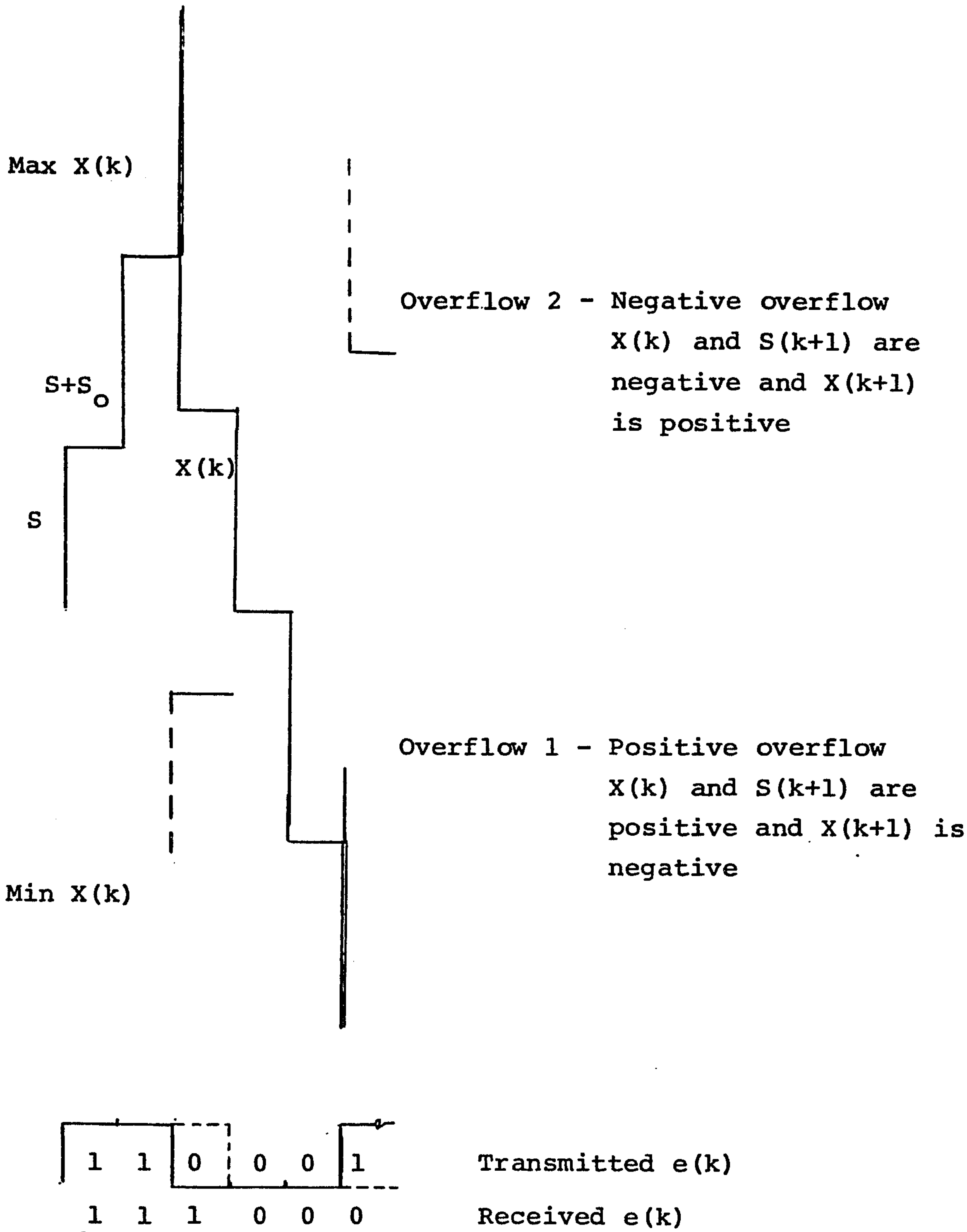


Fig. 3.4.1 Overflow conditions

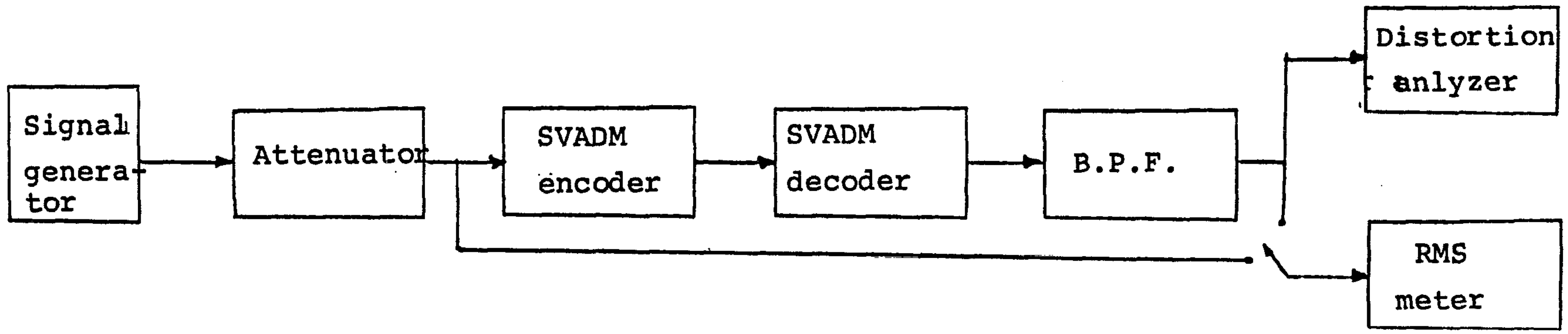


Fig. 3.6.1 Test set up to evaluate the SVADM

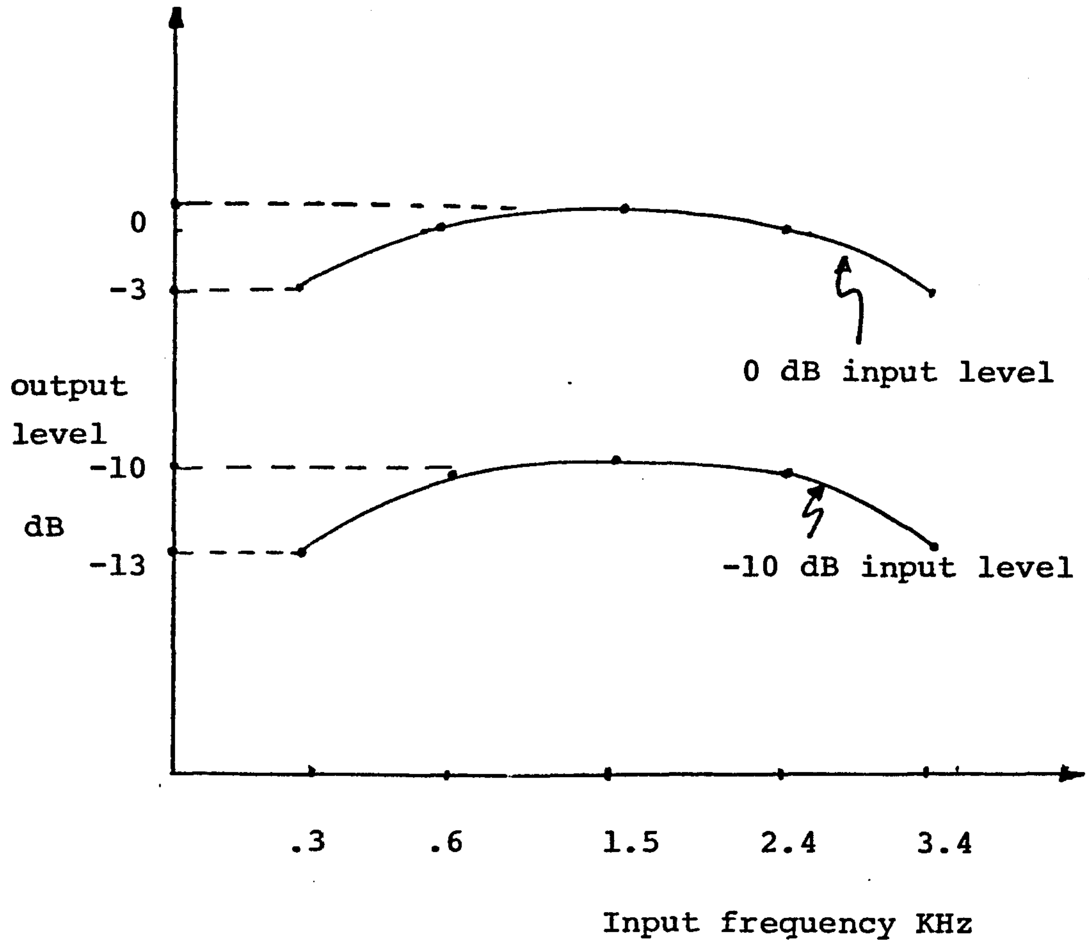


Fig. 3.6.2 Bandwidth test

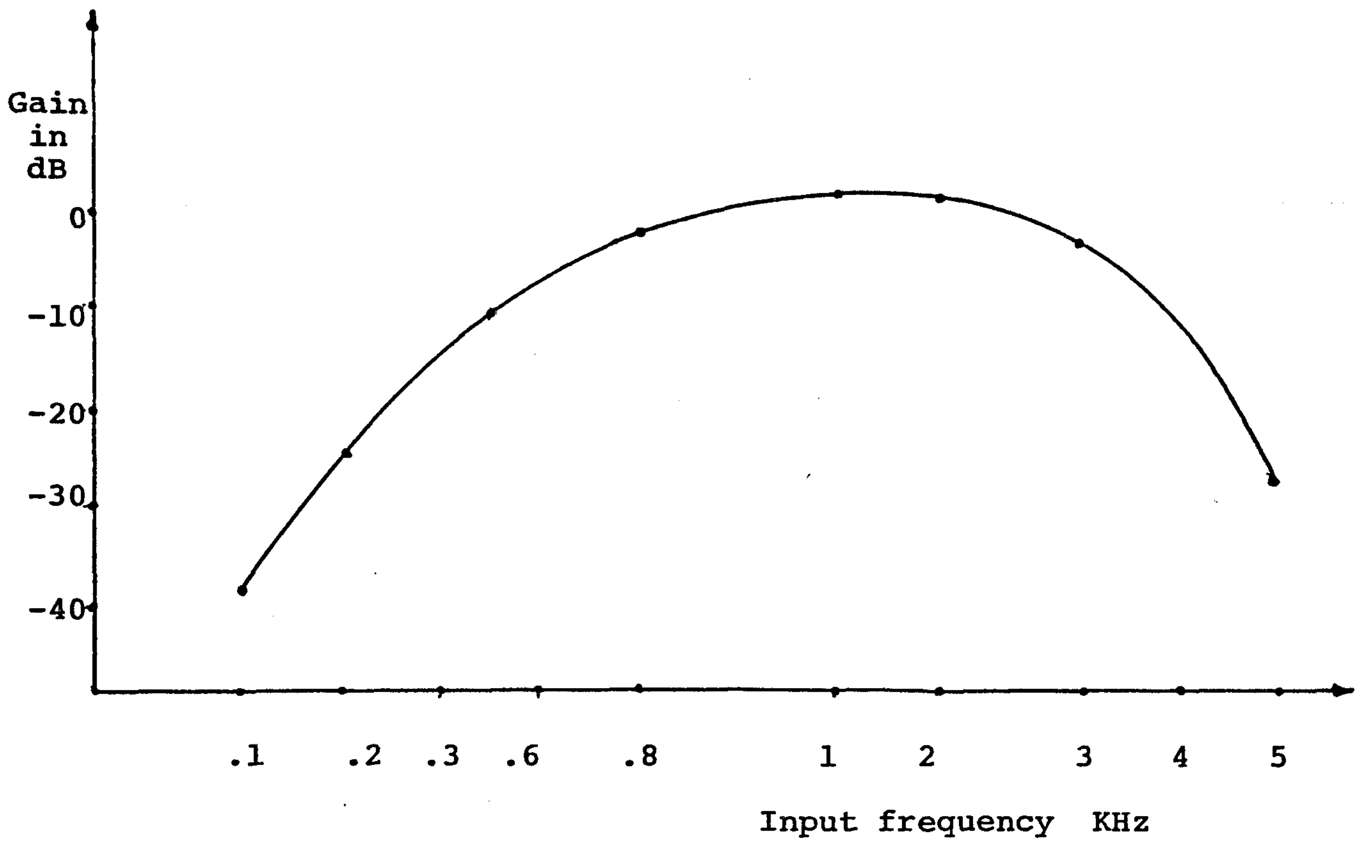


Fig. 3.6.3 C-Message weighted filter

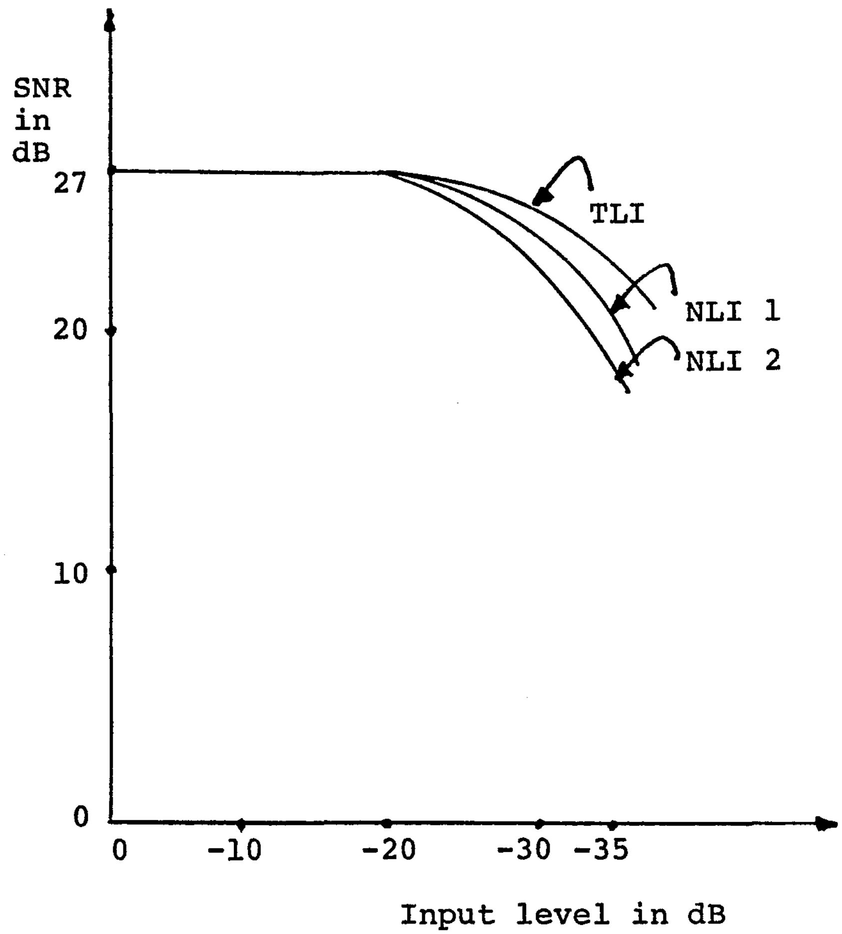


Fig. 3.6.4 Dynamic range of the SVADM

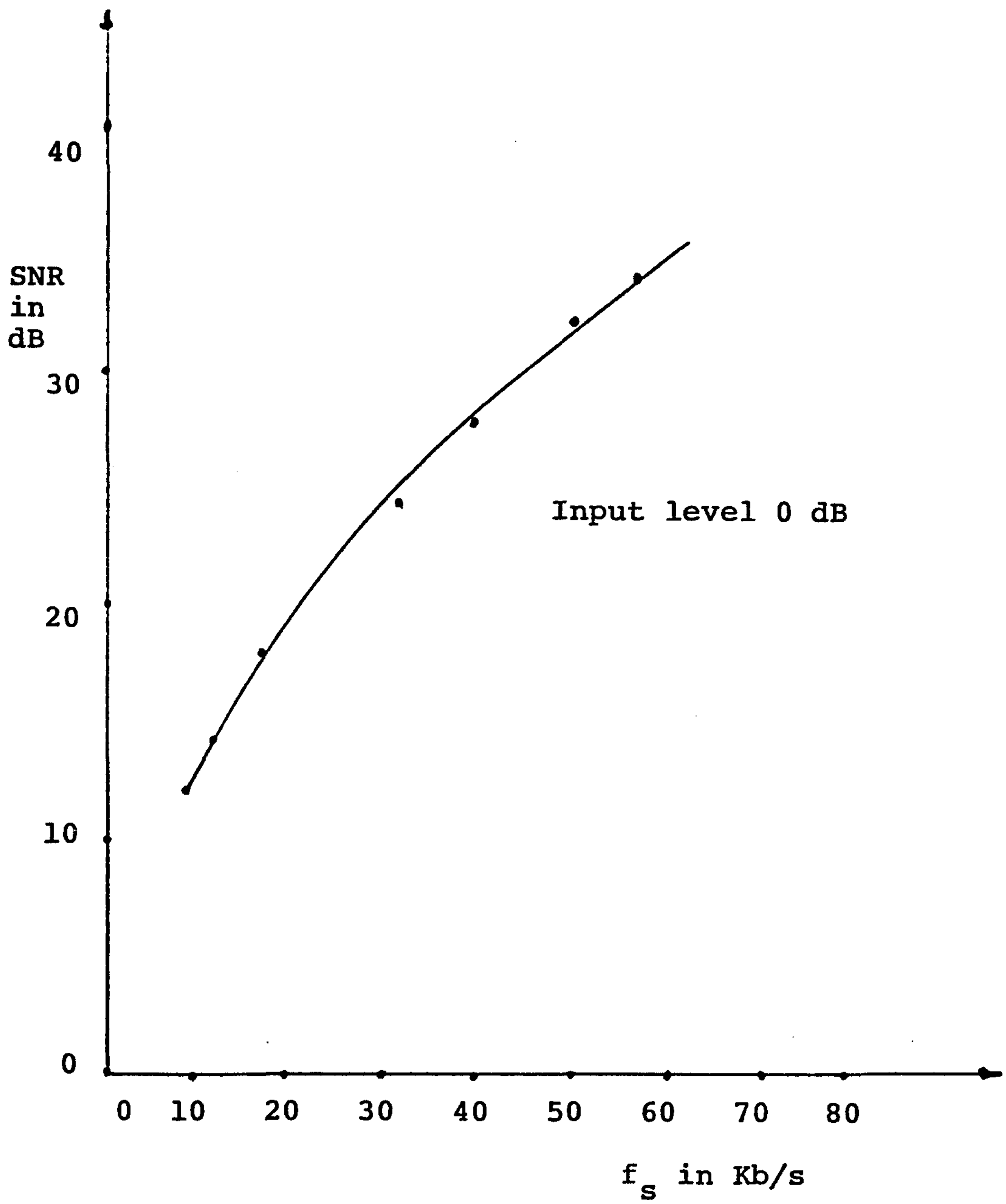


Fig. 3.6.5 SNR vs f_s

Table 3.5.1 Leak factor as a function of bit rate

Bit rate f_s Kb/s	Leak factor L
32	(1-1/128)
16	(1-1/64)
9.6	(1-1/32)
For practical implementation L was chosen to be (1-1/64)	

Chapter 4

Performance Comparison of SVADM and CVSD

The subjective quality of the processed speech is the most important measure in describing the delta modulation system. In this chapter, we describe the subjective comparison of the CVSD and the SVADM.

4.1 Subjective Comparison of the Performances of the CVSD and the SVADM:

Figure 4.1.1 shows the test set up used for the subjective comparison test. It consists of a speech source, a band pass filter (B.P.F.), the SVADM encoder-decoder, the CVSD encoder-decoder, a B.P.F., and monitoring devices. For all the tests, the speech was bandlimited from 300 Hz to 2500 Hz. Three speech tapes were used which are:

1) A Mark Twain story.
2) A radio conversation.
3) A standard set of group words.

Several subjective tests to evaluate the system performance have been performed. All the tests have been repeated at many different times. The number of listeners, who took the test was around forty. The following tests were performed:

4.1.1 Listeners Preference Test

In this test, the performance of the SVADM and the CVSD were compared at different bit rates. During the test, several comments were made by the listeners of the processed voice.

With these comments, we were able to establish the result which has been tabulated in Table 4.1.1.

We defined the intelligibility and the acceptability in the following way.

Intelligibility: The speech is intelligible if it is understandable.

Acceptability: The speech is acceptable if words or syllables are not missing.

For this test, the input level was set to OdB (full load). From the Table 4.1.1, we infer that the performances of the SVADM and the CVSD were the same at f_s = 32 Kb/s and 24 Kb/s. However, at f_s = 16 Kb/s, 10 Kb/s and 8 Kb/s, the listeners showed a clear preference to the SVADM over the CVSD. At the Nyquist rate of 5 Kb/s, the outputs of both the SVADM and the CVSD are not intelligible.

4.1.2 Dynamic Range Test

We have seen a 30 dB dynamic range for the SVADM, when using a sinusoidal tone of 1 KHz. The Motorola CVSD also claims to have a 30 dB dynamic range using single tone as the input to the CVSD . The ultimate test performance, however, is the subjective quality of the processed speech.

In this test, we varied the input signal from OdB (full load) to -40 dB in steps of 10 dB. With the available comments from the listeners, the results are tabulated in Table 4.1.2. From the Table, we note that the SVADM exhibited a dynamic range of 40 dB at f_s = 32 Kb/s, while the CVSD has a 30 dB dynamic range at the same bit rate. As we lower the bit rate, the dynamic decreased for both systems. Figure 4.1.2 shows the subjective dynamic ranges of the CVSD and the SVADM as a function of the bit rate. As we see from the figure, the dynamic range of the SVADM is about 10 dB higher than that of the CVSD.

4.1.3 Cascading (Tandem Connection)

To determine the performance of a working system, two delta modulators were cascaded. The speech processed by the first DM system is then processed by the second DM system. The subjective quality of the speech processed by the second DM system was studied. This type of tandem connection is referred to as cascading. Cascaded CVSD loses intelligibility for input signal levels of -30 dB and -40 dB at all bit rates from 32 Kb/s down 9.6 Kb/s. The cascaded SVADM performs similar to the SVADM without cascading at all bit rates. In addition the cascaded SVADM has a much higher intelligibility as compared to the cascaded CVSD even at 10 dB and -20 dB input signal levels. This test is very useful to study the feasibility of using the delta modulators in repeater stations.

4.2 Subjective Comparison of the Performances of the CVSD and the SVADM in the Presence of Channel Errors

In this experiment we simulated channel errors using the following method:

4.2.1 Generation of Channel Errors

Figure 4.2.1 shows a method of generating random errors. This scheme comprises a noise generator, a comparator and combinatoric logic. The noise generator produces an analog gaussian noise voltage of the comparator and is varied to generate different error rates. The noise voltage is compared to V_t and if it exceeds V_t, the D flip-flop shown in Figure 4.2.1 is set, causing an inversion of the logic state of the transmitted e (k). In order to determine the error rate, it is necessary to detect the state of the D flip-flop at every clock cycle of the delta modulator. The error rate is given by the ratio of the error count to the clock count. The CVSD and the SVADM were compared at different error rates.

4.2.2 Subjective Comparison of the CVSD and the SVADM as a function of error rate.

Figure 4.2.2 shows the test set up for subjective evaluation of the CVSD and the SVADM in the presence of errors. The input speech signal was bandlimited from 300Hz to 2500Hz. The error rate is varied by varying the threshold voltage in the error generator. The error rates chosen for the test were 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1} and $2(10^{-1})$. The parameters varied in this test were the input level, error rate and the bit rate. The results are tabulated in Table 4.2.1. From the table, we infer that at odB input level, fs = 32 kb/s and the error rate of 10^{-1}, the CVSD was preferred to the SVADM. Under all other operating conditions, the SVADM was preferred to the CVSD. At lower input levels i.e. from - 20dB and down, the SVADM is significantly better than the CVSD at all error rates. In addition, the subjective dynamic ranges of both the CVSD and the SVADM decrease with the error. Using the results from Table 4.2.1., we also plotted the dynamic range as a function of the error rate for both the CVSD and the SVADM. The results are shown in Fig.4.2.3. From Fig. 4.2.3., we conclude that the SVADM has about 10-15 dB higher dynamic range over that of the CVSD even in the presence of errors. We have already shown in Fig.4.1.2, that the SVADM has a 10-15 dB higher dynamic range over that of the CVSD with no channel errors.

We have tested the CVSD developed by two different manufacturers and the SVADM to compare them in terms of bit rate, bit error rate, input level etc. We have used a special purpose micro-computer that was designed and developed in the Communication Laboratory[1]to realize the SVADM encoder or decoder or both for comparison with the CVSD. Appendix B describes the SVADM encoder program just to illustrate the use of the micro-computer. We have also used the SVADM developed by the Delta Modulation Incorporated for our experiments. In the following chpters, we present the use of delta modulators as source encoders in packet voice networks.

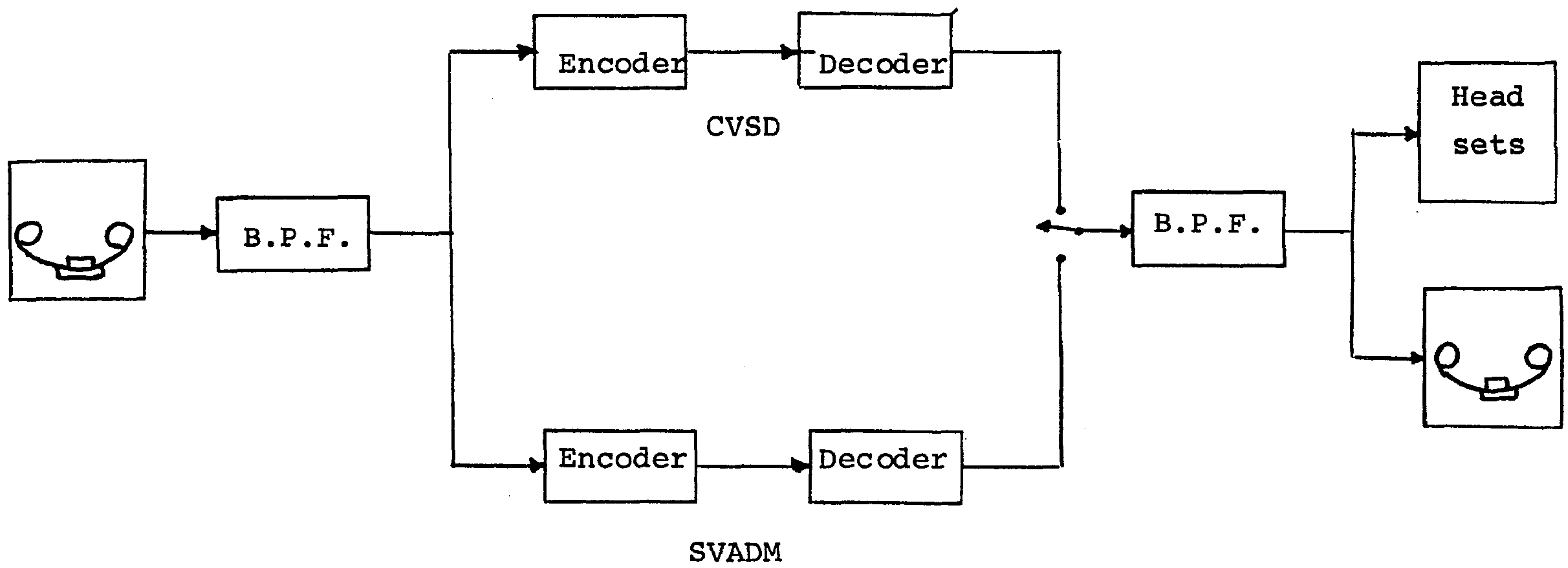


Fig. 4.1.1 Test set up for subjective comparison of the CVSD and the SVADM

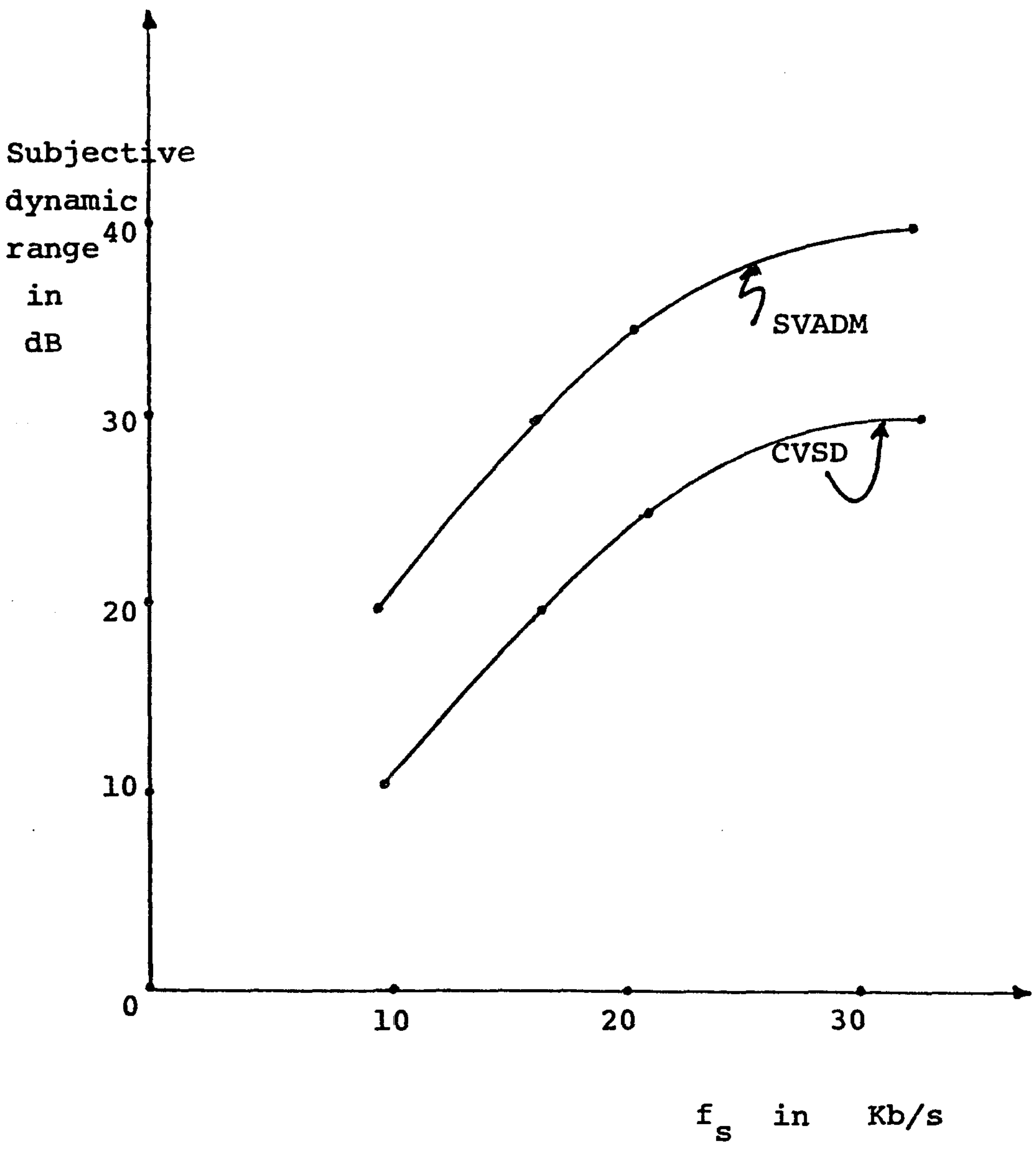


Fig. 4.1.2 Dynamic range as a function of bit rate

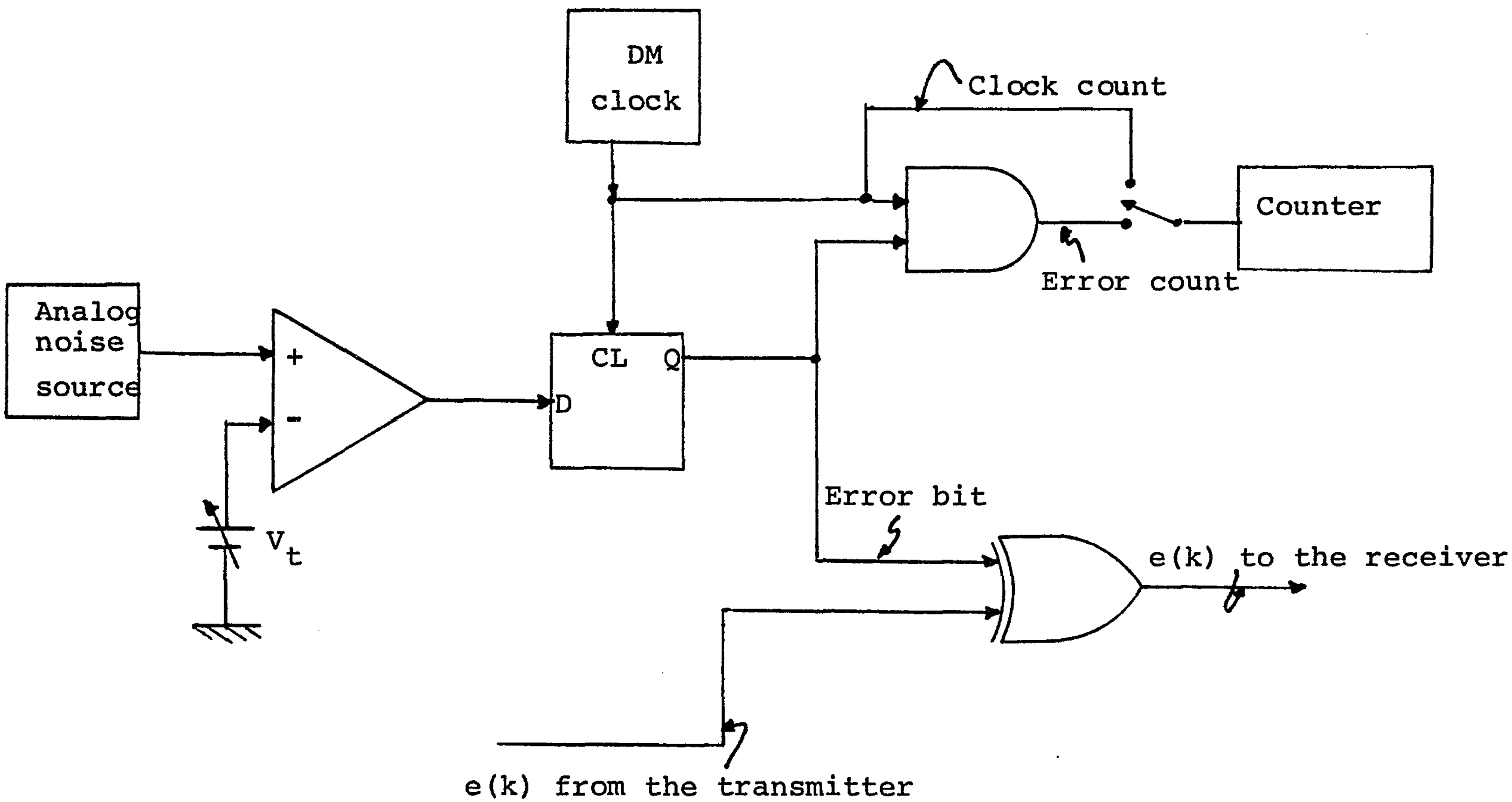


Fig. 4.2.1 Generation of random errors

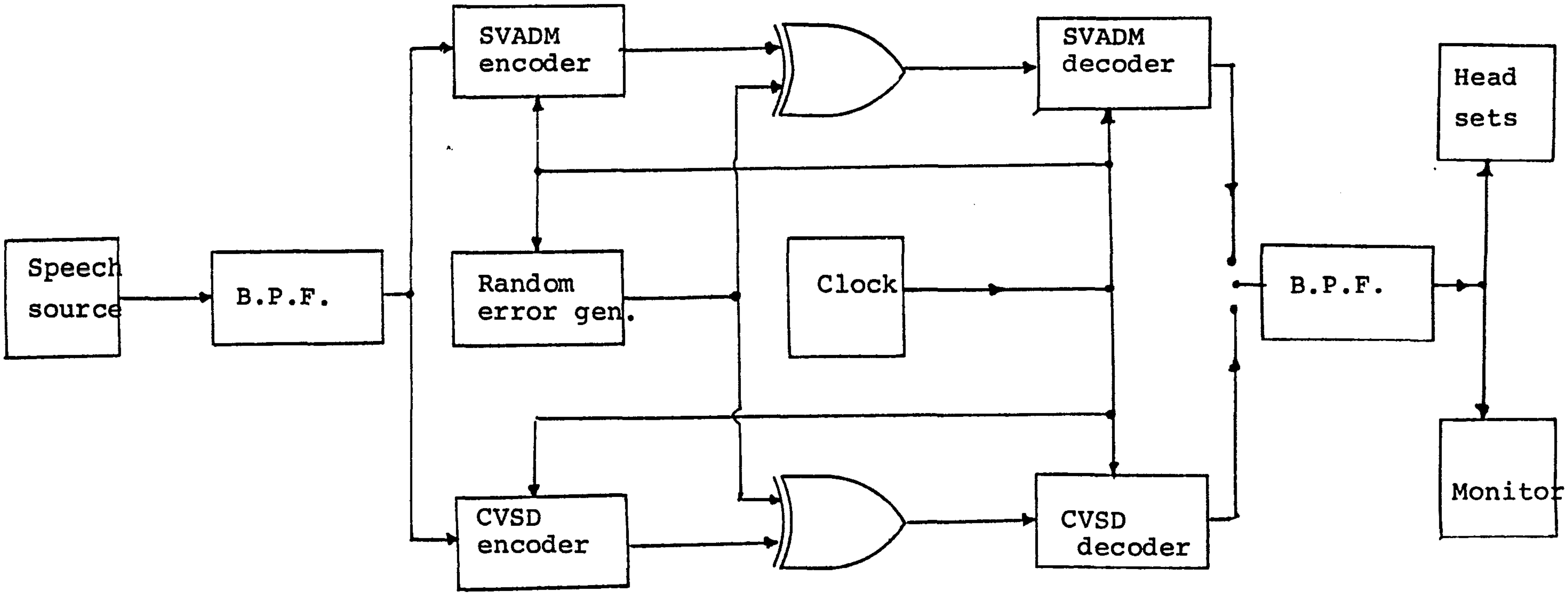


Fig. 4.2.2 Test set up for comparison of the CVSD and the SVADM in the presence of bit errors

(a) SVADM at f_s = 32 Kb/s
(b) CVSD at f_s = 32 Kb/s
(c) SVADM at f_s = 16 Kb/s
(d) CVSD at f_s = 16 Kb/s
(e) SVADM at f_s = 9.6 Kb/s
(f) CVSD at f_s = 9.6 Kb/s

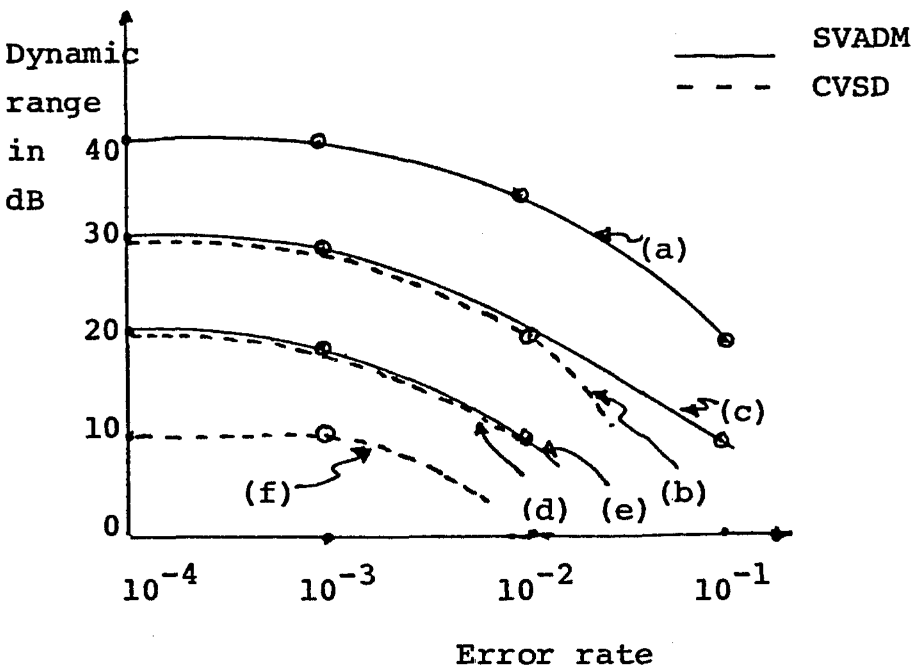


Fig. 4.2.3 Dynamic range as a function of error rate

Table 4.1.1 Subjective comparison of the CVSD and the SVADM as a function of bit rate

Input speech level 0 dB and speech bandlimited from 300 Hz - 2500 Hz

f_s Kb/s	Motorola CVSD Subjective	SVADM Subjective	Comments
32	Intelligible Acceptable	Intelligible Acceptable	Both are same
24	Intelligible Acceptable	Intelligible Acceptable	Both are same
16	Intelligible Acceptable	Intelligible Acceptable	Very slight preference to SVADM
10	Buzzy has slope over-load noise	Smooth back-ground noise due to gran-ularity	SVADM preferred
8	Noisy	Less noisy	SVADM preferred

Table 4.1.2 Subjective comparison of dynamic ranges of CVSD and SVADM.

f_s Kb/s	Input dB	CVSD	SVADM	Comparison
32	0	Intelligible	Intelligible	No difference
	-10	Intelligible	Intelligible	No difference
	-20	Intelligible	Intelligible	No difference
	-30	Intelligible: Voice is breaking and buzzy	Intelligible	SVADM preferred
	-40	Not intelligible	Intelligible: Voice is buzzy	SVADM preferred
16	0	Intelligible	Intelligible	No difference
	-10	Intelligible	Intelligible	SVADM preferred
	-20	Intelligible: Voice is buzzy	Intelligible	SVADM preferred

Table 4.1.2 continued

f_s Kb/s	Input dB	CVSD	SVADM	Comparison
	-30	Not intelligible	Intelligible: has background noise	SVADM preferred
	-40	Not intelligible	Not intelligible	
9.6	0	Intelligible	Intelligible	No difference Granularity due to f_s exists
	-10	Intelligible	Intelligible	SVADM preferred
	-20	Not intelligible	Intelligible: Noisy	SVADM preferred
	-30	Not intelligible	Not intelligible	
	-40	Not intelligible	Not intelligible	

Table 4.2.1 Subjective comparison of the CVSD and the SVADM at different error rates

f_s Kb/s	Input level dB	Error rate	CVSD	SVADM	Comparison
32	0	10^{-4}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
		10^{-3}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
		10^{-2}	Intelligible: With back ground noise (smearing)	Intelligible: With back ground noise	No preference
		10^{-1}	Intelligible: More noise	Intelligible: More noise	CVSD preferred
32	-20	10-4	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference

Table 4.2.1 continued

f_s	Input level dB	Error rate	CVSD	SVADM	Comparison
		10^{-3}	Barely intel-ligible	Intelligible: Same as at no errors	SVADM preferred
		10^{-2}	Barely intel-ligible	Intelligible	SVADM preferred
		10^{-1}	Not intel-ligible	Not intel-ligible	
16	0	10^{-4}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
		10^{-3}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference

Table 4.2.1 continued

f_s	Input level dB	Error rate	CVSD	SVADM	Comparison
16	0	10^{-2}	Intelligible	Intelligible	SVADM preferred
		10^{-1}	Not intelligible	Not intelligible	
16	-20	10^{-4}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
		10^{-3}	Barely intelligible	Intelligible	SVADM preferred
		10^{-2}	Barely intelligible: Not acceptable. Heavy background noise	Intelligible: Fluttering noise	SVADM preferred
		10^{-1}	Not intelligible	Not intelligible	

Table 4.2.1 continued

f_s Kb/s	Input level dB	Error rate	CVSD	SVADM	Comparison
9.6	0	10^{-4}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
		10^{-3}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
		10^{-2}	Intelligible: Noisy	Intelligible: Noisy	No preference
		10^{-1}	Not intel- ligible	Not intel- ligible	
9.6	-20	10^{-4}	Intelligible: Same as at no errors	Intelligible: Same as at no errors	No preference
9.6	-20	10^{-3}	Not intel- ligible	Intelligible	SVADM preferred

Table 4.2.1 continued

f_s Kb/s	Input level dB	Error rate	CVSD	SVADM	Comparison
		10^{-2}	Not intelligible	Not intelligible	
		10^{-1}	Not intelligible	Not intelligible	

5. Introduction to Packet Switching Networks

Computer communication networks were originally designed for data transfer. However, in recent years, there has been a growing need for communication of real time interactive voice over computer networks.

Initially, the Advanced Research Project Agency (ARPA) installed a network called ARPANET mainly for data transfer. This network, today, connects about 100 research organizations such as universities and military establishments mainly as a resource sharing facility. Also, several institutions are using this network to study the feasibility of transmitting packet voice along with data. A way to envision a design process is to decide whether the switching facility will be designed to handle only voice or only data or both. Since data networks already exist it is of interest, to design voice-data networks. As part of this thesis we have studied the feasibility of using adaptive delta modulators as source encoders in packet voice networks.

5.1 Why Packet Switched Networks

Packet networks have gained a wide spread acceptance with the application of teleprocessing and networking by the business sector. Thus, there is an interest to have a long haul inexpensive communication media amongst the giant country. Introduction of sophisticated packet satellite radio data communication systems is a present trend. In addition, there is a local access problem at the periphery of the network which lends itself to the use of packet switching in the form of the use of a multi-access broadcast channel in a local environment commonly known as ground radio packet switching.

A different reason for a packet switched network is to facilitate teleconferencing amongst different users. In a packet network , packets are labelled and transmitted to different destinations. Teleconferencing (conversation among several users) is accomplished by transmitting packets of a customer to several different customers. This process does not require separate lines as in the case of conventional telephony.

In order to characterize the satellite and ground radio switching systems, we describe such schemes as multi-access broadcast channels in a distributed environment. The main object is to properly share the communication channels among a collection of users. The adjective "multiaccess" refers to the fact that many users are trying to share the channel simultaneously in some cooperative fashion. The adjective "broadcast" refers to the fact that every user can transmit on a channel. The adjective "distributed" refers to the fact that users are geographically distributed in a way which makes controlling their behaviour an issue of importance An important parameter describing this notion of distributed sources is the ratio α of the propagation delay, pd (the time it takes electromagnetic energy moving at the speed of light to pass between two separated terminals) to the transmission time, pt, of a packet. As an example, the transmission time of a packet is 10 ms when 1000 bits packets are transmitted over 100 Kb/s channel. For transmission of packets from the source to the destination 10 km apart, the packet propagation delay (speed of light) is approximately 30μs. This is a typical example of a ground radio packet switching system. Thus pd $<<$ pt. However, for a satellite environment, pd is of the order of 300 ms. ARPA = has conducted several tests [15] in satellite environment as well as for ground radio. The measurements and implementation results have been claimed to indicate that packet switched systems are feasible and effective.

In spite of the results, it is only fair to say that packet switching is relatively a new technology and the increased use of low cost computers gives a promising start to this new field.

5.2 Terminology in Packet Switched network

In computer communication systems, we have a great need for sharing expensive resources a collection of high peak to average (i.e. "bursty") users. In Fig.5.2.1. we display the skeleton of a packet switched network in which we can identify three kinds of resources:

1. Each user has access, to a terminal. Every terminal is connected to its Host computer with a required communication resource. Usually this is an expensive portion of the system.

2. The Host computers which provide information processing services - have multi-access time sharing provide the mechanism for efficient resource sharing.

3. The communication subnetwork, consisting of communication trunks and software switches, whose function is to provide the data communication service for the exchange of data and control among the other devices.

The Host computers mentioned above contain hardware and software resources (in the form of application programs and data files) whose sharing comes under the topic of time sharing. As far as the packet communication is concerned, the most visible impact has been the development of the communication subnetwork. Here packet communication first demonstrated its enormous efficiencies in the form of ARPANET in the early 1970's (15). The communication resources to be shared in this case are the storage capacity at the Interface Message Processors (IMPs), processing capacity at the IMP and communication capacity of the trunks connecting the IMPs. Packet switching in this environment has proven to be a major technical breakthrough in providing cost effective data communications.

A packet is a group of binary digits including data and call control signals which is switched as a composite whole. Packet switching is the transmission of data by means of addressed packets whereby a transmission channel is occupied for the duration of transmission of the packet only. The channel is then available for use by packets being transferred between different data terminal equipment. The IMP is a device at each node of the ARPANET which performs message switching and interconnects the "Hosts" with the high bandwidth leased lines.

In order to understand the function of a packet voice network, we think of the example illustrated in Fig. 5.2.1. When user A asks for a connection to user B, the customer A's data are then assembled into packets. These packets are then provided with headers and transmitted, interleaved with other packets from one IMP to another depending on the availability of a channel at that instant. When the packets reach the IMP for which user B has the access, the connection is made between A and B. Since packets are transferred in different routes, the connection between A and B is referred to as 'virtual connection'. This kind of virtual connection is extremely advantageous to design an efficient packet switching network.

There are some packet networks which have neither interleaving facility nor virtual connection facility nor both. A particular case is the ETHERNET developed by the Xerox corporation. ETHERNET is a broadcast communication network, which is more like a BUS structure. At any given time, the packets belonging to only one customer(BIU*) are available on the network and every user has the access to those packets. Thus, it does not have a virtual connection facility.

For speech transmission using packet networks, it is advantageous to have interleaving facility, since there are silent periods in speech. We can detect silent periods and not transmit any packet during these periods and thus allow some other user to transmit packets. Interleaving facility would allow the packets from different sources to be transmitted on a first come first serve basis. In addition, for efficient networking, provision of a virtual connection facility is made available to facilitate large number of users to access a small number of channels in some networks such as ARPANET. In this thesis, we recognize that a packet switched network in general might contain both interleaving and virtual connection facilities.

With this basic background in packet switching networks, we have studied the use of delta modulators as source encoders in a packet voice network. We have studied the effect of packet loss, packet size and bit rate on the processed speech with the SVADM and the CVSD as source encoders. In order to improve the efficiency of a network we developed algorithms for detecting silent periods and not transmit any packet during these periods.

* BIU -Bus Interface Unit.

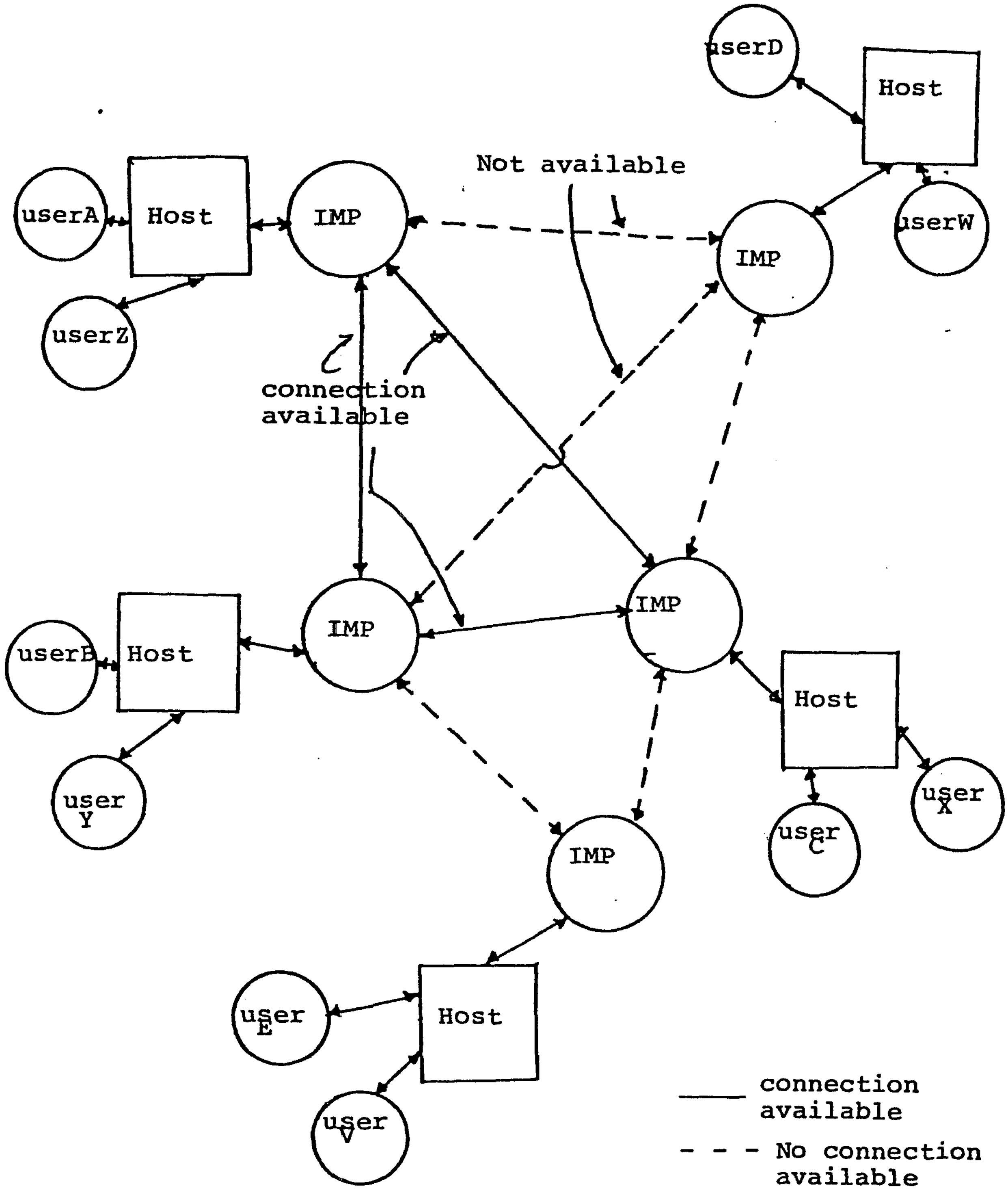


Fig. 5.2.1 Skeleton of a packet switching network

Chapter 6

Delta Modulators as Source Encoders in Packet Voice Networks

Current methods used for digitizing voice packet networks are the Pulse Code Modulation (PCM), Adaptive Delta Modulation (ADM) and the Linear Predictive Coding (LPC). If PCM is used to encode 2.5 KHz voice, one would require a bit rate of at least 40 Kb/s to produce good quality voice. A packet size of 1000 bits requires that the PCM packets be transmitted at the rate 40 packets/sec. The ADM systems produce good quality voice, when operated at 10-16 Kb/s. For the same packet size the ADM packets can be transmitted at the rate of 10-16 Kb/s. The ADM is also preferred to the LPC, since the LPC is still a relatively high cost and complex system. The ARPA network is currently employing the CVSD algorithm to digitize voice. Therefore it is appropriate to compare the use of the SVADM with that of the CVSD in a packet voice network. We have already shown that the performance of the SVADM is preferred to that of the CVSD when operated at bit rates of 16 Kb/s and lower. We have compared the performance of the SVADM in a packet voice network interms of packet size (P), bit rate (f_s) and packet loss rate (r), with that of the CVSD.

6.1 Concept of Packet Loss

In a packet switched network when a customer A (source) asks for a connection to a called party B (destination), the customer's packets are then transmitted, interleaved with other packets from one exchange to another, thus giving a virtual connection. Once the contact has been established between A and B, B would be receiving a virtually continuous stream of packets as long as A is active. As the packets arrive the destination B processes them. Thus while the i^{th} packet is being processed, B looks for the $(i+1)^{st}$ packet. If the $(i+1)^{st}$ packet is not available for processing after B has completed processing the i^{th} packet, then we recognize the $(i+1)^{st}$ packet as being lost. In a normal operation the destination B can lose the $(i+1)^{st}$ packet in one of two different ways as follows:

a) The $(i+1)^{st}$ packet actually arrived at B but was rejected as non-valid. When a non-valid packet is received, the request for retransmission is not required in voice transmission since voice systems using ADMs generally tolerate reasonable error rates and besides, the delay constraints preclude the use of retransmission of packets any way.

b) The $(i+1)^{st}$ packet has not arrived (i.e. it is late) even after B has completed processing the i^{th} packet. After waiting for an appropriate period the destination B then, will decide that the $(i+1)^{st}$ packet is lost and starts looking for the $(i+2)^{nd}$ packet.

6.2 Effect of Packet Loss

When the destination B decides that a packet is lost and starts processing the next packet, the reproduced speech signal exhibits loss of speech. If, for example, the speech is encoded at 16 Kb/s and the packet size is 1 kbits, the fraction of speech lost due to a single packet loss is $(1/16)^{th}$ of a second or approximately 60 msec. The degradation of the quality of the speech due to 60 msec. of speech loss is minimal. This is because the human ear is insensitive to the small amount of degradation. Also, if one of every hundred packets is lost, then 60 msec. of speech is lost in 6 sec. of speech and this too does not adversely affect the quality of the processed speech.

When a packet is lost, the state of the delta modulator decoder is different from that of the encoder (similar to bit error described earlier). However, this will be corrected by the error correction logic as described earlier.

6.3 Compensation Algorithms at the Receiver

In addition to the earlier mentioned error correction technique, in order to help the receiver in its correction process, we have developed compensation algorithms for use by the receiver during the length of the packet loss. Three different compensation algorithms have been studied.

Algorithm 1: Freeze the decoder.

In this algorithm, the state of the decoder remains constant or is frozen during the packet loss period. This is done by inhibiting the sampling clock pulse to the decoder during the entire length of the missing packet. This enables the decoder to remain at the same state; i.e. the receiver step size and the estimate remain the same until a new packet is received. The encoder, however is changing its state continuously. Thus the state of the decoder is different from that of the encoder when a new packet arrives. This will be eventually corrected by the leaky integrator routine. During the packet loss, freezing the decoder usually creates a large step size error.

Algorithm 2: Generate a local periodic 1 1 0 0 1 1 0 0 · · · steady state pattern at the receiver.

In this method, the receiver will locally generate a 1 1 0 0 1 1 0 0 · · · pattern for the entire packet loss period. The steady state pattern for the entire packet loss period. The steady state pattern at the decoder input would enable the receiver estimate to leak to zero level, during the period of a lost packet. However, the step size error remains unchanged. It must be noted that the steady state pattern 1 1 0 0 1 1 0 0 · · · is only applicable to the SVADM decoder and not the CVSD decoder. The steady state pattern generates an oscillation at $f_s/4$ and is usually heard at low bit rates.

<u>Algorithm 3</u>: Generate a local periodic 1 0 1 0 1 0 · · · steady state pattern at the receiver.

In this algorithm, the receiver will locally generate a 1 0 1 0 1 0 · · · pattern instead of 1 1 0 0 1 1 0 0 · · · as in algorithm 2. This pattern at the input of the decoder enables the step size to become smaller. However, the estimate error remains approximately the same. The smaller step size in the decoder is extremely advantageous, since it will prevent any large variation of the magnitude of speech due to an error at the input. This is particularly more pronounced at high error rates. In addition, at low bit rates, the oscillations at $f_s/2$ is not heard. Even though the step size due to this pattern reaches a minimum, the adaptive step size algorithm enables the decoder step size to grow fast once the new packets are processed.

Figure 6.3.1 displays the receiver estimates of the three methods during a packet loss period.

6.4 Experimental Results

Figure 6.4.1 describes the system used for packet loss studies. Speech output of the tape recorder is band limited from 300 Hz to 2500 Hz and used as the inputs to the SVADM encoder and the CVSD encoder. The packetizer, the depacketizer and the loss of packets were simulated by the PDP-11/34 computer for real time operation. The output bits of the depacketizer are then encoded by the respective decoders. Two types of speech tapes were used.

1) A Mark Twain story
2) A general radio conversation.

The parameters for the subjective quality test are

a) Packet size P, where P = 2048, 1024, 512, 256 bits
b) Packet loss rate r, where $r = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 2(10^{-1})$
c) Sampling rate f, where f = 16, 9.6 Kb/s

The subjective comparison of the CVSD and the SVADM in terms of P, r, f_s is tabulated in Table 6.4.1. At the maximum input level (0 dB), the performance of the packet voice system using the CVSD or the SVADM was found to be about the same. However, at lower input levels, there is a general degradation in the performance of the CVSD as found to be true in the subjective dynamic range test described earlier.

There was no difference in the performance regarding the intelligibility using the three receiver algorithms for packet loss. However, for the SVADM encoder-decoder, using algorithms 1 and 2, there was a large change in the estimated speech due to large step size errors. This change was some times annoying to the listeners, particularly at high packet loss rates ($r = 10^{-1}$). This effect was not found when using algorithm 3.

As seen from the Table 6.4.1, a loss rate of 10^{-2} was not noticeable. The breaks in speech were distinguishable only at loss rates of 10^{-1} and $2(10^{-1})$. However, the speech was intelligible even at loss rates of 10^{-1}. This result is true for $P = 2048, 1024, 512, 256$ and $f_s = 16, 9.6$ Kb/s..

From our results, we conclude that the packet voice network using delta modulation source encoders can safely operate at loss rates of 10^{-2}.

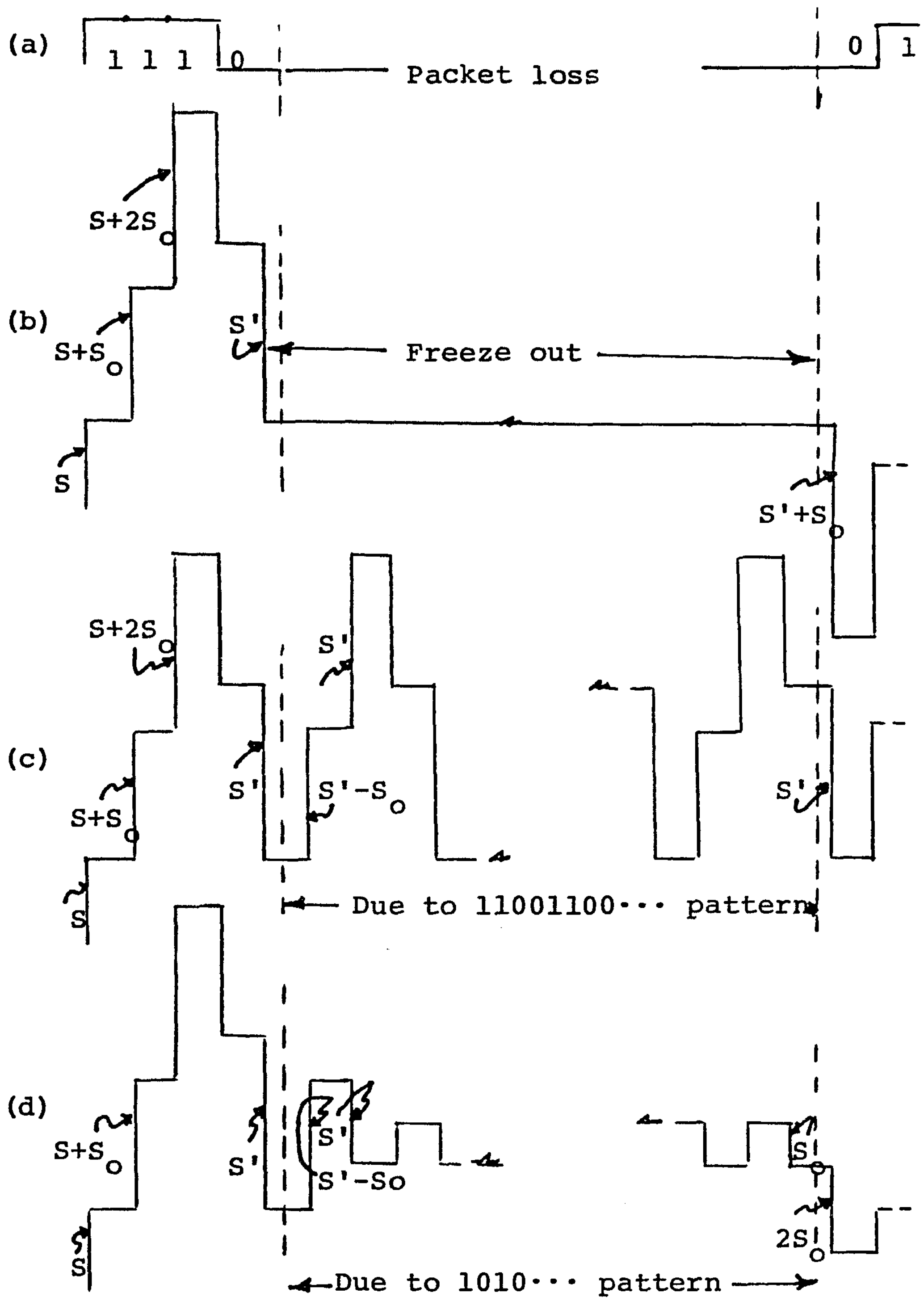


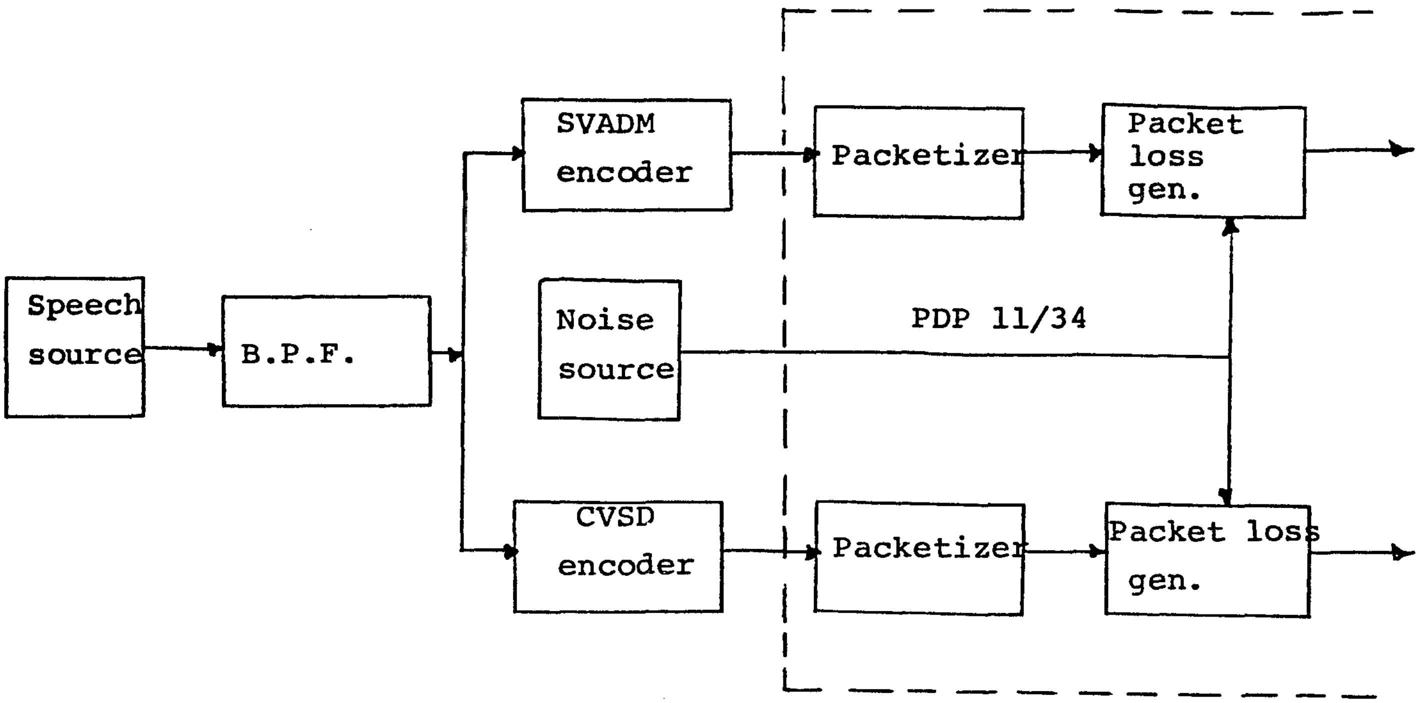
Fig. 6.3.1 Receiver estimates using algorithms 1,2 & 3

(a) Input bit pattern at the decoder

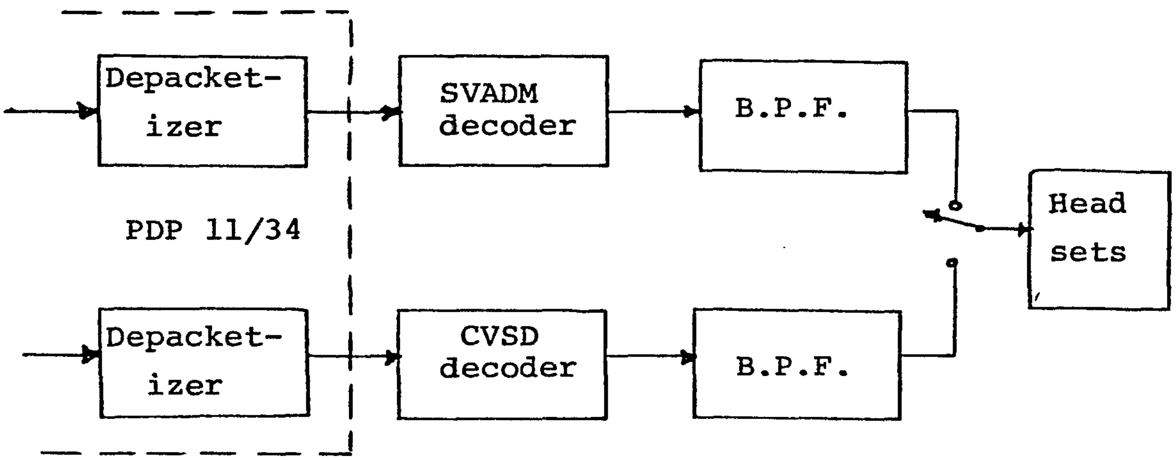
(b) Estimate using algorithm 1

(c) Estimate using algorithm 2

(d) Estimate using algorithm 3



(a)



(b)

Fig. 6.4.1 Test set up for packet loss studies
(a) Transmitter section
(b) Receiver section

Table 6.4.1 Subjective comparison of the CVSD and the SVADM in terms of P, f_s and r.

Criteria for comparison:

1. Intelligibility - The speech is intelligible if it is understandable.
2. Acceptability - The speech is acceptable if words or syllables are not missing.

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
0dB	2048	16	0	Intelligible Acceptable	Intelligible Acceptable	No preference
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when r = 0.
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
			$2x(10^{-1})$	Intelligible Not acceptable	Intelligible Not acceptable	

Table 6.4.1 continued

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
	2048	9.6	0	Intelligible Acceptable	Intelligible Acceptable	No preference
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when $r = 0$.
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
			$2x(10^{-1})$	Not intel-ligible. Not acceptable.	Not intel-ligible. Not acceptable.	
0db	1024	16	0	Intelligible Acceptable	Intelligible Acceptable	No preference

Table 6.4.1 continued

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when r = 0.
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
			$2x(10^{-1})$	Intelligible Not acceptable	Intelligible Not acceptable	
	1024	9.6	0	Intelligible Acceptable	Intelligible Acceptable	No preference. Granularity is heard.
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when r = 0.

Table 6.4.1 continued

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
			$2x(10^{-1})$	Not intelligible. Not acceptable.	Not intelligible. Not acceptable	
	512	16	0	Intelligible Acceptable	Intelligible Acceptable	No preference
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when $r = 0$.

Table 6.4.1 continued

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
0dB	512	16	10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
			$2x(10^{-1})$	Intelligible Not acceptable	Intelligible Not acceptable	
	512	9.6	0	Intelligible Acceptable	Intelligible Acceptable	No preference. Granularity is heard.
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when r = 0.
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.

Table 6.4.1 continued

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
			$2x(10^{-1})$	Not intel-ligible. Not acceptable.	Not intel-ligible. Not acceptable.	
	256	16	0	Intelligible Acceptable	Intelligible Acceptable	No preference
			10^{-4}, 10^{-3} 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are are similar to when $r = 0$.
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
OdB	256	16	$2x(10^{-1})$	Intelligible Not acceptable	Intelligible Not acceptable	

Table 6.4.1 continued

Input level	P bits	f_s Kb/s	r	CVSD	SVADM	Comparison
	256	9.6	0	Intelligible Acceptable	Intelligible Acceptable	No preference. Granularity is heard.
			10^{-4}, 10^{-3}, 10^{-2}	Intelligible Acceptable	Intelligible Acceptable	No preference. Performances are similar to when r = 0.
			10^{-1}	Intelligible Acceptable	Intelligible Acceptable	No preference. Breaks in speech are noticed.
			$2x(10^{-1})$	Not intelligible. Not acceptable.	Not intelligible. Not acceptable.	

Table 6.4.1 continued

For Input level -10, -20, -30 dB	for all conditions of P, f_s, r the SVADM is preferred to the CVSD

CHAPTER 7

Silence Detection and Speech Initiation

It has been an established fact that conversational speech contains many silent periods, sometimes as high as 50% [12]. As such, the detection of silence would enable us to reduce the packet transmission rate by not transmitting silent periods. Thus, we achieve an efficient packet voice transmission. One of the ways of detecting silence is to use an analog level detection technique such as used in the time Assigned Speech Interpolation (TASI) system. In TASI system, if the input amplitude level is below a certain specified level, it is determined as the silent period. Above that level, it is recognized as the active period. However, we have developed a digital technique of silence detection, where we recognize a certain bit pattern at the output of the encoder. We have shown earlier that the SVADM produces a periodic 11001100... in the steady state for a constant input. We detect this pattern when the input signal has a silent period.

We have developed two schemes of packet formation. In the first scheme, the packets assembled are of fixed size. Also, we introduce the concept of"silent packet", by keeping track of the number of silent bits in a packet. In the second scheme, the packets assembled vary in size between a maximum and a minimum.

7.1 Fixed Packet Size Scheme

In this scheme, the packet size is kept constant. We determine whether each packet assembled is a silent packet or a speech packet. All silent packets are not transmitted. Figure 7.1.1 shows the flow diagram of the fixed packet size scheme. In this scheme, there are two steps involved in determining a silent packet. The first step is to determine whether the input speech is in active mode or in silent mode. The second step is to keep track of the number of silent bits and speech bits in every packet to determine whether the packet is worthy of transmission or not. When the packet is determined as a silent packet, then we consider it as not worthy of transmission, otherwise the packet is worthy of transmission. The algorithms involved in the decision process is described in the following sections.

7.1.1. Algorithm for Silence Detection

All delta modulators produce a periodic output for a constant input. The SVADM produces a 11001100... pattern in the steady state for a constant input. On the other hand the CVSD encoder produces a 10101010... pattern. In order to detect the onset of silence, we shall employ an algorithm which will detect these steady state patterns.

For the SVADM, in order to determine the start of a silent period, is was decided that we shall observe eight consecutive bits of the encoder output to see if they have a 11001100 pattern (or any of the three other possible permutations of 11001100 for eight bits). If this pattern was detected, a decision that a silent period has begun was made.

The reason for choosing eight bits for detection of silence rather than four consecutive bits is due to the fact that the SVADM encoder output may have a 1100 or any one of the other permutations at the peak of the input signal and create false silence periods. Also, we have found that, when the input signal varies over the full range, no difference exists, whether we use eight or twelve consecutive bits for detection of silence. Thus, we have used a minimum of eight consecutive bits to detect the onset of silence.

Having entered a silent period, it was decided that we shall consider the signal in silence until three consecutive output bits are 0 0 0 or 1 1 1. The SVADM produces a minimum of three bits of 0 0 0 or 1 1 1 at the onset of speech. Using more than three consecutive bits of the same sign may cause the initial part of the talk spurt to be clipped. Detection of the onset of speech is not feasible using only two bits of the same sign.

For detecting the onset of silence in the case of the CVSD encoder, we look for eight bits of 1 0 1 0 1 0 1 0 1 0, since the output of the CVSD encoder in the steady state is 1 0 1 0 1 0···. Here too, we remain in the silent period until three consecutive bits of 1 1 1 or 0 0 0 are detected for speech initiation. Figure 7.1.2 shows the timing diagram for silence detection and speech initiation.

7.1.2. Silent Packets

As the transmitter assembles a packet, we keep track of the number of silence (steady state) bits by using a counter. To determine whether a packet is silent or not, we set up a threshold parameter T_p, which is a number assigned to a packet. If the ratio of the number of silence bits, S, to the total number of bits in a packet, P, exceeds T_p, then we say, the packet is a silent packet, i.e., we consider this packet not to have enough useful information to make it worthy of transmission. As such all silent packets are not transmitted. Clearly, this reduces the packet transmission rate. Figure 7.1.3. shows the discarding of silent packets. In Figure 7.1.3. p_1 and p_5 are speech packets, p_2 and p_4 are silent packets since $S/P \geqslant T_p$ and $S''/P \geqslant T_p$ respectively. However, $S'/P \leqslant T_p$ and therefore the packet p_3 is not a silent packet. In this case only p_1, p_3 and p_5 are transmitted.

By not transmitting p_2 and p_4, we loose some speech bits.For example, the initial part of the speech in p_2 is lost. Experiments have been performed to evaluate the effect of such a speech loss during transmission. The results will be presented later.

We have described the process of packetization and determination of silent packets. The packet size is kept constant and packetization is performed for every P consecutive bits. We refer to this method of packetization as Non-Repacking. Another scheme we have used is called Repacking.

7.1.3 Repacking

By repacking, we refer to the idea in which the transmitter having currently detected a silent period, halts its packetization process until such time as it detects a new speech period. Only then, will the transmitter begin the formation of a new packet. Figure 7.1.4 (a) and (b) illustrate the non-repacking and the repacking schemes respectively.

In Fig. 7.1.4 (a), we show that p_1, p_2, p_3, p_4 and p_5 are packets of size P bits. The shaded area corresponding to S, S' and S" represent silence bits in each of the packets p_2, p_3 and p_4 respectively. P_1 and p_5 are speech packets. p_2, p_3 and p_4 are silent packets, since $S/P \geqslant T_p$, $S'/P \geqslant T_p$ and $S''/P \geqslant T_p$. Thus only p_1 and p_5 are transmitted. By not transmitting p_2, p_3 and p_4, some speech bits are lost in those apckets. The speech bits lost in p_4 can be saved if the repacking scheme is used as shown in Fig. 7.1.4(b).

In the repacking scheme, after determining that p_2 and p_3 are silent packets, the transmitter recognizes that the encoder output still has silent bits and therefore will halt its packetization process. It will start packetization once it detects that the speech has been initiated and therefore the new packet is now p_4' and not p_4. Thus the repacking scheme transmits the speech bits contained in p_4 which was lost when the non-repacking scheme was used. Therefore, there is less chance of losing the onset of speech. However, the speech bits lost in p_2 and p_3 cannot be recovered in either of the schemes. It was found that repacking vastly enhances the quality of the processed speech.

7.1.4 Receiver During Silent Periods

When the transmitter decides that a packet (silent packet is not worthy of transmission, it will not send the packet. When the silent periods are not transmitted, a gap is created in the stream of received packets at the receiver. At this point, the receiver recognizes that a silent period has begun at the source. As such, it will now begin to take local compensating action; namely, the receiver will perform one of the following algorithms (as mentioned previously for packet loss):

1) Freeze the receiver; that is, allow the receiver to stay in its current state during the silent period.

2) The receiver locally generates a steady state pattern of 1 1 0 0 1 1 0 0 $\cdots$ to be processed by the SVADM decoder during the silent period.

3) The receiver locally generates a steady state pattern of 1 0 1 0 1 0 · · · to be processed by the SVADM decoder during the silent period.

Experimental tests have been conducted to evaluate the above algorithms and the results will be presented in the following section.

7.1.5 Experimental Results for Packet Voice Channel with Silence Detection and Speech Initiation:

Figure 7.1.5 shows the test up. It consists of a speech source, a band pass filter (B.P.F.), a DM encoder, a packetizer, a silence detector, a depacketizer, a steady state generator, a DM decoder, a B.P.F. and monitoring devices. The packetizer-silence detector and the depacketizer-steady state generator were simulated by a PDP-11/34 computer for real time operation.

The parameters varied in the experiments were:

1) Packet size P, where P = 1024, 512 bits
2) Threshold T_p, where T_p = 1/2, 1/4, 1/8 1/16 and
3) Bit rate f_s, where f_s = 16, 9.6 Kb/s.

Experiment 1: Non-Repacking

We found by subjective comparison that there is a very little difference in the use of the three receiver compensation algorithms. In general, a local generation of a 1 0 1 0 1 0 · · · at the receiver during silent periods was preferred for the reasons mentioned earlier. In addition, the subjective evaluation showed that the listeners of the processed speech found no recognizable degradation at T = 1/2 and 1/4. However at $T_p = 1/8$, they were able to distinguish the breaks in the speech. This was due to the fact that at lower thresholds more silent packets are not transmitted. Also, at $T_p = 1/16$, the processed speech loss intelligibility.

We computed the effective packet rate of transmission (r) by keeping track of the total number of packets (N_p) and the total number of silent packets (S) over a fixed period of time. The total time taken to transmit the packets is given by

$$T = (N_p)\ (P)\ /\ f_s$$

where f is the bit rate

The effective packet transmission rate is given by

$$r_e = (N_p - S_p)\ /\ T$$

Figure 7.1.6 shows the plot of r_e as a function of T_p for f_s = 16 Kb/s. At $T_p = 1/2$, $r_e \approx 13$ packets/sec. and at $T_p = 1/8$, $r_e \approx 12$ packets/sec. For all three values of T_p, the processed speech is intelligible. By detecting silence, the effective packet rate, r_e, is reduced. For example at f_s = 16 Kb/s and P = 1024 bits, r_e is approximately 16 packets/sec., when all the packets are transmitted. However, by detecting the silent periods, we obtain a reduction in the value of r_e. Thus at $T_p = 1/8$, $r_e \approx 12$ packets/sec. constitutes a reduction of 25% while maintaining intelligible speech. This is a substantial reduction since the speech tape used had silent periods of approximately 25%, which was measured experimentally.

Experiment 2: Repacking and generation of a local 11001100... pattern at the receiver, when a silent period is detected

The use of repacking and the introduction of a 1 1 0 0 1 1 0 0 · · · pattern at the receiver, during silence, improved the subjective quality of the processed speech at $T_p = 1/8$. The noticeable breaks in speech heard in experiment 1, were not heard.

Here also, we computed the effective packet rate of transmission bh processing the speech over a fixed period of time. In this experiment, we measured the total time (t) of speech processing. r_e is then given by

$$r_e = (N_p - S_p)/t. \qquad (7.1.3.)$$

Tables 7.1.1 and 7.1.2 illustrate the computation of r_e for different values of T_p in the non-repacking and repacking schemes respectively. Figure 7.5 shows the plot of r_e as a function of T_p. We notice that the values of r_e in the repacking scheme are similar to the values in the non-repacking scheme. Thus, r_e is still reduced compared to transmitting all packets.

The periodic pattern of 1 1 0 0 1 1 0 0 ... at the input to the decoder produces a periodic estimate whose fundamental frequency is equal to a fourth of the bit rate. When f_s<16 Kb/s, this frequency is less than 4 KHz. This unwanted component can be heard at the output. In the next experiment, we overcome this problem by feeding a 1 0 1 0 1 0 · · · instead of a 1 1 0 0 1 1 0 0 . . . to the SVADM decoder.

Experiment 3: Repacking and generation of a local 1 0 1 0 1 0 · · · pattern at the receiver when silence is detected.

The use of a 1 0 1 0 1 0 · · · pattern at the decoder input generates a tone at $f_s/2$ and is not heard. The subjective evaluation showed this scheme performed with approximately the same quality as that of experiment 2, with respect to speech intelligibility.

In Table 7.1.3. we have tabulated a subjective comparison of the non-repacking and the repacking schemes. The results of the experiments 2 and 3 are combined. The two criteria, we use, for subjective comparison are (a) intelligibility and (b) acceptability. At $T_p = 1/2$ and $1/4$, there is no difference in the performances of the repacking and the non-repacking schemes. At $T_p= 1/8$, the repacking scheme is significantly better than the non-repacking scheme. However, at $T_p=1/16$ neither system is intelligible.

7.1.6 Summary

Silence detection was accomplished digitally by using the periodic output of the encoder. We showed that the overall packet transmission rate was reduced by not transmitting silent packets. However, we found that for input levels of -20dB and below, the speech processed was unintelligible. Thus, the dynamic range of this system was only 10dB at f_s=16 Kb/s. However, the delta modulator back to back scheme has a 30dB dynamic range (Chapter 4) at the same bit rate. The loss of dynamic range is due to the fact that the silence is detected too soon at lower input levels and thus we lose the end of the speech. Also, the discarded silent packets consist of some speech bits which were lost when not transmitted. The dynamic range of a packet voice system was improved by using a variable packet size algorithm which is explained in the following section.

7.2 Variable Packet Size Scheme

In this section we present a different concept of packet voice transmission where the packet size (or length) is varied between a minimum and a maximum. We varied the silence detection algorithm slightly. We have used the SVADM as the source encoder. Figure 7.2 ,1 shows the flow diagram of the variable packet size scheme. In this scheme, all packets transmitted have sizes between a minimum and a maximum. The packetization of e(k) stops once a silent period is detected. The algorithms involved in this scheme are described in the following sections.

7.2.1. Silence Detection Algorithm

To determine the start of a silent period, we observe sixteen consecutive bits of the encoder output to see if they have a 1100110011001100 pattern (or any of the three other possible permutations as illustrated in Fig.7.2.2). If this pattern is detected, a decision that a silent period has begun is made. Also, the process of packetization of e(k)s is stopped during the silent period. The reason for choosing sixteen bits instead of eight as described in section 7.1 is to prevent the formation of a large number of smaller packets and thus preventing a higher packet transmission rate. To show this effect, we have plotted the distribution of packet sizes using eight as well as sixteen bits to detect silence. We conducted an experiment to keep track of the number of packets of different sizes and made for a speech period of 10 minutes and the experiment was repeated several times. The silence detection algorithm and process of packetization were simulated by PDP 11/34 computer for real time operation at 16 Kb/s. From the results of this experiment, we have plotted the distribution of packet sizes in Figs.7.2.3 and 7.2.4. It is evident from Fig. 7.2.3 and Fig. 7.2.4 that using sixteen bits to detect silence prevents fhe formation of a large number of smaller packets. Also, the use of sixteen bits was preferable to eight bits at lower input signal levels to prevent the detection of silence too soon.

Having entered a silent period, we look for the onset of speech. To detect the onset of speech we observe three consecutive bits of the encoder output to see if they have a 111 or a 000 pattern. Upon entering the active speech period, we now begin the packetization of e(k). Figure 7.2.5 shows the timeing diagram for detection of the onset of speech and the onset of silence.

7.2.2. Process of Packetization

As long as the speech is active, the process of packetization continues. As we described earlier, the packet size is varied between a minimum and a maximum value. Thus, when the i^{th} packet, assembled is a maximum packet, the formation of the $(i+1)^{st}$ packet begins. If a silent period is detected, the packetization stops and the $(i+1)^{st}$ packet may be of a different size. If the $(i+1)^{st}$ is of a size less than the minimum size, then we define it to be not worthy of transmission and the $(i+1)^{st}$ packet is not transmitted. Figure 7.2.6 shows the process of packetization.

In Fig. 7.2.6, p_1, p_3, p_4 are maximum size packets.

p_2 is of size less than the maximum, but greater than the minimum and p_5 is of size less than the minimum. Therefore, only packets p_1, p_2, p_3 and p_4 are transmitted.

For subjective evaluation of the packet voice system, we have varied the minimum and the maximum sizes of the packet.

When silent periods are not transmitted, a gap is created in the stream of packets received at the receiver. We have developed a receiver compensation algorithm (described in Section 7.1) which maintains a proper receiver output during silent periods. The receiver generates locally a 1010... steady state pattern at the input to the SVADM decoder during silent periods.

7.2.3. Experimental Results

Figure 7.2.7 shows the test set up. It consists of a speech source, a bad pass filter (B.P.F.) a DM encoder, a packetizer, a silence detector, a depacketizer, a steady state generator, a DM decoder, a B.P.F. and monitoring devices. The packetizer silence detector and the depacketizer steady state generator were simulated by PDP-11/34 computer for real time operation. Two types of speech tapes were used:

1. A Mark Twain story,
2. A general radio conversation.

The parameters varied in the experiments were:

a. Minimum packet size P_{min}, where P_{min} = 64, 128, 256 bits
b. Maximum packet size p_{max}, where P_{max} = 1024, 2048 bits

c. Bit rate f_s, where f_s = 9.6, 16 Kb/s

d. Input level I, where I = 0, -10, -20, -30 dB

We experimentally evaluated the distribution of packet sizes by counting the number of packets of different sizes formed. The experiment was conducted for about 10 minutes and repeated several times for different speech tapes. For each speech tape, we plotted the distribution in Fig. 7.2.4. We notice that the probability of the maximum packet size is very high. We also computed from the distribution, the packet transmission rate r, the average packet size P_{av} and the effective bit rate f_e. Table 7.2.1. shows a typical computation. In 10 minutes, we have classified a total of 5187 packets. From Table 7.2.1, we notice that f_e = 11.8 Kb/s when the system was operated at f_s = 16 Kb/s and P_{min} = 256 bits, which constitutes a reduction of about 25% of the original bit rate. This seems to agree with the fact that the speech tape which was used for the experiment has silent periods of about 25% of the total speech time. We also monitored f_e on a counter by generating a clock pulse only during the active speech period and not during silence. We found that the monitored value of f_e agreed with that of our computation.

In order to compare the variable packet voice system subjectively, we used the SVADM encoder-decoder system without packetization and silence detection. We refer to this system (without the PDP-11/34 shown in Fig.7.2.7) as the "back-to-back system". We found subjectively that there is no difference between the performance of the packet voice system and the back-to-back system at f_s = 16 Kb/s and I = 0 dB, -10 dB and -20dB. This is also true at f_s = 9.6 Kb/s and I = 0 dB and -10dB. Thus, the packet voice system offers a 20 dB dynamic range of f_s = 16 Kb/s. However, the back-to-back system offers a dynamic range of -30 dB at f_s = 16 Kb/s (chapter 4). For subjective performance, we have used P_{min} = 256 bits. We found that there is no difference in the subjective quality of speech reproduced when P_{min} is varied from 64 bits to 256 bits. However, the effective bit rate changes as a function of P_{min}. Table 7.2.2 shows f_e as a function of P_{min}, P_{max} for different speech tapes.

7.2.4 Summary

The variable packet size packet voice system has been developed where the effective bit rate is reduced as a function of the total silence available in the speech source. Also, it offers a 10 dB higher dynamic range over the fixed packet size voice system.

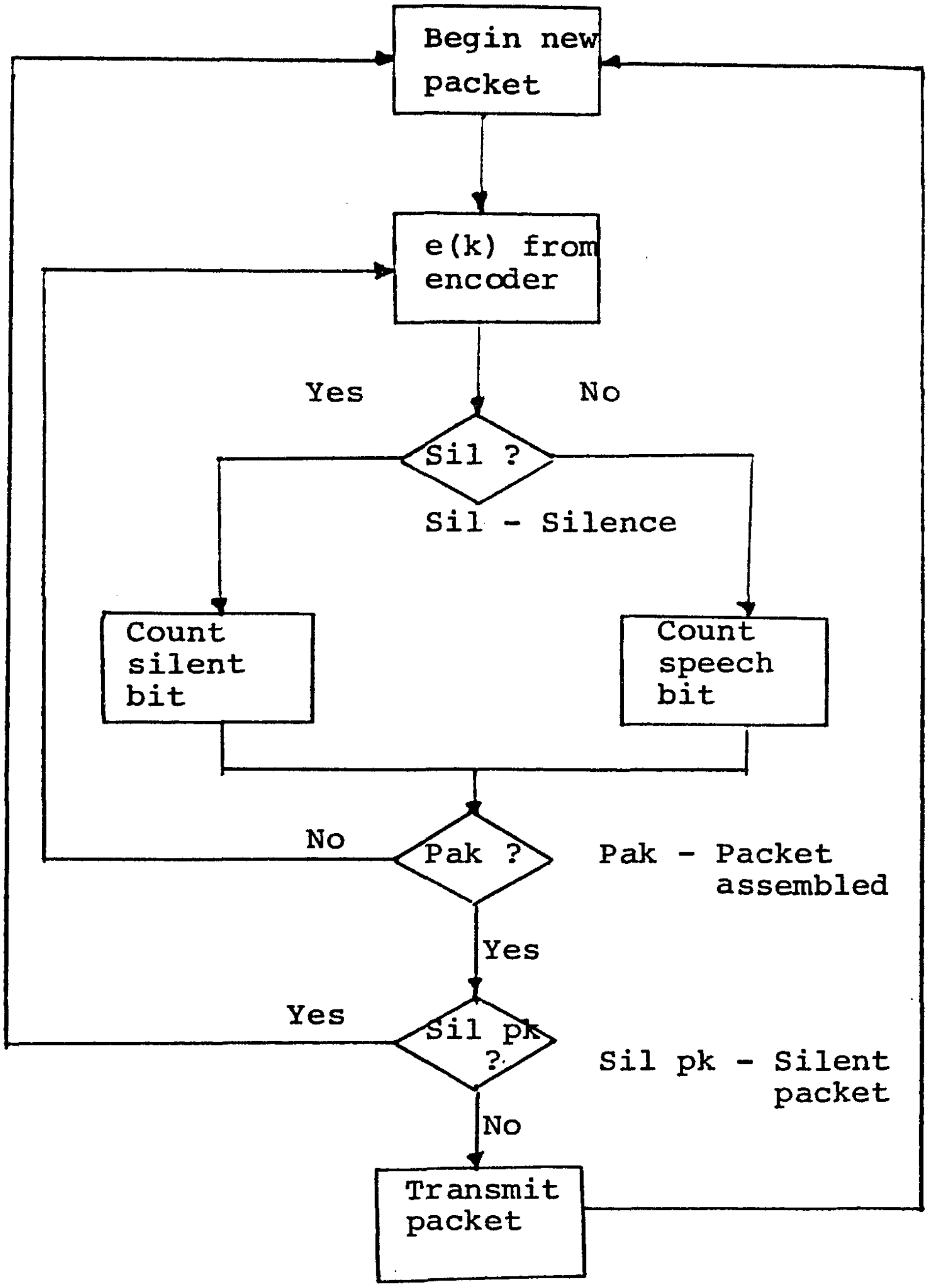


Fig. 7.1.1 Flow diagram for the fixed packet size scheme

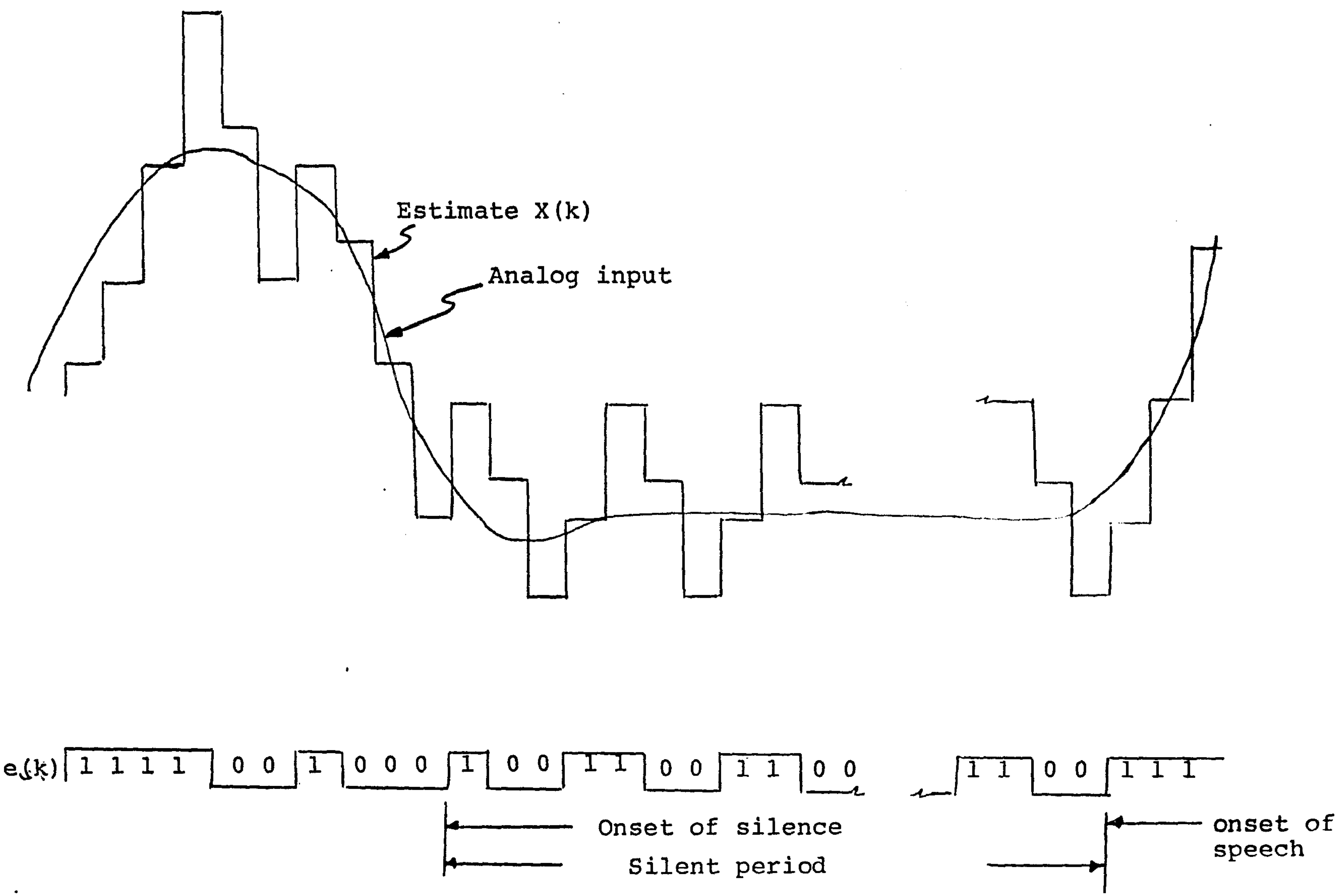
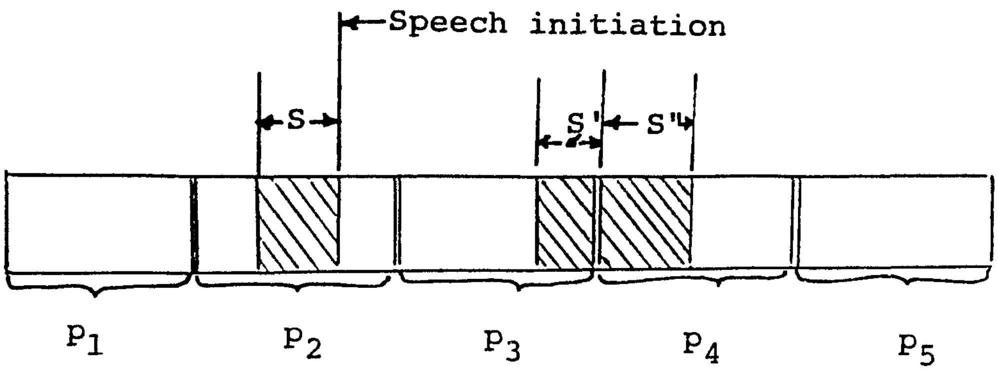


Fig. 7.1.2 Timing diagram for the onset of speech and the onset of silence



$S/P \geq T_p$ $S'/P < T_p$ $S''/P \geq T_p$

P - Packet size of each packet

Fig. 7.1.3 Determination of a silent packet

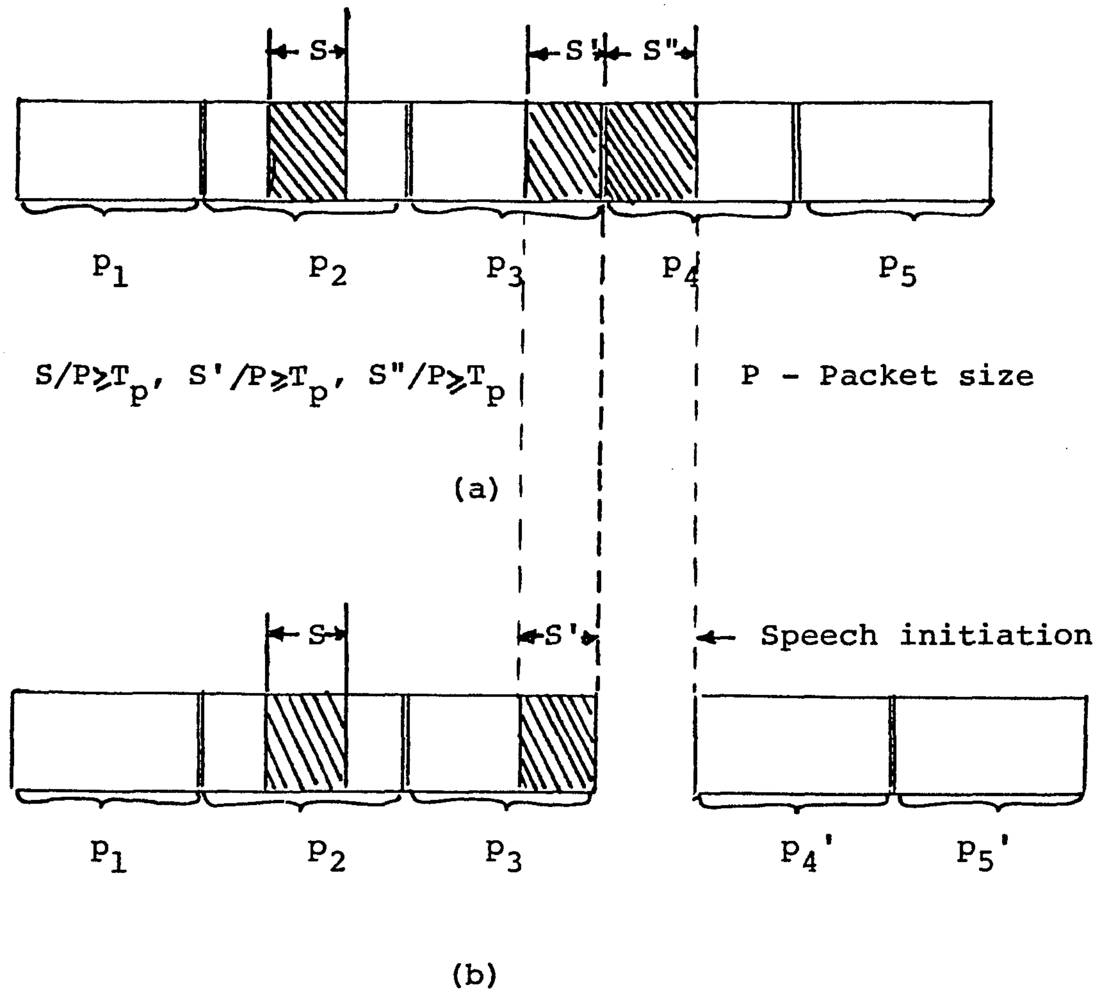
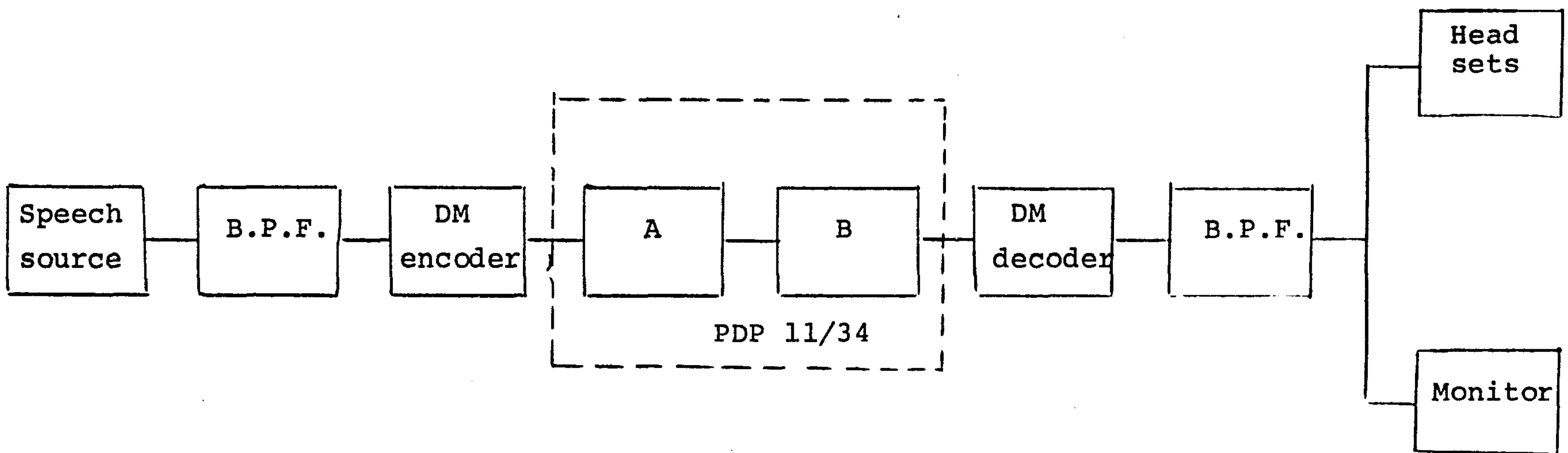


Fig. 7.1.4 (a) Non-Repacking
(b) Repacking



A - Packetizer and silence detector

B - Depacketizer and steady state generator

B.P.F. - Band pass filter set from 300 Hz to 2500 Hz

Fig. 7.1.5 Test set up for silence detection and speech onset detection

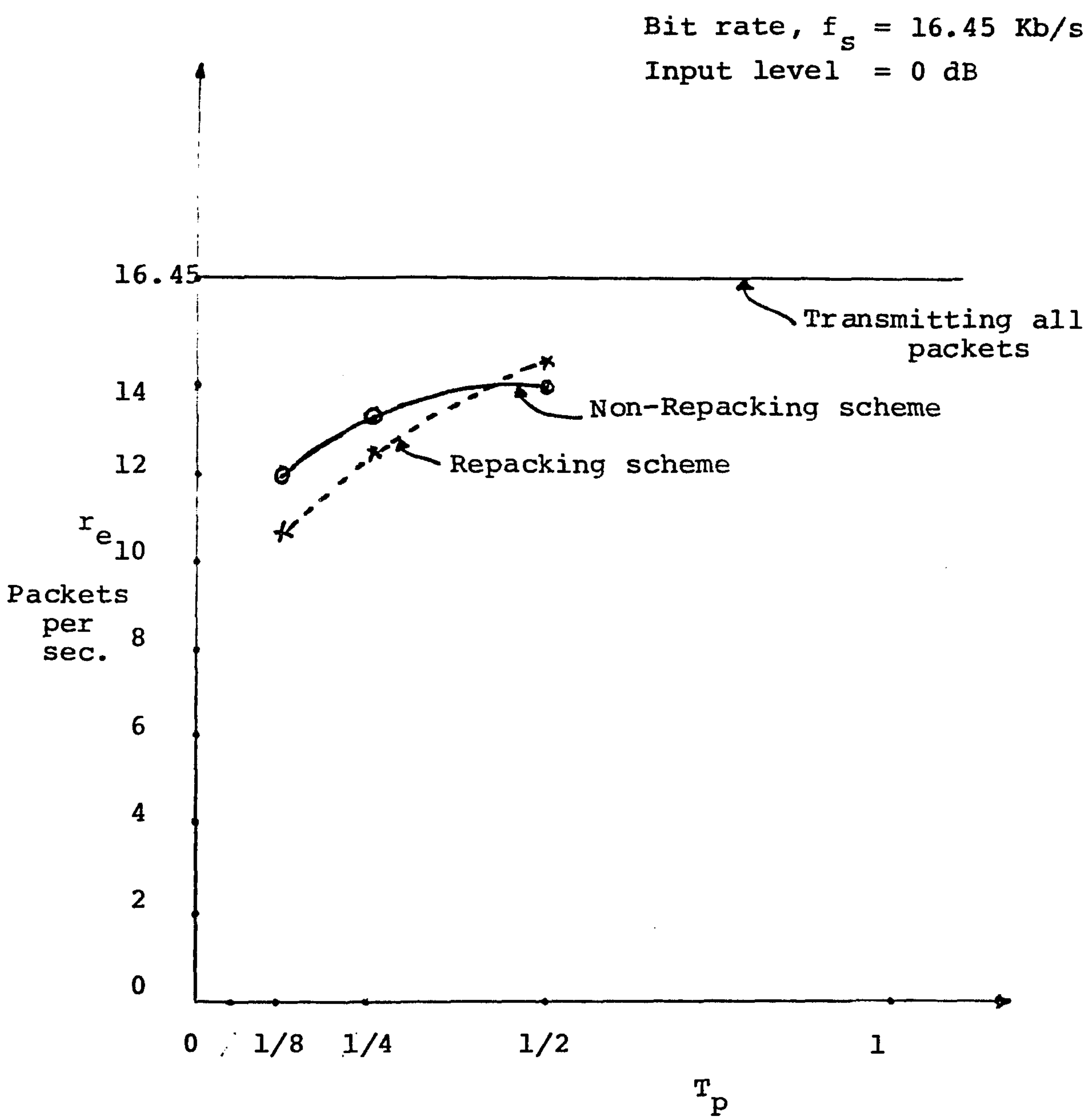


Fig. 7.1.6 Effective packet transmission rate, r_e, as a function of T_p

Table 7.1.1 Computation of the effective packet rate of transmission for "Non-Repacking" scheme.

Input level = 0 dB
Packet size = 1024 bits
Bit rate f_s = 16/452 Kb/s

	Threshold T_p		
	1/2	1/4	1/8
Total number of Packets formed, N_p	19911	16871	14747
Total number of Silent packets, S_p	2654	2762	3703
Total number of packets transmitted $(N_p - S_p)$	17255	14109	11044
Total time taken to transmit the packets, (N_p) $(P)/f_s$ secs.	1239	1050	917.87
Effective packet rate of transmission, r_e	13.9	13.4	12

Table 7.1.2 Computation of the effective packet rate of transmission for"Repacking" scheme.

Input level = 0 dB

Packet sixe = 1024 bits

Bit rate f_s = 16/452 Kb/s

Total time of speech processing = 600 sec.

	Threshold T_p		
	1/2	1/4	1/8
Total number of packets formed, N_p	9492	9561	9515
Total number of silent packets, S_p	1002	2071	3150
Total number of packets transmitted $N_p - S_p$	8490	7490	6365
Effective packet rate of transmission r_e	14.15	12.4	10.6

Table 7.1.3 Subjective comparison of "Non-Repacking" and "Repacking" schemes

I - Intelligible; A - Acceptable; NI - Not intelligible; NA - Not acceptable
SC -Speech is clipped; BN - Breaks are noticed

Input level dB	f_s Kb/s	P bits	T_p	Non-Repacking	Repacking	Comparison
0	16	1024	1/2	I,A	I,A	No difference
			1/4	I,A	I,A	No difference
			1/8	I,NA,SC	I,A	Repacking is significantly better
			1/16	NI,NA	NI,NA	Words are missing
-10	16	1024	1/2	I,A	I,A	No difference
			1/4	I,A	I,A	No difference
			1/8	NI,NA	I,A,BN	Repacking is preferred
0	9.6	1024	1/2	I,A	I,A	No difference
			1/4	NI,NA	NI,NA	Words are clipped

Table 7.1.3 Continued

Input level dB	f_s Kb/s	P bits	T_p	Non-Repacking	Repacking	Comparison
-10	9.6	1024	1/2	I,A,BN	I,A,BN Breaks are less than Non-Repacking	Repacking is preferred
			1/4	NI,NA	NI,NA	

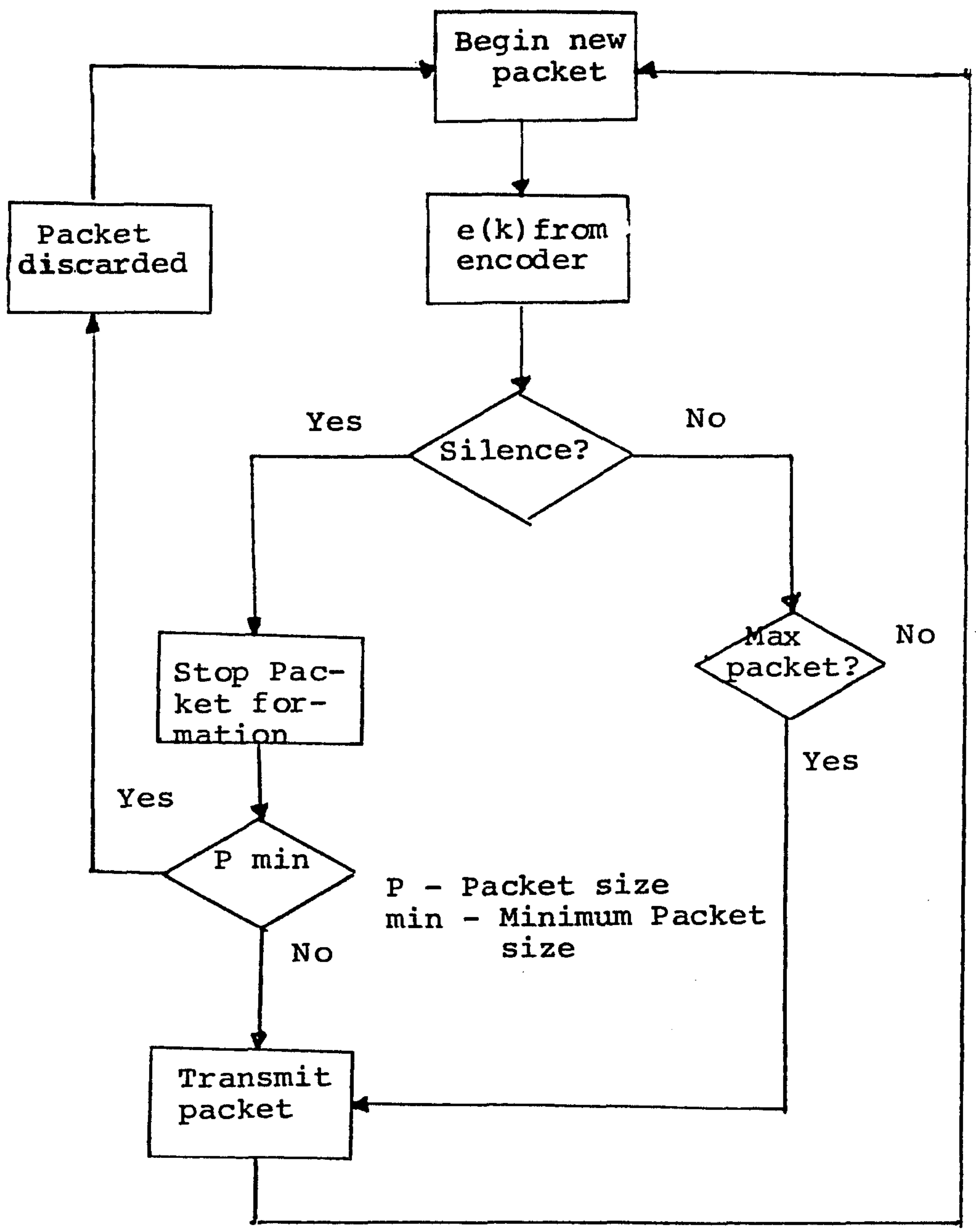


Fig. 7.2.1 Flow diagram of the variable packet size scheme

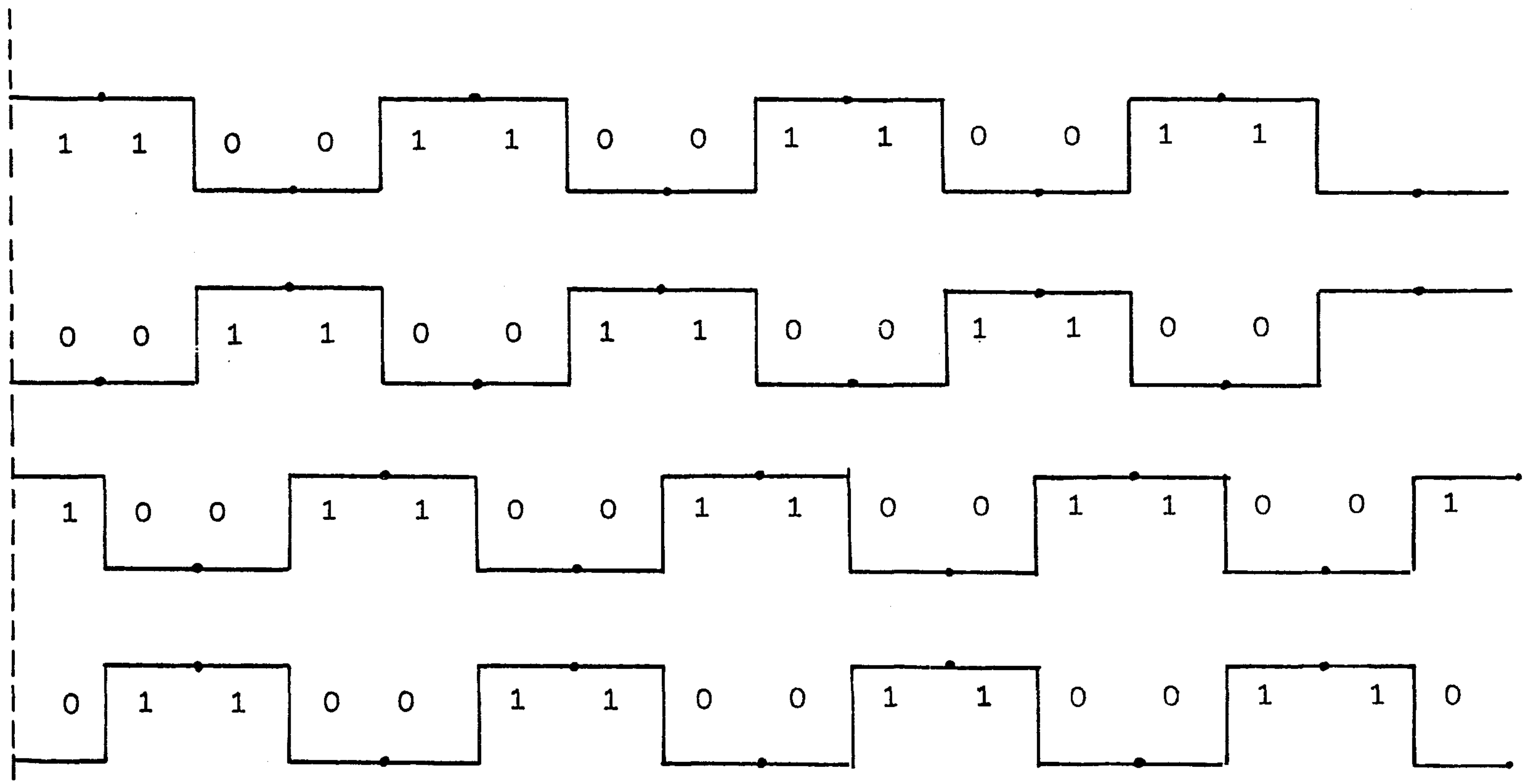


Fig. 7.2.2 Steady state patterns for the SVADM

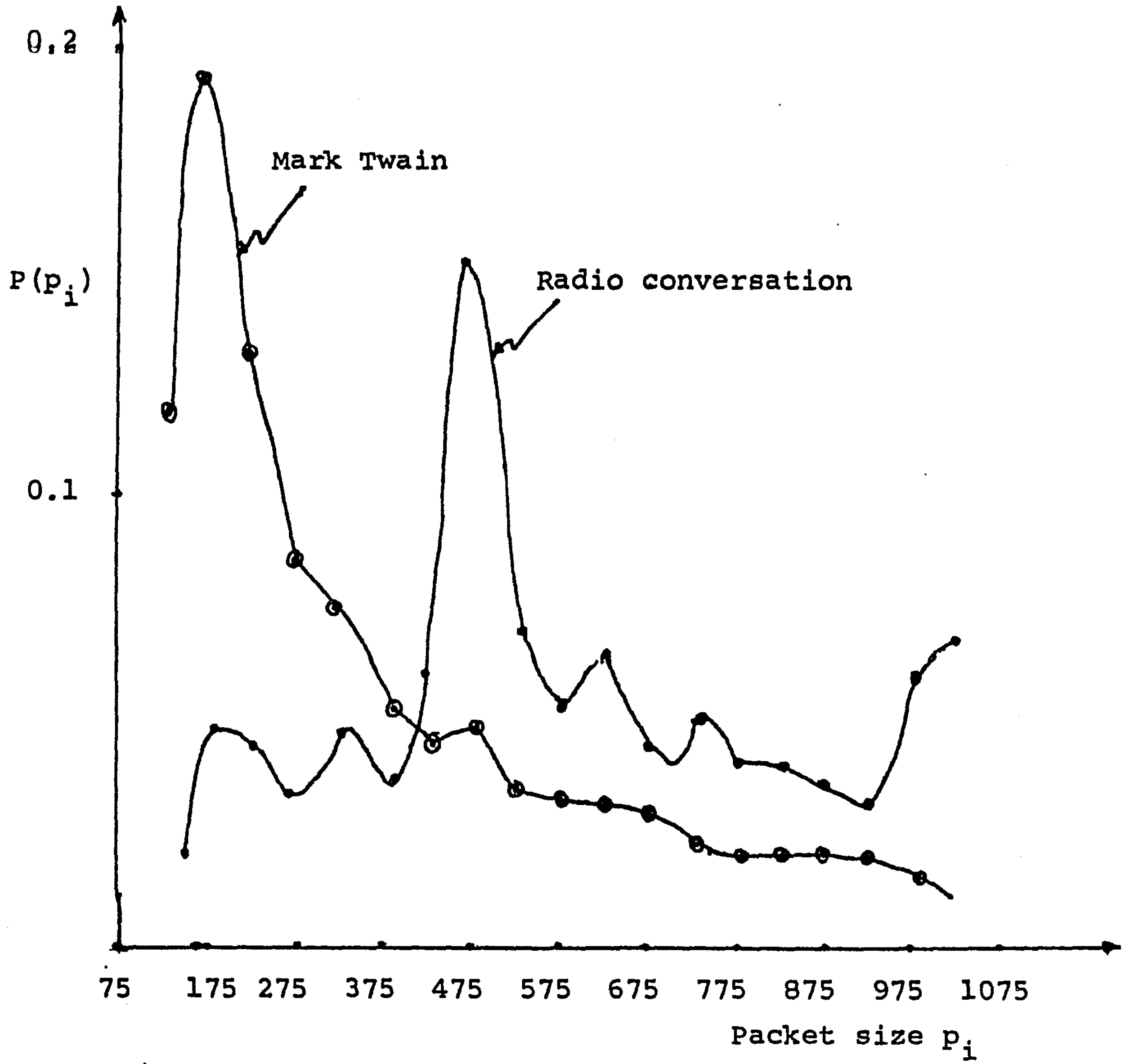


Fig. 7.2.3 Probability distribution using eight bits for silence detection

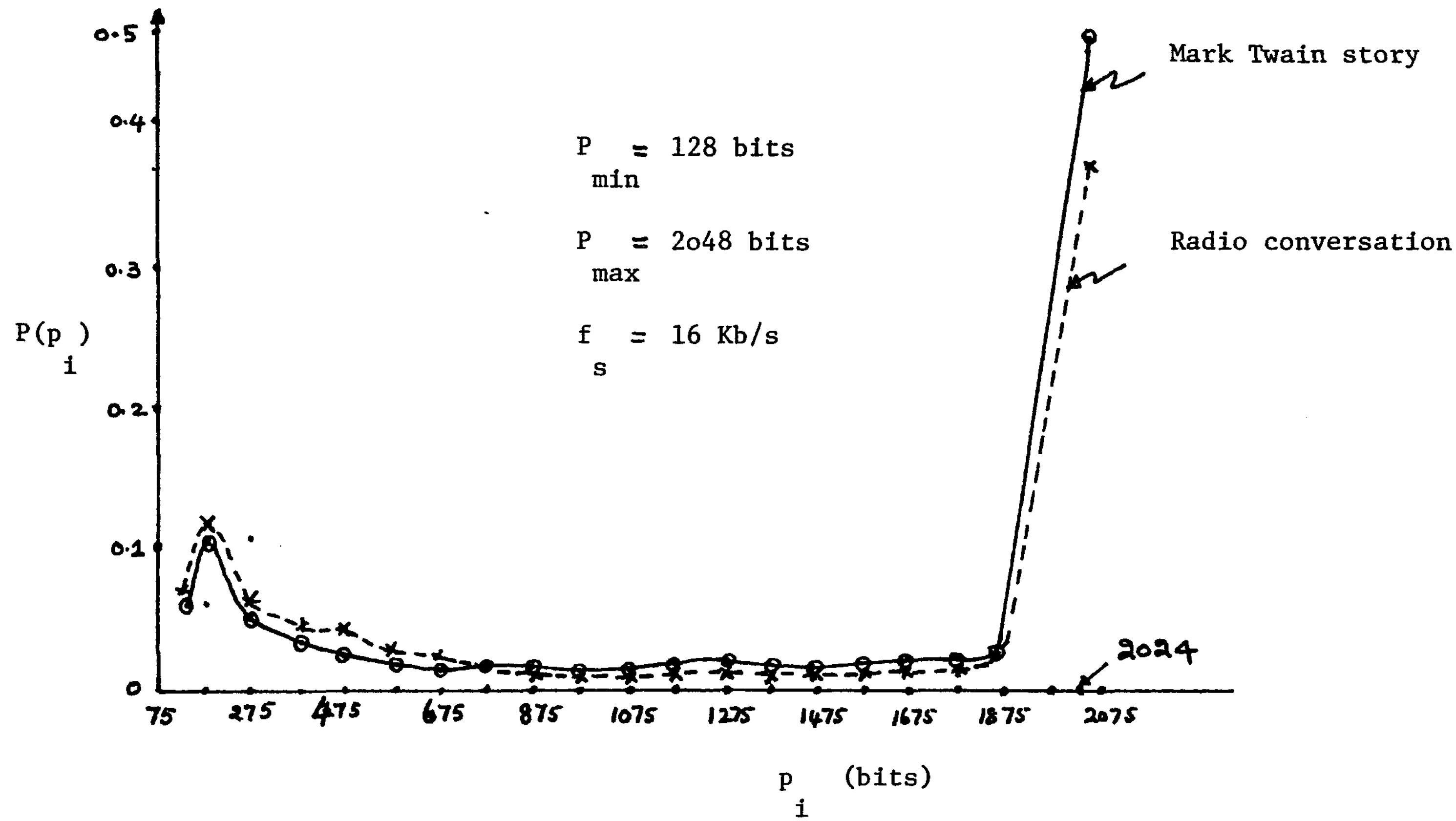


Fig. 7.2.4 Probability distribution using sixteen bits for silence detection

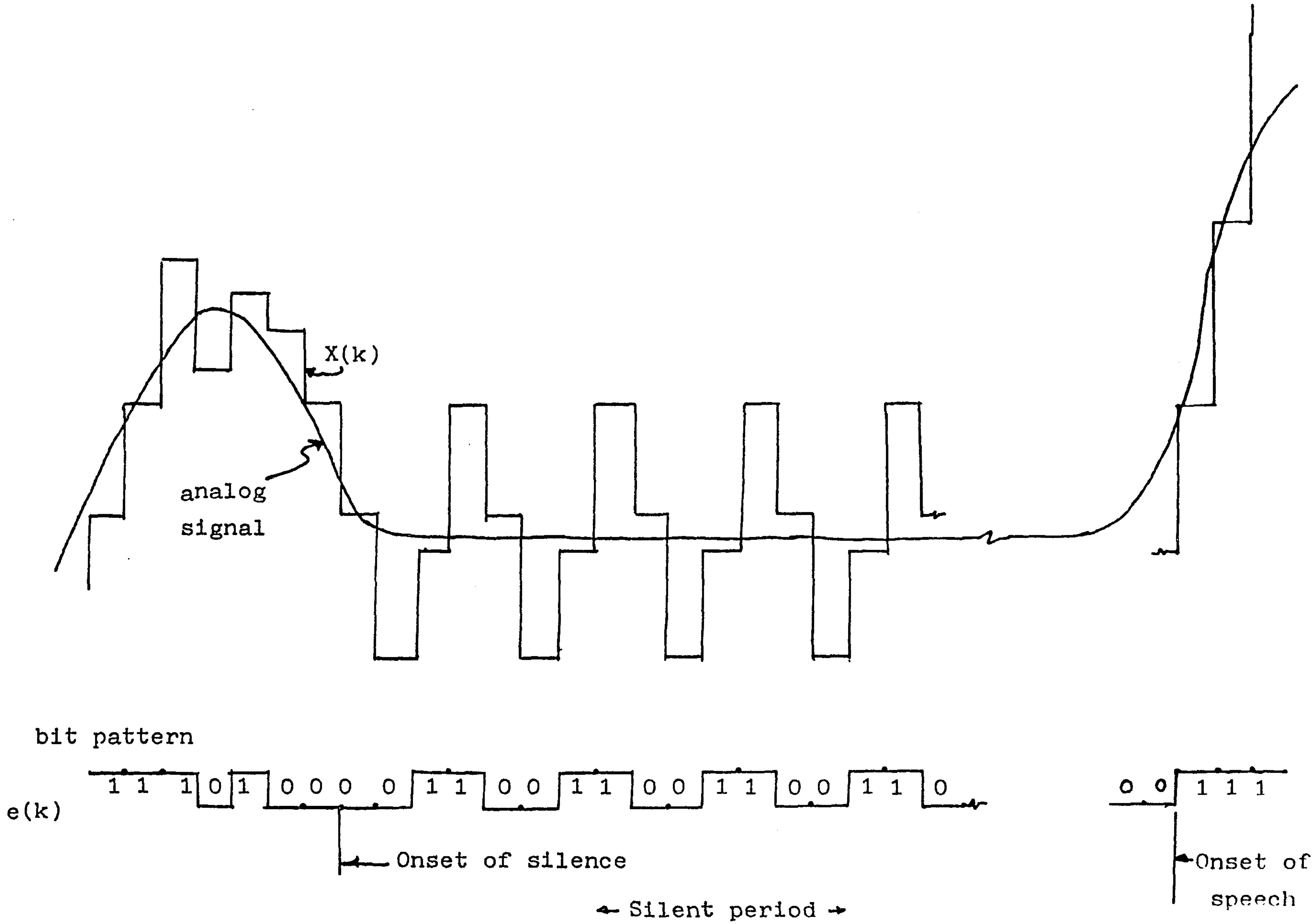


Fig. 7.2.5 Timing diagram to show the onsets of silence and speech periods

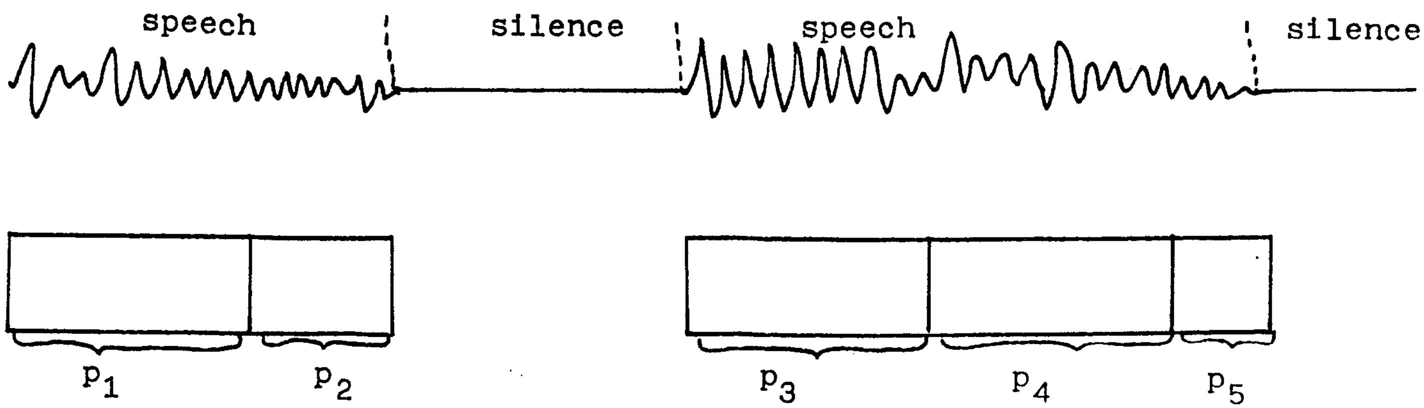
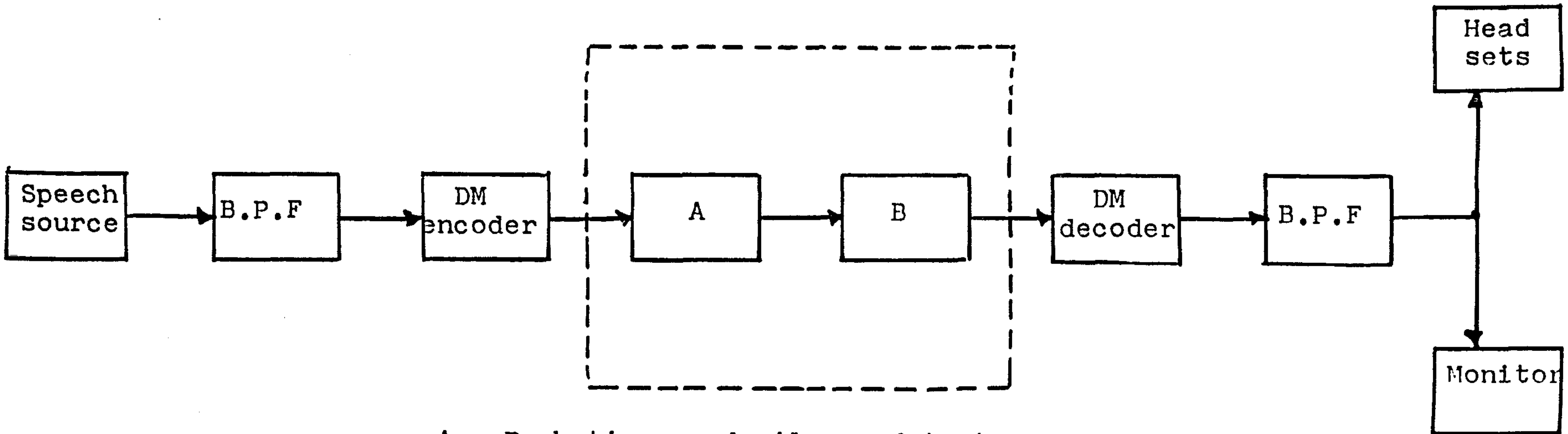


Fig. 7.2.6 Packetization process



A - Packetizer and silence detector

B - Depacketizer and steady state generator

B.P.F - Band pass filter set from 300 Hz to 2500 Hz

Fig. 7.2.7 Test set up for silence detection and speech onset detection

Table 7.2.1 Typical computation of r, P_{av}, f_e from the distribution of packet sizes at f_s = 16 Kb/s

Packet range PR bits	PR bits	# Packets	P_{av} bits	r Paks. per sec.	f_e Kb/s
250-299	275	361	1476	8	11.8
300-349	325	244			
350-399	375	190			
400-449	425	135			
450-499	475	111			
500-549	525	110			
550-599	575	91			
600-649	625	71			
650-599	675	60			
700-749	725	59			
750-799	775	64			
800-849	825	36			
850-899	875	37			
900-949	925	32			
950-999	975	32			
1000-1049	1025	38			
1050-1099	1075	36			
1100-1149	1125	32			
1150-1199	1175	22			
1200-1249	1225	27			
1250-1299	1275	30			
1300-1349	1325	18			
1350-1399	1375	24			
1400-1449	1425	21			
1450-1490	1475	15			
1500-1549	1525	18			
1550-1599	1575	22			
1600-1649	1625	17			
1650-1699	1675	21			

Table 7.2.1 continued

Packet range PR bits	PR bits	# Packets	P_{av} bits	r	f_e Kb/s
1700-1749	1725	20			
1750-1799	1775	19			
1800-1849	1825	18			
1850-1890	1875	17			
1900-1949	1925	22			
1950-1999	1975	16			
2000-2048	2024	3111			

Mathematics of computation

$P_{av} = \sum [p_i/P][\overline{PR}]$

$P = \sum p_i$

$r = P/t$

$f_e = r \cdot P_{av}$

t = total time of forming all the packets

Table 7.2.2 r, P_{av}, f_e as a function of P_{min}, P_{max}, at f_s = 16 Kb/s

	P_{max} bits	P_{min} bits	r Pak./sec.	P_{av} bits	f_e Kb/s
Mark Twain story	2048	129	11.2	1270	14
		200	9.7	1340	12.9
		230	8.	1476	11.8
Radio conversation	2048	128	11.7	1114	13
		200	10	1285	12.8
		250	9.1	1396	12

Chapter 8

Hybrid Adaptive Delta Modulator

The Hybrid Adaptive Delta Modulator (HADM) is a new concept, wherein, the bit rate is varied as a function of the input slope. This is in addition to the adaptation of the step size. In previous chapters, we showed that the transmission rate is reduced when silent periods are not transmitted. Packet networks are very convenient to detect silent periods and not transmit them. When the receiver does not receive a silent period, it locally generates a steady state pattern at the input to the decoder. However, if packet networks are not used, there is a problem for the decoder to reinsert the silent periods, when transmitter does not transmit them. HADM is a convenient way of avoiding such a problem. When the input signal has a silence, the HADM operates at a low bit rate and when the input signal is active, it operates at a high bit rate. The digital output is stored in a buffer and then transmitted at an average rate. Theoretically, if we have an infinite buffer, the transmission can be achieved at a lower bit. The information of bit rate switching is contained in the past digital output and as such there is no need to transmit any additional information. There are two different algorithms studied. The first is the modification of the SVADM and we called it the Hybrid SVADM (HSVADM). The second algorithm has a new step-size algorithm.

8.1 Hybrid Song Voice Adaptive Delta Modulator

The HSVADM has the same step size algorithm as in SVADM and the bit rate changes depending on the two past digital outputs.

.. Algorithm

The HSVADM is described by

$$X(k+1) = X(k) + S(k+1) \tag{8.1.1}$$

$$S(k+1) = |S(k)|e(k) + S_o e(k-1) \tag{8.1.2}$$

$$e(k) = \mathrm{Sgn}\,[M(k) - X(k)] \tag{8.1.3}$$

$$f_s = f_H\, F\,[e(k-1),\ e(k-z)] \tag{8.1.4}$$

where, at kth interval,

e(k), X(k), S(k), So, M(k) are defined as before (Refer Chapter 3),

f_s is the DM bit rate,

f_H is the high bit rate and

F is a variable bit rate function

The F function is defined by

$$F(0,0) = F(1,1) = 1 \tag{8.1.5}$$

$$F(0,1) = F(1,0) = \beta$$

where β (i.e. low bit rate fraction) is less than one.

From Eq.(8.1.5) we observe that when the sign of e(k) is changing, the delta modulator is operating at a lower bit rate. This signifies that the input signal is not rising fast. On the other hand when the sign of e(k) is not changing, the delta modulator is operating at a higher bit rate. This signifies that the input signal is rising fast. Figure 8.1.1 shows the block diagram of the HSVADM. The feedback circuit of the encoder is essentially the decoder. Since the transmission of the data is accomplished at an average bit rate, the output of the encoder is buffered before transmission. The design of the buffer will be dealt in later section. For the moment, we consider that we have infinite buffer length so that we always can transmit the data at an average rate.

Figure 8.1.2 shows the step response of the HSVADM we note that the steady state output of the HSVADM also has a periodic 11001100...(similar to SVADM). However, these outputs follow alternately one high and one low frequency bit. Thus the bit rate for transmission (f_{dm}) is the average of the two bit rates. This is not the most efficient scheme, since the steady state condition occurs during the silent periods of speech and we would expect to have only low frequency bits during silent periods. Thus we developed a new algorithm which has a variable step size algorithm different from that of the HSVADM. The new algorithm is referred to as the "Hybrid Adaptive Delta Modulator (HADM), which will be described in later sections.

8.2 Subjective Comparison of HSVADM and SVADM

The easy way to compare these two systems is by subjective listening. We used the microcomputer designed and developed in the communication laboratory [1] to program both the HSVADM and the SVADM for real time operation. Figure 8.2.1 shows the test set up used for subjective comparison. During the test we monitored the average bit rate (f_{av}) of transmission of the HSVADM. We compared the performance of the HSVADM with that of the SVADM operating at the same bit rate (i.e. f_{av} of HSVADM = f_s of SVADM). The test was performed by varying the following parameters:

High bit rate of HSVADM, f_H where f_H = 32, 24, 16 Kb/s
Low bit rate fraction, β where β = 1/2, 1/3, 1/4

We found from the test, that there is no difference in the performances of the HSVADM and the SVADM under all conditions of operation (input level was set at full load). This is evident from the fact that during silent periods, the HSVADM operates alternately at bit rates f_H and βf_H ($f_L = \beta f_H$) for every e(k) generated. Therefore by slightly modifying the step size algorithm of the SVADM,we developed the new HADM algorithm.

8.3 Hybrid Adaptive Delta Modulator (HADM)

The HADM algorithm allows the steady state output to reach a periodic 101010··· pattern for a constant input.

8.3.1 Algorithm

The HADM algorithm is described by

$$e(k) = \mathrm{Sgn}[M(k) - X(k)] \tag{8.3.1}$$
$$X(k+1) = X(k) + S(k+1) \tag{8.3.2}$$

Let us define a pattern $R = e(k-2)\ e(k-1)\ e(k)$. Then

$$S(k+1) = \begin{cases} \left[|S(k)| + 2\,So\right] e(k) & \text{for } R_1 = \begin{Bmatrix} 000 \\ 111 \end{Bmatrix} \\ |S(k)|\, e(k) & \text{for } R_2 = \begin{Bmatrix} 011 \\ 100 \end{Bmatrix} \\ \left[|S(k)| - So\right] e(k) & \text{for } R_3 = \begin{Bmatrix} 001 \\ 110 \\ 010 \\ 101 \end{Bmatrix} \end{cases} \tag{8.3.3}$$

Also, to prevent a zero step size,

$$s(k+1) = So\ e(k) \qquad \text{for } |S(k+1)| \leq S_o \tag{8.3.4}$$

The frequency adaptation is in general given by

$$fs = \begin{cases} f_H & \text{for } R_1 \\ \alpha f_H & \text{for } R_2 \\ \beta f_H & \text{for } R_3 \end{cases} \tag{8.3.5}$$

where,

f_H is the highest frequency,

α, β are coefficients less than unity to allow lower bit rate.

Before, we used this algorithm in variable bit rate mode, we compared the new step size algorithm with SVADM. We call this new step size algorithm as new step size Adaptive Delta Modulators (NSSADM). The NSSADM is a fixed bit rate algorithm. That means, $\alpha = \beta = 1$. For comparing NSSADM with SVADM, we use test shown in Fig. 8.3.1. Figure 8.3.2 shows the SNR vs bit rate for an input sinusoid of 800Hz and 0 dB(full load). For the test, the band pass filter us set from 300 Hz to 2500 Hz. From Fig.8.3.2 , we infer that both DMs perform about the same. When speech was processed by both Dms and subjectively tested, the performance of the NSSADM was found to be about the same as that of the SVADM with respect to the dynamic range, the bit error rate and the bit rate. Therefore, all the performance curves shown in Chapters 3 and 4 for the SVADM are also valid for NSSADM, with one exception. The steady state output of the SVADM has a 11001100... pattern, whereas, it is 101010... in the case of NSSADM. Thus, when speech bandlimited from 300 Hz to 2500 Hz is processed at a bit rate $f_s < 16$ K b/s, the oscillation in the steady state is not heard in the case of NSSADM. It is essentially the 101010... pattern of NSSADM allows the HADM to produce the low bit rate output during silent periods.

8.3.2. Comparison of HADM and SVADM with sinusoidal input

Eventhough, we emphasize that the HADM operates at low bit rate in silent periods and operates at high bit rate during speech, it is also true that HADM produces some low frequency bits even when the input speech is active to offset the excessive increase of the step size. In order to visualize this, we compared the HADM with the SVADM for sinusoidal inputs. For comparison, the bit rate of the SVADM is made equal to the average bit rate of the HADM which was monitored during measurements. Figure 8.3.1 shows the test set up used for measurements. Table 8.3.1 shows the SNR as a function of the average bit rate and the coefficients α, β. We notice that the SNR is about the same for HADM and SVADM and vary within $\pm$ 0.5 dB. Figure 8.3.3 shows the SNR as a function of α for different values of β. For all values of α and β less than unity, the SNR was about the same and varied within $\pm$ 0.5 dB. We also plotted the SNR at different input levels using the same sinusoid. Figure 8.3.4 shows the SNR as a function of input level. In this case we chose $\alpha = \beta = \frac{1}{2}$. Figure 8.3.5 shows the estimates of HADM for different input signals.

8.3.3 <u>Comparison of HADM and SVADM with Speech Input</u>

Using the same test set up shown in Figure 8.3.1 we compared subjectively the performance of the HADM with that of the SVADM. Here also, we monitored the average bit rate of the HADM. Since HADM operates at only low bit rate during silent periods, the performance of the HADM is expected to be better than that of the SVADM. We found subjectively that $\alpha = \beta$ $(= \delta)$ is better than of other combinations which means that we use only two bit rates for the HADM. Therefore the bit rates for the HADM are f_H (high bit rate and f $(= \delta f_L$; low bit rate). The frequency adaptation is now given by

$$f_s = f_H \, F \left[e(k-1), \, e(k-2) \right] \tag{8.3.6}$$

where

$$\begin{aligned} F(0,0) &= F(1,1) = 1 \\ F(0,1) &= F(1,0) = \delta \quad (\delta < 1) \end{aligned} \tag{8.3.7}$$

For comparison, the speech was bandlimited from 300 Hz to 2500 Hz. We found that when f_L (low bit rate) is less than $\approxeq$ 8 K/bs, the performance of the SVADM was preferred to that of the HADM. Therefore to show a good performance of the HADM, we restricted f_L to be always larger than or equal to 8 K b/s by suitably adjusting f_H and δ. Table 8.3.2 compares the performances of HADM and SVADM as a function of fav, input level. We found that as long as $f_L \geq 8$ Kb/s, the performance of the HADM was preferred to that of the SVADM at input levels from OdB down to -20dB for f_{av} above 16 K b/s and at input levels of OdB and -10 dB for f_{av} below 16 K b/s. Thus, the HADM has a 10 dB lower dynamic range compared to that of the SVADM.

8.4 Conceptual Buffer

In the HADM, the encoder generates a multi bit rate digital output e(k). Therefore, a suitable buffer is required to stack the e(k)s of different bit rates and transmit these e(k)s at an average rate. The length of the buffer depends on the runlengths of e(k). A finite length buffer might tend to overflow. Also, underflow of a buffer is usually not a major problem.

Figure 8.4.1. shows a conceptual buffer. It consists of a serial input-parallel output register, a ring counter and tri state devices. In the ring counter, the state of the flip-flap A is set initially to '1' and other flip-flaps are cleared. The e(k) of the HADM encoder is serially fed into the buffer. They are shifted to right at each input clock-pulse ($f_{in}=f_s$;variable rate). Everytime, the e(k)s are shifted to right, the '1' in the ring counter shifts to right. Whenever the output clockpulse ($f_{out}= f_{av}$; average clock occurs), the e(k) from the buffer is popped out. Since only one tristate device is enabled at a time, the corresponding e(k) is transmitted. Also, the '1' in the counter is shifted to left when the average clockpulse occurs.

Consider that e(k)s are filled up to F and '1' is at C in the counter. Then the tristate device T_3 is enabled. When font clockpulse occurs, the e(k) at F is transmitted. The occurance of f_{av} pulse enables the '1' in the ring counter to shift left. The next e(k) to be transmitted is at (y-1). Also, because of the fact that '1' is now at B, the tri state device T_2 is enabled. However, if f_{in} pulse occurs and e(k) is pushed into the buffer, the '1' is shifted to C and thus enables T_3. Also, the e(k) to be popped out is moved to y and will be transmitted once an f_{out} pulse occurs. Thus the transmission of e(k)s is a synchronously controlled.

The buffer overflows if more e(k)s pushed in and less e(k)s are popped out. In other words, e(k)s of higher bit rate (f_H) are pushed in and the buffer is transmitting at f_{av} ($f_{av}<f_H$). When the buffer is completely filled, the '1' in the counter is at extreme right position (E in Fig. 8.4.1). If the buffer is substantially larger than the maximum run length of e(k)s of high bit rate, this overflow condition is eliminated.

The buffer underflows if less e(k)s are pushed in and more e(k)s are popped out. This condition can exist if the e(k)s of low bit rate (f_L) are pushed in, while the buffer is transmitting at f_{av}. When the buffer is completely empty, the '1' in the counter is at extreme left position (A in Fig.8.4.1). When this condition occurs, it is advantageous to clear the register, so that, the buffer transmits a string of zeros if f_{av} occurs and this would saturate the decoder.

8.4.1. Computation of Average Bit Rate (f_{av})

In order to arrive at an expression for f_{av}, we will consider a time slot T where, the total number of e(k)s to be transmitted =N the total number of e(k)s belonging to low frequency f_L = N1 the total number of e(k)s belonging to high frequency $f_H = N_2$. Then

$$1/f_{av} = \frac{N1}{N} \left[1/f_L \right] + N2 \left[1/f_H \right] \qquad (8.4.1)$$

$$N1 + N2 = N$$

If $N1 = N2 = \frac{N}{2}$, then

$$1/f_{av} = 1/2 \left[1/f_L + 1/f_H \right] \qquad (8.4.3)$$

In our tests, we have monitored the average bit rate and we found that it agrees with Eq.(8.4.3) for speech. Table 8.4.1 illustrates the values of f_{av} for different values of f_L and f_H.

8.4.2 <u>Buffer Length</u>

The bit rates of the HADM are f_L and f_H and the average transmission rate is f_{av}. Let us consider that:

the ,the maximum sequence of e(k)s belonging to $f_H = L_2$
the maximum sequence of e(k)s belonging to $f_H = L_2$

Underflow condition

The underflow can occur if the HADM is operating at f_L and the transmission is done at f_{av}.

The rate at which e(k)s are entering the buffer = $(1/f_L)$ (8.4.4)

The sequence length== L_1 (8.4.5)

The rate at which e(k)s are leaving the buffer = $(1/f_{av})$ (8.4.6)

The total time of presence of e(k)s = $L_1(1/f_L)$ (8.4.7)

The total time required for L_1 e(k)s to leave the buffer = $L_1(1/f_{av})$ (8.4.8 a)

Let the buffer length be = M (8.4.8 b)

Then, assuming the buffer to be empty initially,

$$(1/f_{av})M = (1/f_L)L_1 - (1/f_{av})L_1 \quad (8.4.9)$$

$$= L_1(1/f_L - 1/f_{av}) \quad (8.4.10)$$

or

$$M = L_1(f_{av}/f_L - 1) \quad (8.4.11)$$

Overflow condition

The overflow condition can occur if the HADM is operating at f_H and the transmission is done at f_{av}.

The rate at which e(k)s are entering the buffer = $1/f_H$ (8.4.12)

The sequence length = L_2 (8.4.13)

The total time of presence of L_2 e(k)s = $L_2(1/f_H)$ (8.4.14)

The rate at which $e(k)$s are leaving the buffer

$$= 1/f_{av} \tag{8.4.15}$$

The time required for L_2 $e(k)$s to leave the buffer

$$= (L2)(1/f_{av}) \tag{8.4.16}$$

Assuming the buffer to be full initially

$$(1/f_{av})\ M = (\frac{1}{f_{av}})\ L_2 - (1/f_H)\ L2 \tag{8.4.17}$$

$$M = L2(1 - f_{av}/f_H) \tag{8.4.18}$$

From Eqs. (8.4.11) and (8.4.18),

$$L1\ (f_{av}/f_L - 1) = L_2\ (1 - f_{av}/f_H) \tag{8.4.19}$$

f_{av} L1 = L2,

$$\left(\frac{f_{av}}{f_L} - 1\right) = \left(1 - \frac{f_{av}}{f_H}\right) \tag{8.4.20}$$

$$f_{av}\ (1/f_L + 1/f_H) = 2 \tag{8.4.21}$$

$$1/f_{av} = \frac{1}{2}\ (1/f_L + 1/f_H) \tag{8.4.22}$$

Equation (8.4.22) is same as Eq.(8.4.3).

If f_H = 30 Kb/s and f_L = 10 Kb/s, f_{av} = 15 Kb/s.

We were able to derive the length of buffer using both underflow and overflow conditions. The buffer at the decoder is similar to the one at transmitting end. However, f_{in} at the receiver is f_{av} and f_{out} is f_s of the HADM decoder.

As a part of this dissertation, we simulated this buffer using PDP 11/34 as a non-real/time processor. We used a square wave as the input to the HADM. We also choose $f_H = 2KHz$, $f_L = 500Hz$ $f_{av} = 1.25$ KHz. This value of f_{av} is slightly higher than that of f_{av} from Eq. (8.4.3). $f_{av} = 1.25$ KHz is chosen for convenience and the simulation is just intended to demonstrate the use of the buffer. The hardware implementation of the buffer for real time voice processing is kept as a possible future work.

Summerizing, we were able to show that HADM indeed improves the performance by sampling silent periods at a low bit rate. Also, the receiver does not need additional information for change of bit rates since it is incorporated in the previous bits.

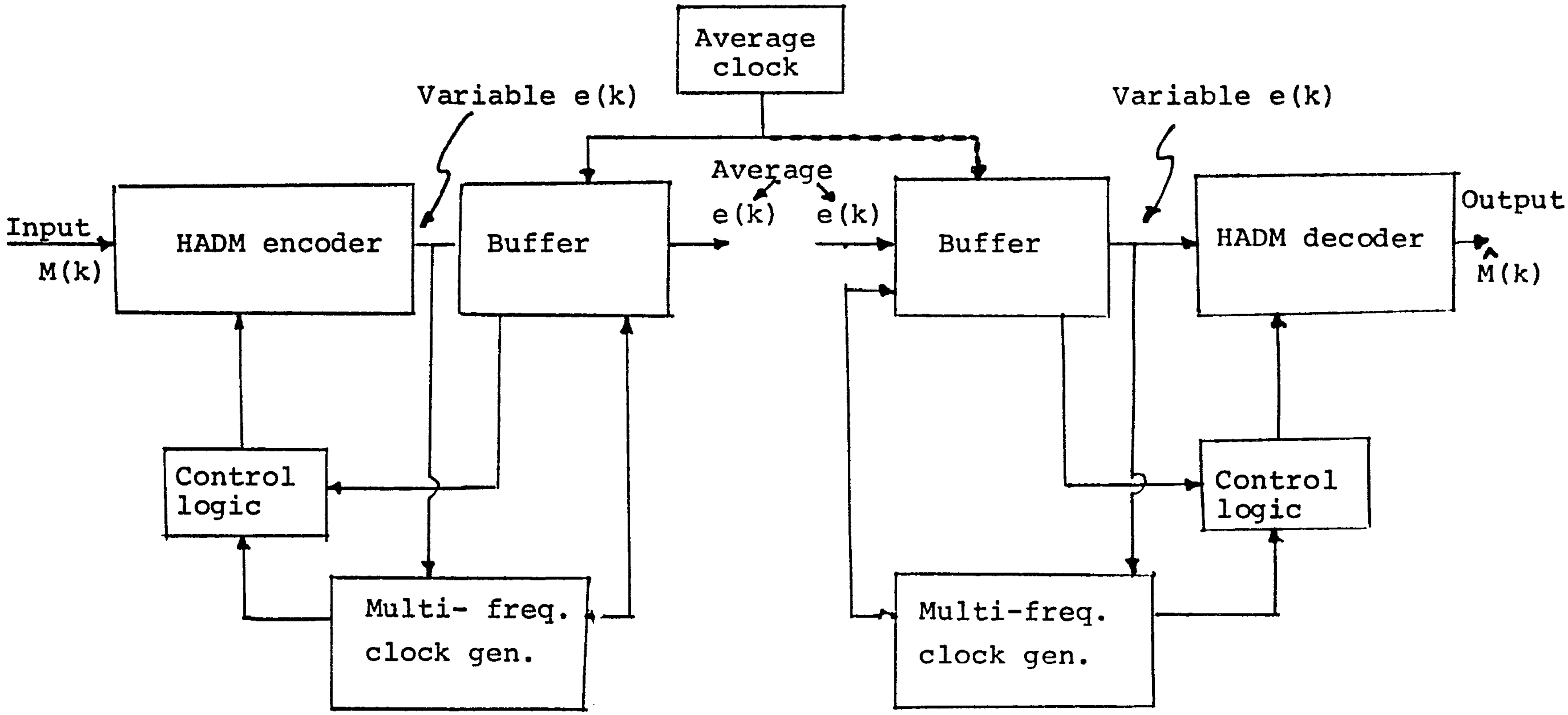


Fig. 8.1.1 Block diagram of the HSVADM

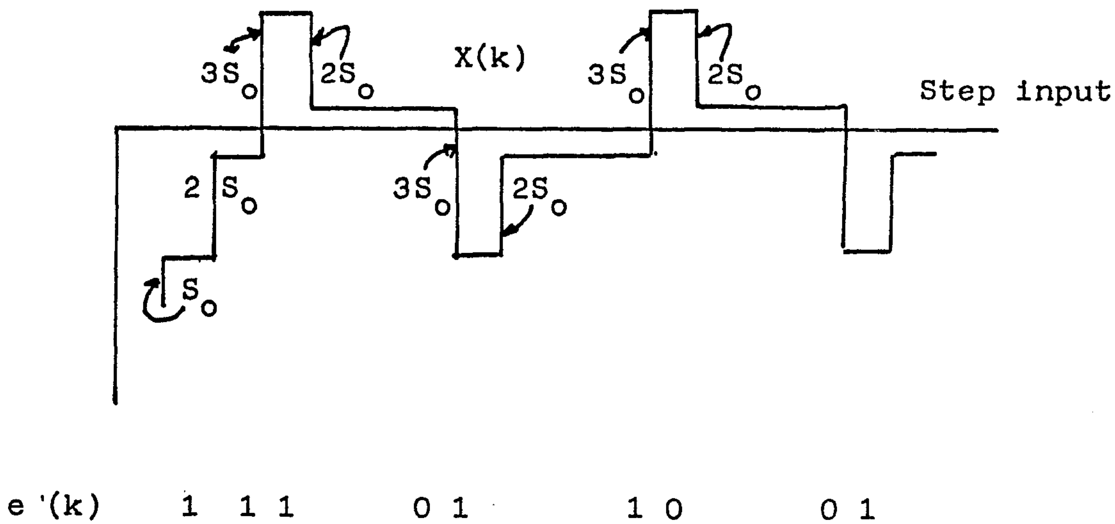


Fig. 8.1.2 Step response of the HSVADM

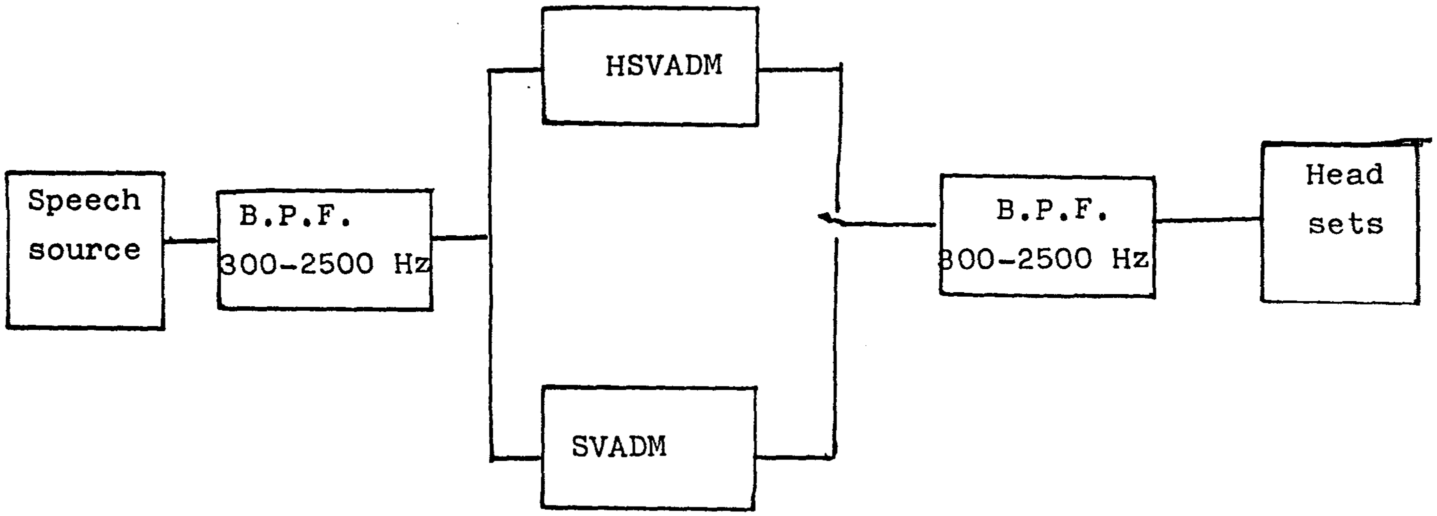


Fig. 8.2.1 Test set up for subjective comparison

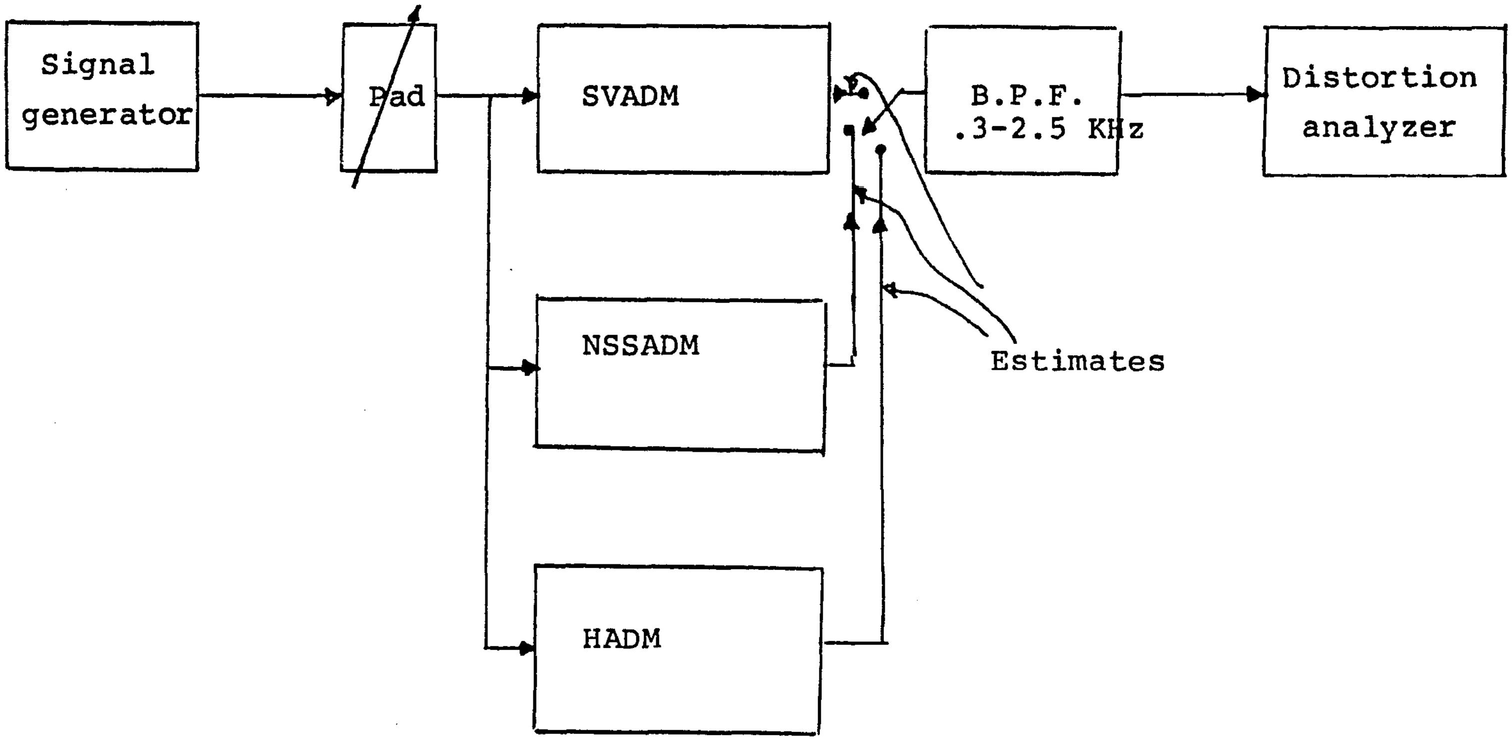


Fig. 8.3.1 Test set up for comparitive measurements

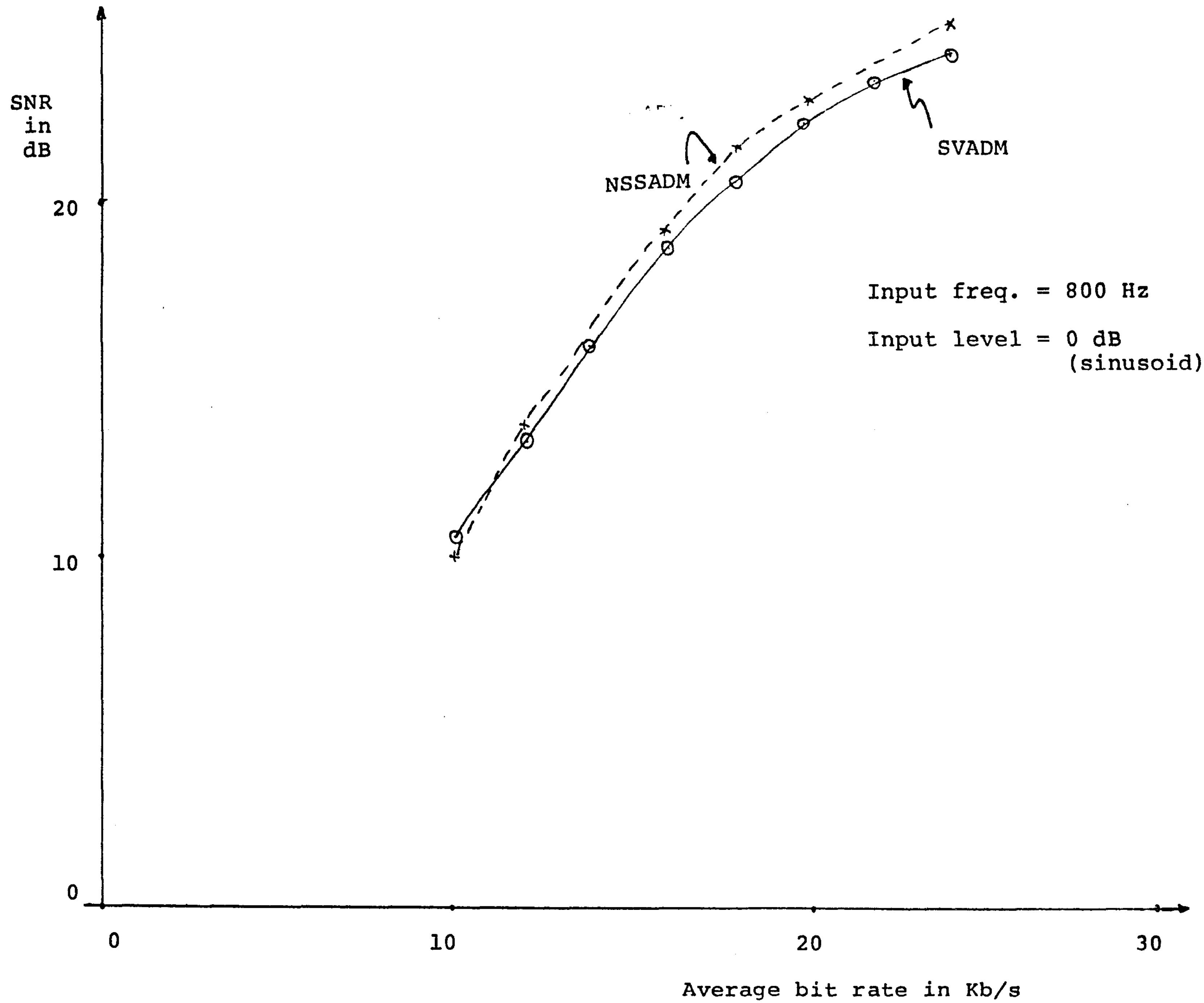


Fig. 8.3.2 SNR vs bit rate

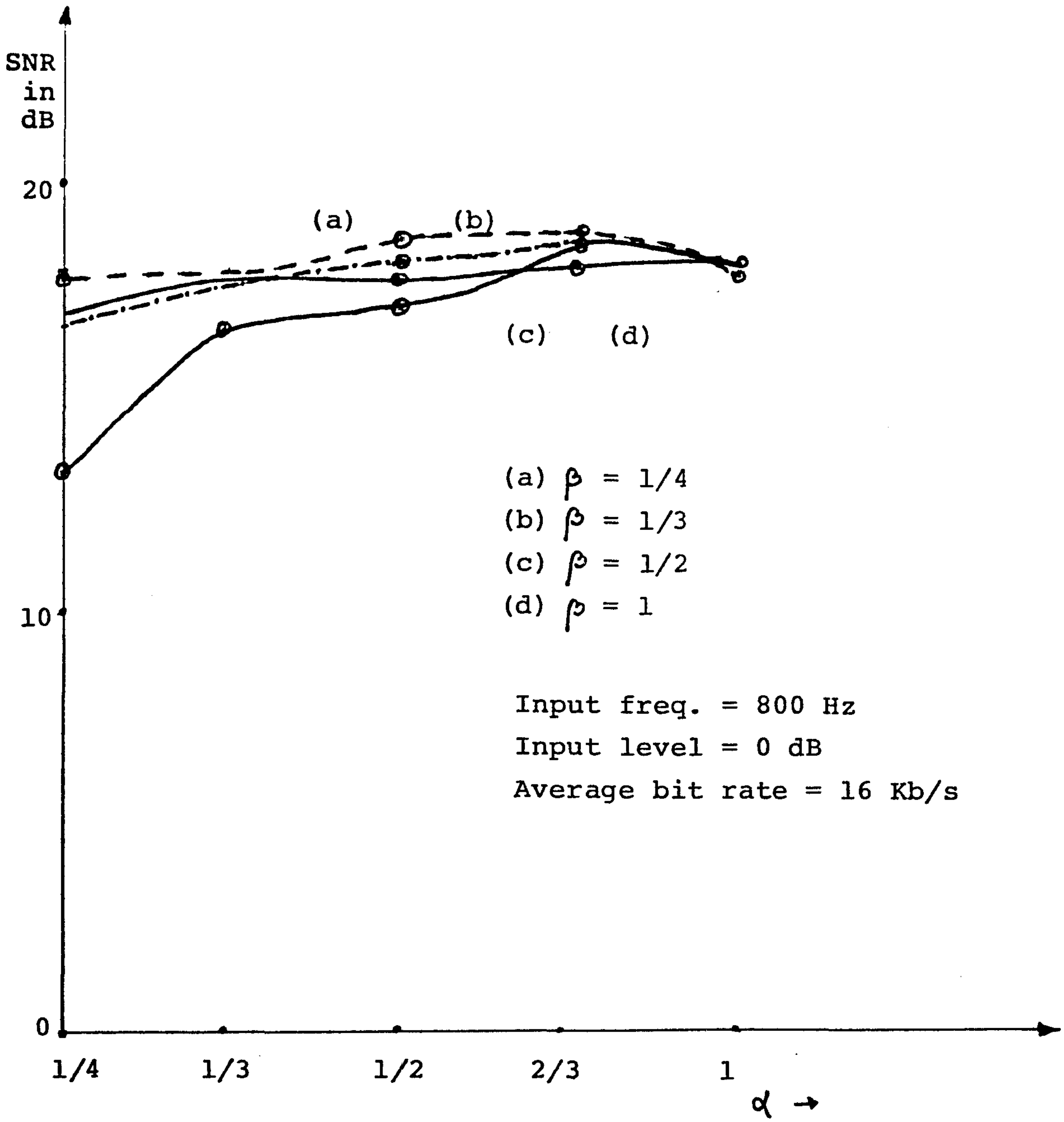


Fig. 8.3.3 SNR as a function of β, α and f_{av}

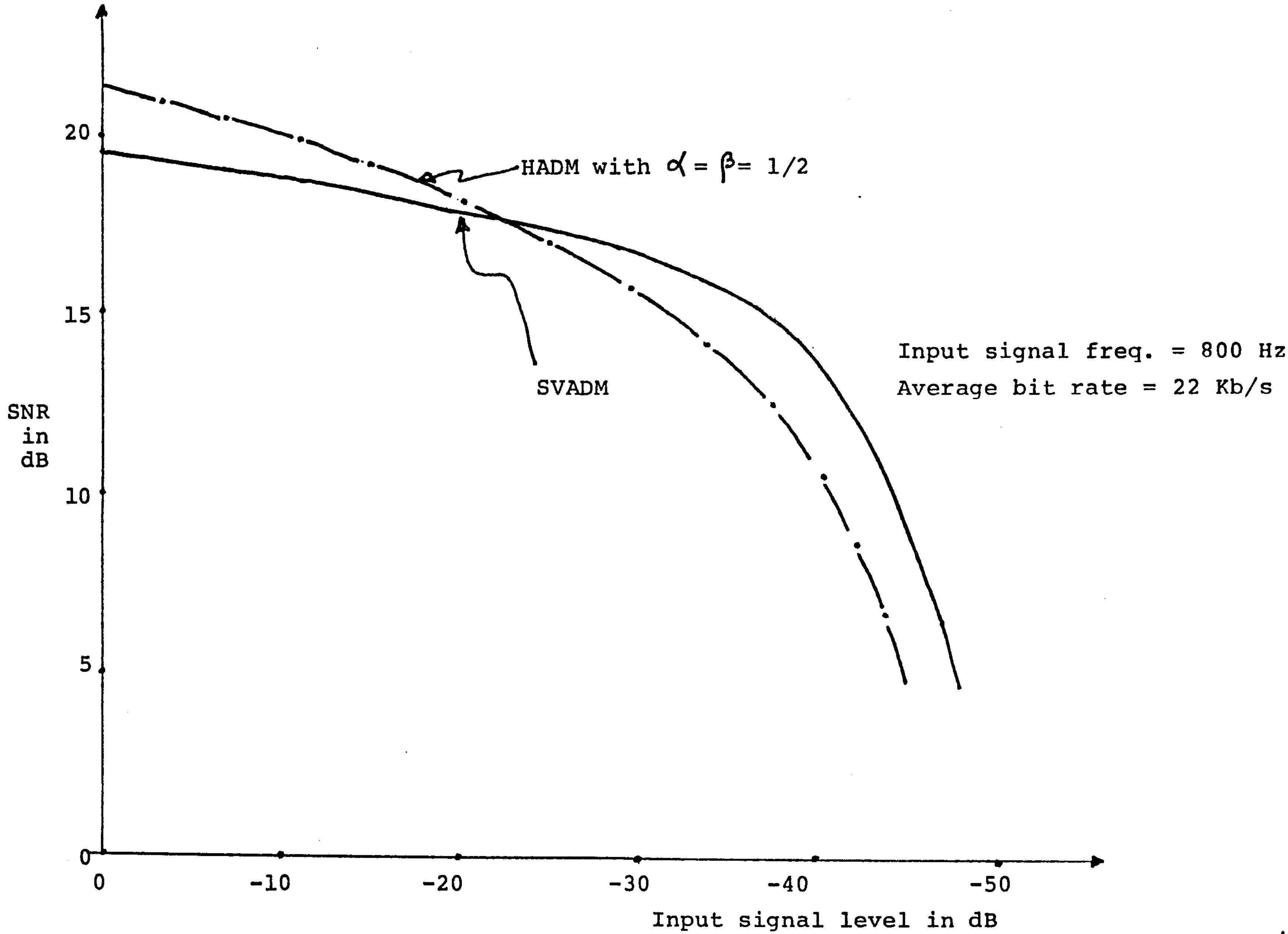
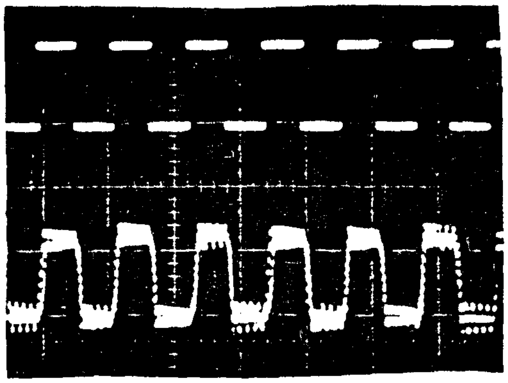
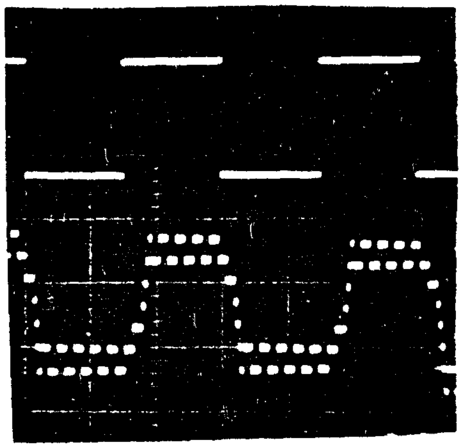


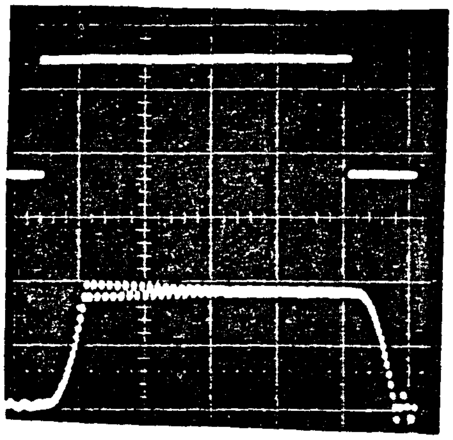
Fig. 8.3.4 SNR as a function of the input level



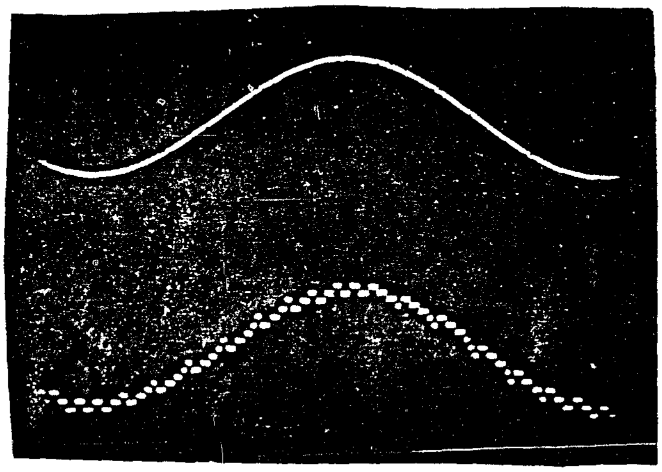
(a)



(b)



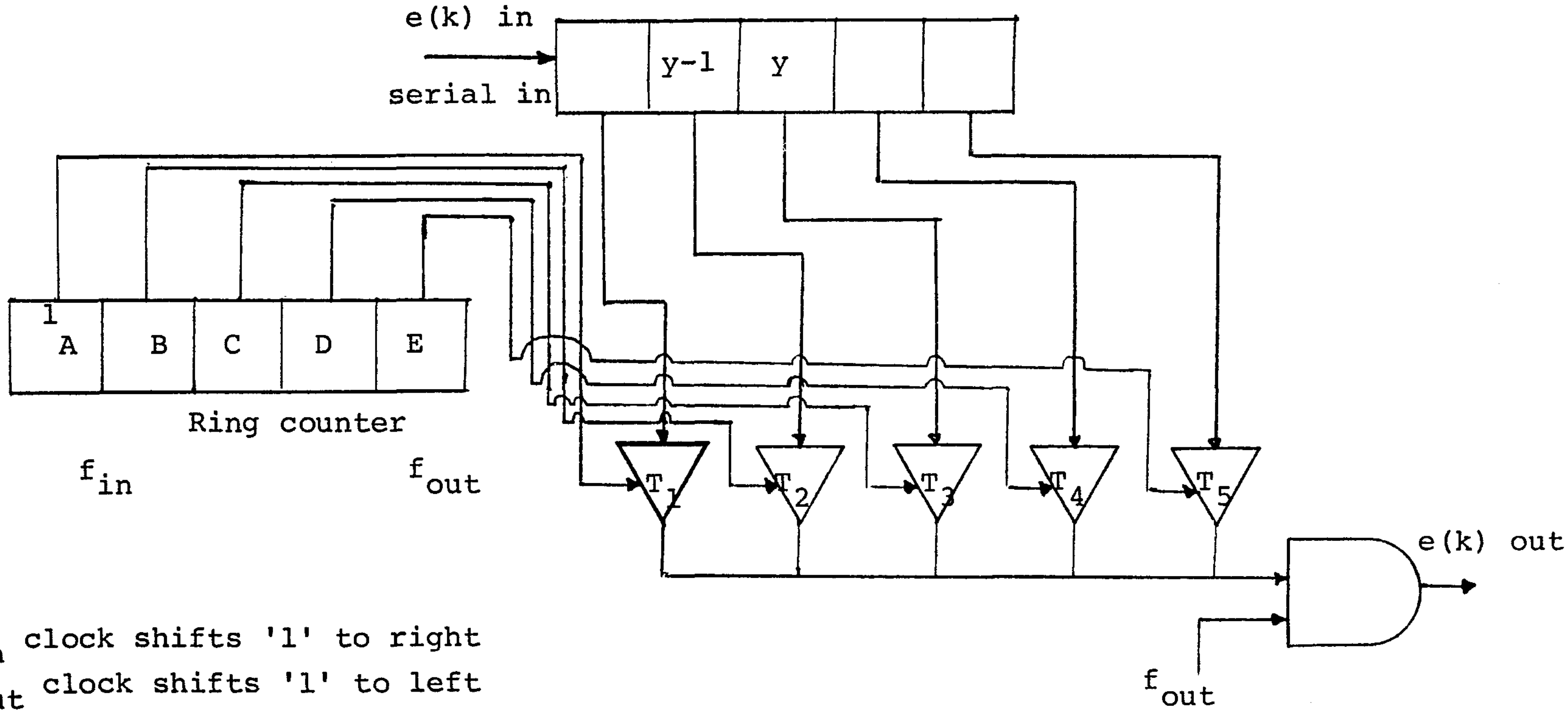
(c)



(d)

The top figure in each section is the input signal
The bottom figure in each section is the estimate of the HADM

Fig. 8.3.5 The input signal and its reproduction at the output of the HADM



* f_{in} clock shifts '1' to right
f_{out} clock shifts '1' to left
** T_1, T_2, T_3, T_4, and T_5 are tri-state devices

Fig. 8.4.1 Schematic diagram of the conceptual buffer

Table 8.3.1 SNR as a function of f_{av}, β and α.

Input signal freq. = 800 Hz (sinusoid)

Input signal level = 0 dB

HADM 1 - Three freq. adaptation with f_H, $1/2f_H$, and $1/4f_H$

HADM 2 - Three freq. adaptation with f_H, $1/2f_H$, and $1/3f_H$

HADM 3 - Two freq. adaptation with f_H, $1/2f_H$

f_{av} Kb/s	SVADM SNR dB	HADM 1 SNR dB	HADM 2 SNR dB	HADM 3 SNR dB
24	24	24	25	25
22	23.3	23	23.6	23.6
20	22	22.8	22.4	22
18	20.3	21.5	21.3	20.5
16	18.5	18	19.1	17.4
12	13.2	13.5	13.5	13
10	10.5	10	10	10.5

Table 8.4.1 Average bit rate for different values of f_H and f_L

f_H	f_L	f_{av}
K b/s	K b/s	K b/s
32	16	21.3
32	8	12.8
24	16	19.2
24	8	12
16	8	10.6

Chapter 9

Conclusions

9.1 Conclusions

In this final and concluding chapter, we present a summary of the contribution of this dissertation. In addition we present a brief discussion of the future research to be done.

The main object of this thesis is to use delta modulators as source encoders in packet voice networks and study the network performance. Initially we outlined the algorithms and performances of the CVSD and the SVADM in chapters 2 and 3 respectively. In chapter 4 we compared the performances of the CVSD and the SVADM as a function of the dynamic range, bit rate and bit error using subjective listening as the criteria. We established the fact that the SVADM offers a 10 to 15 dB higher dynamic range compared to that of the CVSD. Also, the subjective evaluation showed the SVADM was preferred to CVSD.

In Chapters 5,6 and 7, we explored the possible use of ADMs as source encoders in packet voice networks. We performed real time experiments using both the CVSD and the SVADM. The performance of the network showed the ADMs as source encoders can safely operate at 10^{-2} packet loss rate. Also, using the silence detection techniques, the packet transmission rate can be reduced to as low as 10 packets/second by not transmitting silent periods when ADM is operated at 16 Kb/s. Overall, we have studied the effects of two most important parameters on the processed voice, namely (a) packet loss rate (b) silence detection.

We developed algorithms for silence detection and packetization. We also developed receiver compensation algorithms for packet loss and for the gaps in the bit stream caused by not transmitting silent periods. We studied both fixed packet size and variable packet size packet voice networks. We found that the variable packet size scheme offers a higher dynamic range compared to that of the fixed packet size scheme.

In chapter 8, we developed a new delta modulator which has been referred to a the HADM. The HADM has not only step size adaptation but also has frequency adaptation. We were able to show that HADM was preferred to SVADM at bit rates of 16 Kb/s and lower. In addition, we found that the subjective dynamic range of the HADM is 10 dB lower than that of the SVADM unsder different conditions of operation. For comparison, we limited ourself to listening to the estimates of each encoder. We also, showed the design of a conceptual buffer required for the HADM to transmit e(k) at an average rate. At this time, we were able to simulate the buffer for non-real time operation only. Thus, there is an enormous amount of research required to be done in this area.

9.2 Suggestions for Future Work

In this dissertation, we have discussed the effects of packet loss and silence detection on the processed voice in a delta modulator packet voice network. However, another important parameter, i.e. packet delay in the network has not been considered. The future work must recognize this parameter. Furthermore, studies are to be made on packet networks, where both data and voice can be transmitted simultaneously. The effects of data on voice and vice versa have to be realized. This is particularly of interest to voice since delay constraints are heavy on voice.

Another problem to be considered is the voice teleconferencing using packet voice networks. Since the computers are the nodes in the network, all packets are assembled and labelled by them and therefore it appears that packets belonging to one customer could be transmitted simultaneously to several other customers to achieve teleconferencing.

The work on HADM is not complete unless a hardware to realize the buffer is built and the effect of bit loss rate on the processed voice is realized.

All these problems might themselves give out so many problems and thus we limit our discussions at this time to say that above problems have to be examined.

References and Bibliography

1. T.Apelewicz, " The design of a programmable real time ADM voice processor and the formulation of the auto correlation function for the LDM algorithm ", Ph.D. Thesis, Dept. of Elec. Engg., The City University of New York, Sept. 1978.

2. N.Abramson, " The ALOHA system- another alternative for computer communications ", Comput. Conf., AFIPS Conf. proc., Vol. 37, pp 281-285, 1975.

3. -----, " Packet switching satellites ", Nat. Comput. Conf., AFIPS Conf. proc., Vol. 42, pp 695-702, 1973.

4. J.E.Abate, " Linear and adaptive delta modulation ", IEEE proc., Vol. 55, pp 298-308, Mar. 1967.

5. S.J.Brolin, J.M.Brown, " Companded delta modulation for telephone transmission ", IEEE Trans., Vol. 16, Feb. 1968.

6. C.V.Chakravarthy, M.N.Faruqui, " A multi-digit adaptive delta modulation system ", IEEE Trans. Vol. Com-24, pp 931-935, Aug. 1976.

7. R.E.Crochiere, D.J.Goodman, L.R.Rabiner, M.R.Sambur, " Tandem connections of wideband and narrowband speech communication systems ", part 1- Narrowband to wideband link ", B.S.T.J., Vol. 56, No. 9, Nov. 1977.

8. V.R.Dhadesugoor, C.Ziegler, D.L.Schilling, " Delta modulators in packet voice networks ", IEEE Trans. on communication, Jan. 1980 (to be published).

9. V.R.Dhadesugoor, C.Ziegler, C.V.Chakravarthy, D.L.Schilling, " Delta modulation packet voice network with a variable size packet size algorithm ", NTC - 79 in Washington D.C.

10. V.R.Dhadesugoor, C.Ziegler, D.L.Schilling, " Source encoding using delta modulation in packet voice networks ", NTC -79 in Washington D.C.

11. V.R.Dhadesugoor, C.Ziegler, D.L.Schilling, " Digital silence detection in delta modulation packet voice networks ", ICC - 79 in Boston.

12. T.E.Eger, S.J.Campanella, " Study of an adaptive PCM voice communication system ", ICC - 74 in Mineapolis.

13. J.A.Greefkes, F.De Jager, " Continuous delta modulation ", Phillips Res. Rep., Vol. 23, pp 233 - 246, April 1968.

14. N.S.Jayant, " Adaptive delta modulation ", B.S.T.J., Vol. 49, pp 321-327, March 1970.

15. L.Kleinrock, F.Tobagi, " Packet switching in radio channels, part I - Carrier sense multiple access modes and their throuput delay characteristics ", IEEE Trans. Com Tech Vol. 23, pp 1400-1416, Dec. 1975.

16. -----, " Packet switching in radio channels part II - The hidden terminal problem and the carrier sense multiple access mode with a busy tone ", IEEE Trans., Com Tech, Vol. 23, pp 1417-1433, Dec. 1975.

17. J.Lo Locicero, " Ph.D. Dissertation at CUNY ".

18. Motorola Incorporated, " Semiconductor data library ", Vol. 6, Series B, pp 130-134.

19. J.B. O'Neal, " Delta modulation quantization noise and analog computer simulation results for gaussian and television input signals ", B.S.T.J., Jan. 1966.

20. L.G.Roberts, " Dynamic allocation of satellite capacity through packet reservation ", AFIPS Conf. proc., Vol. 42, pp 711-716, 1973.

21. C.L.Song, J.Garodnick, D.L.Schilling, " A variable step size robust delta modulator ", IEEE Trans. Com Tech Vol. 19, pp 1033-1040, Dec. 1971.

22. N.C.Seiler, R.Flower, J.Friedman, " A monolithic implementation of a CVSD algorithm ", ICC-77.

23. A.Tomozowa, H.kaneko, " Companded delta modulation for telephone transmission ", IEEE Trans. Com Tech Vol. 16, Feb. 1968.

24. C.J.Weinstein, " Fractional speech loss and talker activity model for TASI and for packet switched speech ", IEEE Trans. Com Tech Vol. 26, Aug. 1978.

25. M.R.Winkler, " High information delta modulation ", IEEE Int. Conf. Rec., Part 1, pp 285-290, 1965.

Appendix A

Amplitude-Frequency Characteristics of the Digital Low Pass Filter

In this appendix, we derive the expression for $H(w)$, i.e., Eq. (3.2.4), from the digital transfer function,

$$H(z) = (1 + z^{-1} + z^{-2} + z^{-3})/4 \qquad \text{(A-1)}$$

If we set

$$z = \exp(jwT_s) \qquad \text{(A-2)}$$

and separate terms, we obtain

$$H(w) = (1 + e^{-jwT_s})/4 + e^{-j2wT_s}(1 + e^{-jwT_s})/4 \qquad \text{(-3)}$$

Factoring Eq. (A-3), it reduces to

$$H(w) = (1 + e^{-jwT_s})(1 + e^{-j2wT_s})/4 \qquad \text{(A-4)}$$

By removing $e^{-jwT_s/2}$ from the first term of this product and e^{-jwT_s} from the second term, we reveal a familiar form

$$H(w) = e^{-jwT_s/2}\,(e^{jwT_s/2} + e^{-jwT_s/2})/2 \qquad \text{(A-5)}$$
$$\cdot\; e^{-jwT_s}\,(e^{jwT_s} + e^{-jwT_s})/2$$

Recognizing the exponential definition of the cosine, we can now write

$$H(w) = e^{-j3wT_s/2}\cos(wT_s/2)\cos(wT_s) \qquad (A-6)$$

Thus, by observation, we see that the amplitude-frequency characteristics of the digital low pass filter is given by

$$H(w) = \cos(wT_s/2)\cos(wT_s) \qquad (A-7)$$

Appendix B

Program for the SVADM encoder

Loc	Instruction	Code
0	CSR T	010 1 1100 1010
1	ILR R_1	000 1 0001 1010
2	AIA AC	011 0 1111 0010
3	ALR AC-	000 0 1101 1110
4	CIA T	001 1 1110 0010
5	ILR T	000 1 1100 1010
6	SDR R_2	010 0 0010 0010
7	ILR R_0	000 1 0000 1010
8	SDR T	010 1 1100 0010
9	ALR AC-	000 0 1101 1110
10	CIA T	001 1 1110 0010
11	ILR R_2	000 1 0010 1010
12	ALR AC-	000 0 1101 1110
13	CIA T	001 1 1110 0010
14	ILR R_3	000 1 0011 1010
15	ALR T	000 0 1100 1010
16	SDR R_0	010 0 0000 0010
17	ALR R_1	000 0 0001 1010
18	CMA AC	111 1 1111 1010
19	NOP(out)	110 1 0000 1000
20	ILR R_2	000 1 0010 1010
21	SDR R_3	010 0 0011 0010
22	JMP 255	111 1 1111 1011
255	NOP	110 1 0000 1010