RESEARCH ON ADAPTIVE DELTA MODULATORS

Sponsored by the

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

ARPA Order No. 3534

Contract No. MDA 903-78-C-0182

Monitored by Dr. Vinton Cerf

Technical Report ARPA - TR - 1

Date of Contract 3 Jan. 78

Date Submitted 12/15/78

Expiration Date 28 Feb. 79

COMMUNICATIONS SYSTEMS LABORATORY

DEPARTMENT OF ELECTRICAL ENGINEERING

THE CITY COLLEGE OF
THE CITY UNIVERSITY of NEW YORK

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARPA - TR - 1 (18)	2. GOVT ACCESSION NO. (9) Interim rept. Mar - Dec 78	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) RESEARCH ON ADAPTIVE DELTA MODULATORS.		5. TYPE OF REPORT & PERIOD COVERED Interim 3/78 to 12/78
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) V. R. Dhadesugoor, C. Ziegler, D. L. Schilling		8. CONTRACT OR GRANT NUMBER(s) (15) MDA-903-78-C-0182 ARPA Order-3534
9. PERFORMING ORGANIZATION NAME AND ADDRESS ~~Research Foundation of CUNY on behalf of~~ New Communications Systems Lab. The City College New York, N. Y. 10031		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (11) 15 Dec 78
11. CONTROLLING OFFICE NAME AND ADDRESS ARPA (Code HX 1243) 1400 Wilson Blvd. Arlington, Va 22209		12. REPORT DATE 12/15/78
		13. NUMBER OF PAGES 71
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) ONR 715 Seventh Avenue New York, N. Y. 10003 (12) 75p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

2 copies Director, ARPA, Att
1400 Wilson Boulevar
Arlington, VA 22209
12 copies Defense Documentati

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the officicial policies, either express or implied, of the Defense Advanced Research Projects Agency or the United States Government.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

(1) Adaptive Delta Modulator
(2) CVSD
(3) Packet Voice
(4) Silence Detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this report, the performance of the Song Mode Voice Adaptive Delta Modulator (SVADM) and the Continuously Variable Slope Delta Modulator (CVSD), in terms of dynamic range, sampling rate and the channel errors, are compared. The use of the SVADM and the CVSD in a packet voice channel, the algorithms for digital detection of periods of silence and the performance of a packet voice channel using the SVADM and the CVSD as source encoders are presented. The parameters employed for subjective evaluation of the packet voice channel are packet size, silence detection algorithm, bit rate and packet loss rate.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

411173

RESEARCH ON ADAPTIVE DELTA MODULATORS

Sponsored by the

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

ARPA Order No. 3534

Contract No. MDA 903-78-C-0182

Monitored by Dr. Vinton Cerf

Technical Report ARPA - TR - 1

Date of Contract 3 Jan. 78 Date Submitted: 12/15/78

Expiration Date 28 Feb. 79

Donald L. Schilling

Donald L. Schilling

(212) 862-3737

Principle Investigator

COMMUNICATIONS SYSTEMS LABORATORY

DEPARTMENT OF ELECTRICAL ENGINEERING

Table Of Contents

ABSTRACT

In this report, the performances of the Song Mode Voice Adaptive Delta Modulator (SVADM) and the Continuously Variable Slope Delta Modulator(CVSD) in terms of dynamic range, sampling rate and the channel errors are compared. The use of the SVADM and the CVSD in a packet voice channel, the algorithms for digital detection of periods of silence and the performance of a packet voice channel using the SVADM and the CVSD as source encoders are presented. The parameters employed for subjective evaluation of the packet voice channel are packet size, silence detection algoritm, bit rate and packet loss rate.

I. INTRODUCTION

The use of Delta Modulators as source encoders have been emphasized in recent times. The search for techniques which will lower the bit rate and hence increase efficiency without significant loss of quality has yielded several adaptive delta modulators (5 - 11). Our discussion is limited mainly to the Song Mode Voice Digital Adaptive Delta Modulator (SVADM) algorithm for digitizing speech (1). The SVADM algorithm is easily implemented and produces good quality speech at fairly low bit rates. Another processing device which produces good quality speech is the Continuously Variable Slope Delta Modulator (CVSD) (2). Unlike the CVSD which is specifically designed to encode speech, the SVADM responds also to non-speech signals.

Section II of this report describes the algorithm used in the SVADM. This algorithm is extremely inexpensive to implement and has graceful degradation of voice quality in the presence of transmission errors. It also has a 40 dB dynamic range and 90% of word intelligibility at a bit rate of 9.6 Kb/s.

Section III describes the implementation of the CVSD. Several CVSD processors have been developed (2, 14, 15) and each one of them is slightly different. We only describe the principle of the CVSD algorithm.

Section IV compares the performances of the SVADM and the CVSD in terms of dynamic range, transmission errors and sampling rate. For the experiments, the Harris CVSD and the Motorola CVSD are used.

Section V introduces the concept of packet voice in packet networks. It describes briefly the ARPA packet radio network.

Section VI describes the concept of using the delta modulators as source encoders in packet radio networks. The performances of the SVADM and the CVSD are studied in terms of packet size and packet loss rate.

Section VII describes the silence detection and speech initiation algorithms which have been succesfully used to reduce the packet rate transmission. It is well known that speech signals contain a large percentage of quiet periods. Eger and Campenella(12) claim that 50 % of conversational speech is quiet. In a general network like the time assigned speech interpolation (TASI) network (13), the system permits a number of sources to share a smaller number of channels through voice activated switching. The use of silence detection would enable each individual source to transmit at lower packet rates by not transmitting during quiet periods. The SVADM encoder generates a periodic output during the quiet period of speech. Silence detection will be shown to be accomplished by detecting, digitally, the periodic output of the SVADM rather than the analog speech detection used in the TASI system (13).

II. THE SONG MODE VOICE DIGITAL ADAPTIVE DELTA MODULATOR (SVADM)

The SVADM encoder - decoder is a robust delta modulator system, with a dynamic range of 40 dB and word intelligibility of 99% at 16 Kb/s bit rate, and more than 90% of word intelligibility at 9.6 Kb/s bit rate. It is easy to implement digitally.

ALGORITHM:

The Algorithm of the SVADM is

$$X(k+1) = X(k) + S(k+1) \qquad (2.1)$$

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

$$S(k+1) = |S(k)|\, e(k) + S_o\, e(k-1) \qquad (2.2)$$

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 is employed and therefore $S_o \approx 10$ mv. If M (k) is the signal value at time k/f_s, then

$$e(k) = \mathrm{sgn}\left[M(k) - X(k)\right] \qquad (2.3)$$

The new step size S (k+1) differs, in magnitude from the old step size by $\pm S_o$ as evident from equation (2.2).

The complete block diagram of the SVADM is shown in fig. 2.1. Note that the feedback circuit of the encoder is essentially the decoder.

OSCILLATIONS:

If the input is constant, the SVADM reaches a steady state condition. Figure 2.2 shows the response of the SVADM to a step input. In the steady state, the response of the SVADM is an estimate signal which exhibits a periodic pattern repeating after every four samples. Also, a periodic e (k) pattern of 11001100... is generated.

Thus the reconstructed output of the SVADM oscillates with a fundamental frequency of $f_s/4$. The amplitude of the oscillation depends on the step size at the time of oscillation and usually is only 2 S_o when voice signals are encoded. The effect of these oscillations are eliminated by using a digital low pass filter at the output of the SVADM encoder.

THE DIGITAL LOW PASS FILTER:

The Digital Low Pass Filter (DLPF) shown in Fig. 2.1, is a four term non-recursive filter. The output of the DLPF is $\hat{\hat{M}}(k)$ which is given by

$$\hat{\hat{M}}(k) = (1/4) \left[X(k) + X(k-1) + X(k-2) + X(k-3) \right] \tag{2.4}$$

If the steady state output of the SVADM shown in Fig. 2.2 is passed through the DLPF, then $\hat{\hat{M}}(k) = X_q$ for all k, where X_q is the quantized signal. Thus after a four term averaging, the SVADM output produces a constant D.C. level in the steady state. The frequency tone of $f_s/4$ is eliminated. To illustrate the frequency response of the DLPF, it is possible to determine the digital transfer function H(z) where $z = \exp(jwT_s)$. Figure 2.3 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 has been shown that it is the first zero that eliminates the periodic steady state component in the SVADM response.

The DLPF is necessary to eliminate the tone generated at $f_s/4$ only when the SVADM is operated at twice the Nyquist rate. Otherwise, at higher sampling rates , the DLPF need not be used. For bandlimited speech signals (300 Hz-2500 Hz), if the SVADM operates at f_s = 10 Kb/s, the DLPF has the first zero at $f_s/4$ = 2.5 Kb/s and therfore it eliminates the tone at $f_s/4$. However, because of its low pass characteristics, it also attenuates some baseband frequencies. Therefore, the output of the DLPF is then passed through a preemphasis filter to boost the attenuated baseband frequencies.

PREEMPHASIS FILTER:

Figure 2.4 shows the implementation of the preemphasis filter and its frequency response. The frequency response is plotted only up to 4 KHz. The typical high pass characteristics has been achieved by varying the capacitance.

It should be noted, however, that the output characteristics has to be adjusted depending on the input speech bandwidth and the sampling rate. The frequency response shown in Fig. 2.4 has been adjusted for input speech from 300 Hz to 2500 Hz and the sampling rate of f_s = 10 Kb/s. When f_s is larger than 10 Kb/s, it should be noted that both the DLPF and the preemphasis filter are not required at the output of the decoder. Subjective evaluation showed a significant improvement in the performance of the SVADM system operating at f_s = 10 Kb/s when using the DLPF and the preemphasis filter.

OVERFLOW DETECTION LOGIC:

In addition to the above modification to the basic SVADM algorithm an overflow detector is needed in the receiver in conjunction with the estimate X(k). In the presence of channel noise, the estimate X(K+1) may overflow, when X(k) is large and S(k+1) is large. This results in a large change in the estimate and is not acceptable during speech.

The logic is very simple to implement digitally. In the SVADM, the new estimate is realized by

$$X(k+1) = X(k) + S(k+1) \tag{2.5}$$

Overflow occurs when,

$$\overline{[\mathrm{Sgn}\ X(k) \oplus \mathrm{Sgn}\ S(k+1)]}\ [\mathrm{Sgn}\ X(k+1) \oplus \mathrm{Sgn}\ X(k)] \tag{2.6}$$

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).

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 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:

$$X(k+1) = X(k) + S(k+1) \tag{2.7}$$

$$S(k+1) = |S(k)|\, e(k) + S_o\, e(k-1) \tag{2.8}$$

$$e(k) = \mathrm{Sgn}\left[M(k) - X(k) \right] \tag{2.9}$$

DECODER:

$$X'(k+1) = X'(k) + S'(k+1) \tag{2.10}$$

$$S'(k+1) = |S'(k)|\ e'(k) + S_0\ e'(k-1) \tag{2.11}$$

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 transmitter and the receiver estimates, as

$$N(k+1) = X'(k+1) - X(k+1) \tag{2.12}$$

$$= X'(k) - X(k) + S'(k+1) - S(k+1) \tag{2.13}$$

therefore,

$$N(k+1) = N(k) + S'(k+1) - S(k+1) \tag{2.14}$$

Equation 2.14 shows that the noise voltage accumulates as the errors come through the system. If we use a leaky integrator with a leak factor $0<L<1$, we can rewrite the equations (2.7), (2.10), and (2.14) as

$$X(k+1) = L \cdot X(k) + S(k+1) \tag{2.15}$$

$$X'(k+1) = L \cdot X'(k) + S'(k+1) \tag{2.16}$$

$$N(k+1) = L \cdot N(k) + S'(k+1) - S(k+1) \tag{2.17}$$

The value of L usually depends on the sampling rate. It has been found experimentally that for input speech of 300 Hz - 2500 Hz band, the leak factor L should be as shown in Table 2.1.

In Table 2.1, 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 subtracting the shifted estimate from the original estimate. The values of L listed in Table 2.1 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}, 10^{-2} 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

$$X(k+1) = X(k) + S(k+1) + \beta S_0 \tag{2.18}$$

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

$$X(k) = x_0 . x_1 x_2 x_3 \ldots x_{N-1} \tag{2.19}$$

and

$$S(k+1) = s_0 . s_1 s_2 s_3 \ldots s_{N-1} \tag{2.20}$$

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,

$$\beta = \begin{cases} +1 & \text{if } x_0 = s_0 = 1 \text{ and } x_{N-1} \oplus s_{N-1} = 0 \\ -1 & \text{if } x_0 = s_0 = 0 \text{ and } x_{N-1} \oplus s_{N-1} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.21}$$

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 leaky 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

$$X(k+1) = X(k) + S(k+1) + \beta S_0 \tag{2.22}$$

where, for X(k) negative

$$\beta = +1 \text{ when } x_0 = s_0 = 1,\ x_1 = x_2 = x_3 = 0 \text{ and } x_{N-1} \oplus s_{N-1} = 0 \tag{2.23}$$

for X(k) positive

$$\beta = -1 \text{ when } x_0 = s_0 = 0,\ x_1 = x_2 = x_3 = 1 \text{ and } x_{N-1} \oplus s_{N-1} = 1 \tag{2.24}$$

and

$$\beta = 0 \text{ otherwise.} \tag{2.25}$$

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 -40 dB and thus has a larger dynamic range over the non-linear leaky integrator 1. For comparing SVADM with the CVSD we shall use the non-linear leaky integrator 2.

PERFORMANCE OF SVADM ENCODER-DECODER:

Figure 2.5 shows the test set up used. It should be noted that, in this experiment each of the leaky integrator algorithm was tested and measurements were made. In addition, a sinusoidal signal was used for the experiment. In all, five different performance measures were tested.

1) Bandwidth:

we coducted a selective test which showed that at $f_s = 37.5$ Kb/s, the output level varied within $\pm$ 1 dB over an input frequency range of 600 Hz - 2400 Hz and within $\pm$ 3 dB for input frequency range of 300 Hz - 3400 Hz (the band pass filters were set for 300 Hz - 3400 Hz). This test was performed at input levels of 0 dB and -10 dB. Figure 2.6 is the plot of output level vs the input frequency. The same result was obtained for all the three types of leaky integrators.

This result shows that the SVADM offers a good bandwidth for speech inputs.

2) Idle Channel Noise:

For this test, we used a C message weighted filter shown in Fig. 2.7 at the output of the decoder. In order to measure the Idle channel noise, we terminated the input signal at the encoder and then we measure the noise level at the decoder output using an RMS meter. The experiment showed that

a) The idle channel noise with no channel errors is 65 dB below the maximum input signal.

b) The idle channel noise with 10^{-3} channel error rate is 55 dB below the maximum input signal.

3) Dynamic range using a single tone of 1 KHz :

The dynamic range is defined as the input range for which the output signal to noise ratio (SNR) is greater than or equal to 25 dB. There are several definitions existing for dynamic range. We use the definition as described above. For this test, a 1 KHZ sinusoidal signal was used. For this test, the bandpass filters were set from 300 Hz to 3400 Hz. All three leaky integrators were considered. The input signal level was varied from 0 dB down to - 40 dB in steps of 10 dB and the SNR at the output of each of the decoder was measured. The results of the three SVADMs are plotted in Fig. 2.8 (a). The SVADM using the true leaky integrator (TLI) offers higher SNR at lower input levels compared to the other two. However, the SVADM using the non-linear integrator 2 (NLI 2) has a better dynamic range over the SVADM using non-linear integrator 1 (NLI 1). We see from Fig. 2.8 (a), that the dynamic range, as defined above is 30 dB.

The ultimate test for the dynamic range, however, is the subjective quality of the processed speech, We will present the subjective dynamic range measurement for the SVADM in Section IV.

4) SNR vs bit rate f_s:

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. We vary the bit rate f_s down to 6 Kb/s and measure the SNR using the distortion anlyzer. Figure 2.8 (b) shows the graph of the SNR vs f_s of the SVADM (NLI 2), which will eventually be used for subjective comparison with the CVSD.

The SVADM appear to degrade almost linearly with the bit rate, f_s, down to 8 Kb/s. However, when the SVADM is operated at bit rates lower than 8 Kb/s, the degradation increases significantly.

5) Linearity:

This test is performed to show that the output to input amplitude ratio of the SVADM is constant for different input levels. A test signal of 1 KHz was used.

For input levels from 0 dB to 35 dB below maximum, the output level, usually, is desired to be within ± 0.5 dB for each level of the input. It was found that, all three SVADMs met this requirement.

In Section IV, we shall compare the performances of CVSD and SVADM for subjective quality of the processed speech.

III CONTINUOUSLY VARIABLE SLOPE DELTA MODULATOR (CVSD)

The CVSD is an adaptive delta modulator specifically used as a 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 3.1 shows the block diagram of the CVSD in general. The general algorithm is given by

$$X(k+1) = \alpha X(k) + \left| (1-\alpha)\ \Delta(k) \right|\ e(k) \tag{3.1}$$

where,

$$e(k) = \mathrm{Sgn}\left[M(k) - X(k) \right] \tag{3.2}$$

and

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

α is the leakage factor associated with the estimate integrator,

$\Delta(k)$ is the k^{th} step size and

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

Furthermore, $\Delta(k)$ is generated by syllabic companding and is given by

$$\Delta(k+1) = \beta\ \Delta(k) + (1-\beta)\ (V+V_1) \tag{3.3}$$

where,

V is a constant voltage when three consecutive outputs from the CVSD encoder are identical (Sometimes this number could be two or four (2, 14, 15, 16)). V_1 is just a constant voltage added to V to ensure that the minimum step size is non zero.

In Fig. 3.1, the output of the overload detector is either 0 or V volts depending on the three consecutive digital outputs of the CVSD. The feedback circuit of the encoder is the CVSD decoder.

In particular, the CVSD described in (16) 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

$$\beta = \exp\left[(-1/f_s) / (5.69 \times 10^{-3})\right] \tag{3.4}$$

$$\alpha = \exp\left[(-1/f_s) / 10^{-3}\right] \tag{3.5}$$

When $f_s = 16$ Kb/s, $\beta = 0.99$ $\alpha = 0.94$

It is of interest to note that the coefficients α, β 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 1msec. which corresponds to 1000 Hz. Figure 3.2 describes the simple circuit implementation of CVSD encoder.

IV PERFORMANCE COMPARISON OF SVADM AND CVSD

The ultimate performance measure of the system is the subjective quality of the processed voice. Therefore, in this section we describe the subjective comparison of the CVSD and the SVADM for voice signals.

Figure 4.1 shows the test set up used for the subjective comparison of theCVSD and the SVADM. The speech was bandlimited from 300 Hz to 2500 Hz by a four pole (Butterworth) filter. Two speech tapes were used, a Mark Twain story and a taped radio conversation and in addition, a third tape consisting of a set of group test words was used. The following tests were performed.

Listeners preference test of performance of delta modulators as a function of sampling rate:

Several listeners participated in the test at different stages. With the comments available from the listeners, the result has been tabulated in Table 4.1. The Table describes the performances of the Motorola CVSD and the SVADM. Listeners preferred the Motorola CVSD over the Harris CVSD. For the comparison with the SVADM, we, therefore, used the Motorola CVSD.

At the maximum input signal level (i.e. 0 dB), for bit rates of 32 Kb/s and 24 Kb/s, the performance of the SVADM and the CVSD were the same. However, at bit rates of 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.

Dynamic Range(Subjective):

We have seen a 30 dB dynamic range for the SVADM, when using a single sinusoidal tone of 1 KHz. The Motorola CVSD also claims to have a 30 dB dynamic range using a single tone as the input to the CVSD (14). The ultimate test performance, however, is the subjective quality of the processed speech.

Using the test set up shown in Fig. 4.1, we attenuated the input signal in steps of 10 dB. The results of the test performance of both the CVSD and the SVADM are described in Table 4.2. At f_s = 32 Kb/s, the SVADM exhibited a 40 dB dynamic range while the has a 30 dB dynamic range. As we lower the bit rate, the dynamic range decreased for both the systems. Figure 4.2 shows the dynamic range 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.

Cascading:

To determine the performance of a working system, the delta modulators were cascaded. The estimate of the first DM system is processed by the second DM system. The subjective quality of the output of the second DM 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 to 9.6 Kb/s, whereas, the Cascaded SVADM recovers the signal under the same conditions, although some quantization noise is present. At signal levels of -10 dB and -20 dB, Cascaded SVADM has a much higher intelligibility as compared to the Cascaded CVSD. This test is very useful to study the feasibility of using the delta modulators in repeater stations.

Channel Noise:

Figure 4.3 shows a method of generation of random errors. The system consists of a noise generator, a comparator, and combinatoric logic. The noise generator produces an analog gaussian noise voltage. V_t is the threshold voltage of the comparator and is varied to generate different error rates. The noise voltage is compared to the threshold voltage and if it exceeds the threshold, the D flip- flop shown in Fig. 4.3 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 for error rates of 10^{-4}, 10^{-3}, 10^{-2} and 10^{-1}.

Figure 4.4 shows the test set up for the subjective evaluation of the CVSD and the SVADM in the presence of errors. The input speech signal was bandlimited from 300 Hz to 2500 Hz. Table 4.3 describes the performances of the CVSD and the SVADM in the presence of errors. At 0 dB input level, f_s = 32 Kb/s and the error rates of 10^{-1}, the CVSD was preferred to the SVADM. Under all other conditions the SVADM was preferred to the CVSD. At -20 dB input level, 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 seems to decrease with the errors. Figure 4.5 shows the dynamic ranges of the CVSD and the SVADM as a function of error rates at different sampling rates. From Fig. 4.5, we conclude that the SVADM has about 10 - 15 dB higher dynamic range over that of the CVSD even in the presence of errors.

V. INTRODUCTION TO PACKET VOICE NETWORKS

With the advent of packet-switching technology, the economic sharing of computer resources over a wide geographic area has become possible through the use of computer-communication networks. As such networks grow in size and coverage, the need to provide inexpensive, long-haul, high-capacity communications channels becomes more pressing. In addition, there is also the local interconnection problem, that is, the problem of providing inexpensive communications from the users, possibly mobile, terminals into the high-level network itself. In response to these growing needs, ARPA has undertaken the development of new techniques which include the use of packet-switching over a broad-band satellite channel as a solution to the long-haul problem and also the use of ground radio packet-switching for local access.

In the use of a broad-band satellite channel for packet-switching, there are two extremely interesting characteristics that are of great importance. First is the long propagation delay in a roundtrip transmission (e.g. source-satellite-destination) to a satellite repeater in synchronous orbit some 36,000 km above the earth; this delay is approximately 0.25 s. Second, the repeater can retransmit back to earth in a broadcast mode to all earth stations in its broadbeam "shadow". Thus, each transmitter can listen to his own transmission, thus providing "perfect feedback" that gives us automatic acknowledgements.

There are many ways to use a given satellite channel for data communications. However, because of the above mentioned characteristics and the bursty (i.e. high ratio of peak-to-average) nature of the traffic, random access schemes have been used to yield the most efficient use of the channel.

One of the first random access systems developed was the ALOHA system. In the "pure ALOHA" system(17), the users transmit packets any time they desire. If after one propagation delay they "hear" their successful transmission, they can assume that no conflict occurred; otherwise, they know a collision or some other source of noise caused partial or complete destruction of the packet, and they must retransmit. If all users retransmit upon determining a collision, then another collision is certain to occur. Hence, a random retransmission delay must be introduced to avoid such a possibilty. A second method for using the satellite channel is called "slotted ALOHA" (18). In this system, time is slotted into segements whose duration is equal to the transmission of a single packet, time being referenced to the satellite. In such a system, all collisions are total and not partial and the achievable system efficiency is increased by a factor of two. Yet another method (19) for using these channels is to employ a reservation system in which time slots are reserved on a fixed or demand basis for specific users' transmissions.

For ground radio systems, in which the roundtrip propagation delay is small compared to a packet transmission time, a fourth method for using the packet-switched channel has been developed; namely, carrier sense multiple access (CSMA) (20). In CSMA, the terminal listens to ("senses") the channel, and if the carrier signal is heard, the terminal realizes that the channel is busy and will postpone its own transmission until the channel is sensed idle. One problem with CSMA is the assumption that all terminals are in line-of-sight and within range, not only of the central station but also with each other. For terminals within range of a central station but out of range ("hidden") of each other, busy tone multiple access (BTMA) (20) has been developed. In this scheme, as long as the central station senses a carrier on the incoming message channel, it transmits a busy tone (sine wave) on the busy tone channel. It is by sensing a signal on this busy tone channel that terminals determine when the message channel is busy. However, as mentioned above, BTMA still assumes that all terminals are in line-of-sight and within range of the central station.

During the past several years ARPA has been developing a packet-switched radio communication system called The Packet Radio Network (PRN). Packet radio was developed to permit packet-switched radio communication among geographically distributed, fixed or mobile-user terminals and to provide improved frequency management strategies to meet the critical shortage of the RF spectrum. The system is to be capable of providing real-time voice communication as well as essentially error free data communication services. In the Packet Radio network there are many terminals, repeaters and stations. Generally no particular station is in line-of-sight of all the terminals and repeaters. (See Fig. (5.1). Thus to get from source "A" to destination "B" it may be neccessary to employ several repeaters. In intitial tests of the Packet Radio Network data flows from the user (terminal) via a series of repeaters to a central station and then from the station through a second series of repeaters to the final destination (terminal) thus providing communication between "users" that are out of direct range with each other.

When the packet-switched radio channel is operating well below capacity, it can be used to transmit voice or slow-scan video signals using the same packet-switching systems. Advantages of digitizing the voice and video signals rather than employing an analog channel for their transmission are the ability to maintain a high signal-to-noise ratio and the ease of securing the signals (as it is far easier to secure a digital signal than an analog signal). A major objective of ARPA's Network Secure Communications (NSC) project is to develop and demonstrate the feasibility of secure, good-quality, real-time, low packet rate, full-duplex digital voice communication over a packet-switched computer-communications network. The system is to operate in the field as a mobile radio system under conditions of high background noise. It should be capable of handling conferencing and therefore be able to encode a group of speakers.

We have studied the feasibility of using deltamodulators as source encoders in packet voice networks and the performance of such a system is described in the following sections.

VI. DELTA MODULATORS AS SOURCE ENCODERS IN A PACKET SWITCHED NETWORK

Current methods used for digitizing voice include Pulse Code Modulation (PCM), Adaptive Delta Modulation (ADM) and Linear Predictive Coding (LPC). In as much as voice transmission over a packet switched network requires the use of a shared channel with possible traffic congestion, bandwidth considerations are extremely important. If PCM is used to encode 2.5 KHz voice one would require a bit rate of at least 40 Kb/s to reproduce good quality voice. Assuming a packet size of 1000 bits requires that the PCM packets be transmitted at the rate of 40 packets/sec. ADM systems reproduce 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 packets/sec. In addition, since voice contains a large number of silence periods, the ADM packet rate can be further reduced by not transmitting during the silent periods of the voice. Thus, ADM is preferred to PCM.

The ARPA network is currently employing the CVSD and the LPC to digitize voice. As seen in the previous sections, subjective evaluation has shown that the SVADM is preferred to the CVSD when operating at or below 16 K b/s. Furthermore, the CVSD has a relatively narrow dynamic range compared to that of the SVADM. This becomes an important factor when a packet switched radio is operating with variable speaker levels. The SVADM is also preferred to the LPC, since the LPC is still a relatively high cost and complex system. We have studied the performance of the SVADM in a packet voice network, in terms of packet loss rate, bit rate and packet size.

CONCEPT OF PACKET LOSS:

A packet network has been shown in Fig. 5.1. We consider speech transmission from source A to destination B. At source A, the speech is encoded by an adaptive delta modulator encoder (SVADM or CVSD) and then packetized. Each packet consists of a header and information. The header includes the packet number and the destination. While source A is active, destination B should be receiving a virtually continuous stream of packets. Thus, while the i^{th} packet is being processed, destination B looks for the $(i+1)^{st}$ packet. If the $(i+1)^{st}$ packet is not available for processing after B finished 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:

(1) The $(i+1)^{st}$ packet actually arrived at B, but was rejected as <u>non-valid.</u> It is to be noted that unlike in data transmission where B requests a resending of the packet rejected, retransmission is not needed for speech transmission. A single lost packet will not degrade the quality of speech processed by the delta modulators. Also, the step-size error and the estimate error in SVADM decoder will be corrected by the error correction algorithms described in Sec. II above.

(2) B has completed processing packet i and $(i+1)^{st}$ <u>packet has not arrived</u> (i.e. it is late). The receiver B then will decide (after an appropriate waiting period) that $(i+1)^{st}$ packet is lost and starts looking for the $(i+2)^{nd}$ packet.

EFFECT OF PACKET LOSS:

Consider that the speech is encoded at 16 Kb/s and the packet size is 1 K bits. If a packet is lost, the fraction of the speech lost is $(1/16)^{th}$ of a second or approximately 60 msec. The degradation of performance due to 60 msec. of speech loss is minimal in the processed speech. 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 loss occurs in 6 seconds of speech and this too does not adversely affect the quality of the received speech.

When a packet is lost, the delta modulator decoder exhibits errors in both the signal estimate and the step size parameter as new packets are received. These errors will be corrected by the error correction algorithms described earlier. In addition, to help the receiver in its correction process, during the length of the packet loss, the receiver will compensate for the packet loss. Three different compensation algorithms have been studied.

ALGORITHM 1: FREEZE THE DECODER

In this algorithm the state of the receiver 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 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 the new packet arrives and will be eventually corrected by the error correction logic described earlier (see Sec. II).

This method of freezing has an advantage of providing a quiet period during the packet loss. The main disadvantage of a freeze out is the presence of a large step-size error which requires several sampling periods for correction. The estimate error causes only a D.C. shift of the speech waveform.

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. Generation of a steady state pattern locally at the input of the decoder would enable the receiver estimate to leak to the zero level during the period of the lost packet. However, the step size error remains unchanged. Also, generating a 1 1 0 0 1 1 0 0 · · · during a packet loss provides the sound of the quiet period instead of a freeze out. Listeners have shown preference to this algorithm over the freeze out algorithm even though, at low bit rates, oscillations of $f_s/4$ is heard.

ALGORITHM 3: GENERATE A LOCAL PERIODIC 1 0 1 0 1 0 · · · STEADY STATE PATTERN AT THE RECEIVER

In this algorithm, the receiver will locally generate 1 0 1 0 1 0 · · · pattern instead of 1 1 0 0 1 1 0 0 · · · pattern mentioned 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 D. C. shift due to an error in the estimate is basically corrected once the new packets are received. The performance of this algorithm is similar to the above two algorithms. A smaller step-size in the decoder is extremely advantageous. It will prevent the large variation of the magnitude of speech due to an error at the decoder input. This is particularly more pronounced at high error rates. In addition, at low bit rates, the oscillation at $f_s/2$ is not heard. The adaptive step size algorithm enables the decoder to correct itself within a few sampling intervals. Figure 6.1 displays the receiver estimates of the three methods during a packet loss period.

EXPERIMENTAL RESULTS

Figure 6.2 describes the system used for packet loss studies. Speech output of the tape recorder is bandlimited from 300Hz to 2500Hz and used as the inputs to SVADM encoder and the CVSD encoder. The packetizer, the depacketizer and the loss of packets were simulated using PDP-11 computer. The output bits of the depacketizer were then decoded respectively by the SVADM and the CVSD decoders and the estimates were band limited from 300 HZ to 2500 Hz and heard by using head sets. Two types of speech tapes were used.

1. a Mark Twain story read by Ed Begley.
2. a General radio conversation.

The parameters for subjective quality test are

a. Packet size P (2048, 1024, 512, 256 bits)
b. Packet loss rates r (10^{-4}, 10^{-3}, 10^{-2}, $2x10^{-1}$)
c. Sampling rate f_s (16, 9.6 Kb/s)

At the maximum input level, the performance of the packet voice system using the SVADM encoder-decoder or the CVSD encoder-decoder was found to be about the same. However at lower levels of input, there is a general degradation in the performance of the CVSD as found to be true in the test performances on the dynamic range of CVSD described in Section IV.

There was no difference in the performance regarding the intelligibility using the three receiver algorithms for packet loss. However, the peak to peak variation of the magnitude of the estimated speech due to large step size errors of method 1 and 2 is minimized in method 3. The subjective results are tabulated in Table 6.2. The fraction of speech (q) lost due to a packet loss is expressed in terms of packet size P and sampling rate f_s in Table 6.1.

From the results, we derived the following conclusions:

a. A packet loss rate up to 10^{-2} is not noticeable.
b. At packet sizes of 2048 bits, 1024 bits and f_s = 16 Kb/s, the talk spurt break of 128 msec. and 64 msec. respectively is noticed predominantly at error rates of 10^{-1} and $2x10^{-1}$. This is true because of the fact that the human ear notices any speech loss over 30 msec. duration. However, overall intelligibility was still acceptable.
c. The results show that packet switching network using delta modulation source encoders can safely operate at loss rates of 10^{-2}.

VII. SILENCE DETECTION AND SPEECH INITIATION

It is known that the speech has many quiet periods, as high as 50%. As such, the detection of silence would enable us to significantly reduce the packet transmission rate by not transmitting silent periods. One of the ways of detecting silence is to use an analog level detection technique such as used in the TASI system. By using delta modulation techniques, it is possible to detect the silence digitally rather than using the conventional analog level detection technique. The SVADM and the CVSD produce a periodic output in the steady state for a step input. This kind of output is particularly useful in detecting the silence periods.

ALGORITHM FOR SILENCE DETECTION:

All delta modulators produce a periodic output for a constant input. The SVADM produces a 1 1 0 0 1 1 0 0 · · · pattern in the steady state for a constant input. On the other hand the CVSD encoder produces a 1 0 1 0 1 0 1 0 · · · 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, it was decided that we shall observe eight consecutive bits of the encoder output to see if they have a 1 1 0 0 1 1 0 0 pattern (or any of the three other possible permutations of 1 1 0 0 1 1 0 0 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 1 1 0 0 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 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, since the output of the CVSD encoder in the steady state is 1 0 1 0 1 0 1 0 · · · ·. Here too, we remain in the silent period until the three consecutive bits of 1 1 1 or 0 0 0 are detected for speech initiation. Figure 7.1 shows the timing diagram for silence detection and speech initiation.

SILENT PACKETS:

As the transmitter forms a packet, we shall keep track of how many bits are steady state bits (bits that are generated in a silent period). This is done by using a counter. To determine whether the packet is a silent packet or not, we set up a parameter which is defined as the Threshold (T_p). The threshold T_p is a number assigned to a packet. If the ratio, ξ, of the number of silence bits to the total number of bits in a packet exceeds the threshold T_p, then, we say the packet is a silent packet; that is, 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 rate of transmission.

RECEIVER DURING SILENT PERIODS:

When the transmitter decides that a packet (silent packet) is not worthy of transmission it will not send the packet. This will cause a gap in the stream of packets received at the destination. At this point, the receiver will recognize that a silent period has begun at the source. As such, the receiver 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 · · · to be processed by the SVADM decoder during a silent period.
3) The receiver locally generates a steady state pattern of 1 0 1 0 1 0 1 0 · · · to be processed by the SVADM decoder during a silent packet.

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

REPACKING:

By repacking, we refer to the idea where the transmitter, having detected that it is currently in a silent period, will halt its packetization process until such time it detects the initation of a new speech period. Only then, will the transmitter begin the formation of a new packet. Figures 7.3 (a) and (b) illustrate the no re-packing and the repacking schemes respectively.

The advantage of repacking is that the beginning of a speech period is not lost. In Fig. 7.3 (b), suppose p_2 and p_3 are silent packets and there is no repacking, then, p_2 and p_3 are not transmitted and thus the beginning of the speech period which is contained at the end of p_3 will be lost. However, if the repacking scheme is used and p_2 ended in silence, packet p_3 is not formed until the onset of speech as shown in Fig. 7.3(b). The determination of whether the packet is silent or not will be made for p_3' and not p_3. Thus, there is less chance of losing the onset of a speech period. It was found that repacking vastly enhances the quality of the received speech.

EXPERIMENTAL RESULTS FOR PACKET VOICE CHANNEL WITH SILENCE DETECTION AND SPEECH INITIATION:

Figure 7.2 shows the test set up. It consists of a packetizer, silence detector, de-packetizer, steady state generator, which were all simulated by a PDP-11 computer for real time operation at f_s = 16 Kb/s. All other components shown in the block diagram were also real time systems.

For effcient silence detection using the output bits of the SVADM encoder requires an input noise voltage less than the minimum step size S_o (S_o = 10 mv). The peak to peak input amplitude specification of the SVADM is 8 V_{pp}. This would mean that the input speech signal to the SVADM encoder should have a SNR of approximately 55 dB. The speech source, which was used for the experiments is a tape recorder. The noise voltage at the output of the tape recorder was less than 10 mV.

The parameters varied in the experiments were

1) Packet size P (1024, 512 bits)

2) Threshold T_p (whole $T_p = \frac{\text{silence bits in a packet}}{\text{total bits in a packet}}$; (1/2, 1/4, 1/8, 1/16)

3) Sampling rate f_s (16 Kb/s, 9.6Kb/s)

EXPERIMENT 1 : No Repacking and Freeze the Receiver

The digital output bits of the encoder are assembled into packets of length P and in each packet,silence bits are detected using the silence algorithm described earlier. If there are S bits in a packet that are silence bits and if $\xi = S/P \geq T_p$,then the packet of length P is discarded as unworthy of transmission. Figure 7.3(a) showed this algorithm of discarding a silent packet. When the receiver does not receive a packet, the decoder is allowed to remain in the same state until a new packet arrives. Then the decoder processes the new packet arrived. For the tapes used in the experiments, the amount of silence present was measured. In our experiment, we then kept a count of how many packets were deemed silent and therefore not sent. From this, we were able to find what percentage of the total silence present on the tape was detected and eliminated from transmission. Figure 7.4 shows the percentage of silence detected (and thus not transmitted) as a function of the Threshold T_p. A 0 dB input level, approximately all of the silence is detected at $T_p = 1/8$. Thus, we eliminate nearly all the silence from transmission. However, as mentioned previously, silent packets might consist of some speech bits along with silence bits. Thus, when a packet is not transmitted, some of the speech bits may be lost and therefore may cause the beginning of a speech period to be clipped. The listeners were able to distinquish the breaks in speech at $T_p = 1/8$. At $T_p = 1/2$, and $T_p = 1/4$, there was no recognizable degradation in the speech processed when compared to transmitting every packet. However, at $T_p = 1/2$ and 1/4, the silent packets discarded constituted a detection of only 40% and 60% of the total silence respectively. The results of this experiment are in Table 7.1.

EXPERIMENT 2: Repacking and Generation of a Local 1 1 0 0 1 1 0 0 • • •
Pattern at the Receiver when a Packet is not Received

In order to improve the performance of the packet voice system at $T_p = 1/8$, a repacking scheme was implemented.

In this experiment silent packets are detected as in experiment 1. In addition, if a silent packet is detected, the repacking of the next packet starts at the time of speech initiation only. This repacking was carried out by the PDP/11 computer. Also, when the packet is not received, the receiver generates a 1 1 0 0 1 1 0 0 • • • pattern locally at the input of the SVADM decoder enabling the system to sound more naturally. In addition, during this period, the estimate of the speech will be leaked off to zero. However, the stepsize does not change by more than S_o during the decoding of 1 1 0 0 1 1 0 0 • • •.

The use of repacking and introduction of 1 1 0 0 1 1 0 0 • • • sequence improved the quality of the speech received particularly at lower threshold levels. Referring to Table 7.1 when f_s = 16 Kb/s, P = 1024 bits and $T_p = 1/8$, the repacking scheme is seen to be significantly better than the no-repacking scheme. The noticeable breaks in the processed voice, heard in experiment 1 were not heard. We have, thus, developed a system in which virtually all of the silence may be detected and eliminated from transmission without loss of significant quality to the received speech. This result is true even when the packet size is P = 512 bits and $T_p = 1/8$. One problem noted was that the steady state pattern of 1 1 0 0 1 1 0 0 • • • being fed to the SVADM decoder generated an output tone of fundamental frequency $f_s/4$. When f_s < 16 Kb/s, $f_s/4$ is a component less than 4 KHz, which may pass through the filters used and was heard. In the next experiment, we have been able to overcome this problem by feeding a 1 0 1 0 1 0 • • • instead of 1 1 0 0 1 1 0 0 • • • to the SVADM decoder in the absence of a packet.

EXPERIMENT 3: Repacking and Generation of a Local 101010••• Pattern when a
Packet is not Received

In this experiment, repacking technique remains the same as before. However, the use of a 1 0 1 0 1 0 • • • pattern causes the SVADM decoder to reduce its step size to a minimum value. Thus, the inband tone due to the $f_s/2$ component will not be heard in the steady state because of the amplitude of the step size. However, in this case, the estimate is not reduced to zero but remains at the level it achieved prior to the onset of silence. This condition, however, was found to be non-critical since the leaky integrator will eventually correct any difference between transmitter and receiver estimates.

The subjective evaluation showed that this scheme performed with approximately the same quality as the scheme which generated a 1 1 0 0 1 1 0 0 ··· pattern with respect to speech intelligibility. The only difference and improvement was the elimination of the tone generated at $f_s/4$ which was heard at the decoder output when operated at 16 Kb/s and at 9.6 Kb/s under the 1 1 0 0 1 1 0 0 · · · algorithm.

In Table 7.1, we have tabulated a comparison of the subjective experimental results for the no-repacking and repacking experiments performed. The results of experiment 2 and 3 were combined since there was no difference in speech intelligibility. In Fig. 7.4 we show the percentage of silence as a function of threshold T_p. We see from the graph that we detected nearly all of the silence at $T_p = 1/8$ and the subjective evaluation of the system showed that at $T_p = 1/8$, with repacking the speech is reconstructed with little degradation.

VIII CONCLUSIONS

From the tests we have performed, several important conclusions can be derived.

The SVADM offers a 10-15 dB higher dynamic range over the CVSD. Both the delta modulators offer good quality speech processing up to 10^{-2} error rates at 0 dB input level. The subjective evaluation showed that the dynamic range reduces as the error rates increase for both the delta modulators. The SVADM offers higher dynamic range even at high error rates over the CVSD.

The use of delta modulation as a source encoding scheme has been shown to be a viable and efficient technique for use in a packet voice system. We have established that, unlike for data packets, there is no need for a receiver to request the retransmission of an invalid packet when speech is processed. The results show that packet voice system using delta modulators can safely operate up to a loss rate of 10^{-2}.

Silence detection has been accomplished digitally by using the periodic steady state output of the delta modulator encoder. It has been established that by not transmitting packets during silence periods of speech, the packet voice network can be built more efficiently, since there will be a decrease in the overall packet transmission rate without loss of speech quality.

As such, further research is being done at this time to refine our methods of using the SVADM as a source encoder for packet voice. Results of this ongoing research will be reported in later reports to ARPA.

REFERENCES AND BIBLIOGRAPHY

1) 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.

2) N. C. Seiler, R. Flower, J. Friedman, " A monolithic implementation of a CVSD algorithm ", ICC conf. Hartford, Connecticut, 1977.

3) T. Apelewicz, D. L. Schilling, " A programmable micro computer for real time speech processing ", ICC conf., Chicago, Ill., 1978.

4) J. Locicero, " Ph. D. dessertation at the City College of New York ".

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

6) A. Tomozova, H. Kaneko, " Companded delta modulation for telephone transmission ", IEEE Trans., COM-TECH, vol 16, Feb 1968.

7) S. J. Brolin, J. M. Brown, " Companded delta modulation for telephone", IEEE Trans. vol 16, Feb 1968.

8) M. R. Winkler, " High information delta modulation ", IEEE Int. con. rec., pt 1., pp 285-290, 1965.

9) J. A. Greefkes, F. De Jager, " Continous delta modulation ", Phillips res. rep., vol 23 pp 233-246, April 1968.

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

11) N. S. Jayant " Adaptive delta modulation ", B. S. T. J., vol 49, pp321- 327, Mar., 1970.

12) T. E. Eger, S. J. Campanella, " Study of an adaptive PCM voice communication system ", papaer 13 F, proc. of 1974, ICC, Mineapolis.

13) 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.

14) Motorola Incorporated, " Semiconductor data library ", vol 6/ series B, paage 5-134.

15) C. V. Chakravarthy, M. N. Faruqui, " A multidigit Adaptive delta modulation system", IEEE Trans. Communications, vol COM-24, pp 931-935, Aug. 1976.

16) R. E. Crchiere, D. J. Goodman, L. R. Rabiner, M. R. Sambur, "Tandem connections of wide band and narrow band speech communication systems , part 1- Narrow band to

wide band link ", B.S.T.J., vol 56, no.9, Nov 1977.

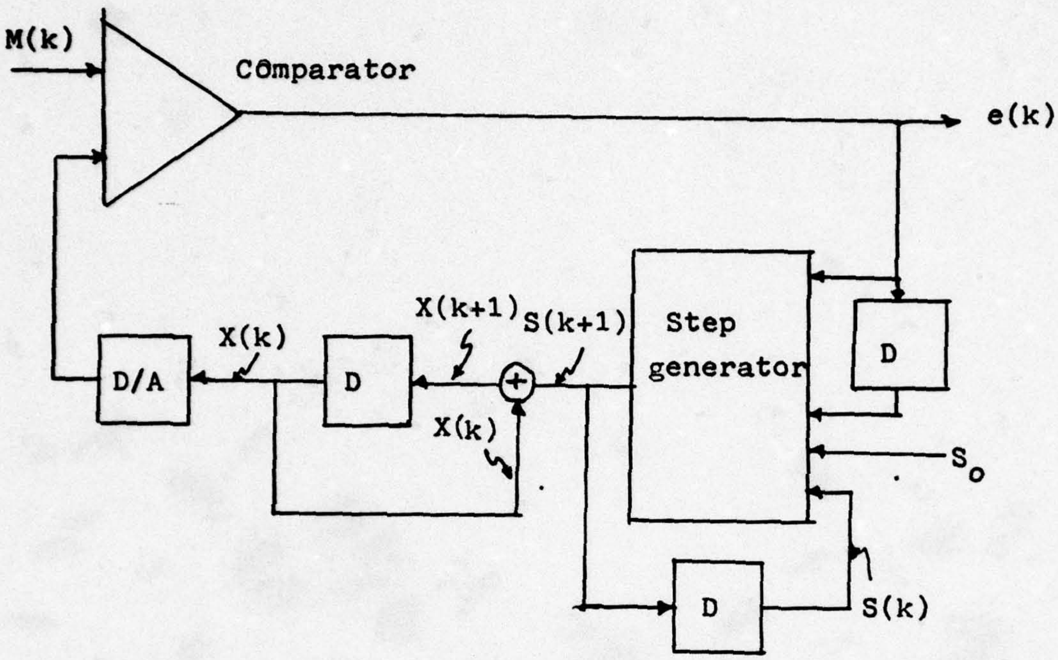
17) N. Abramson, " The ALOHA system- another alternative for computer communications ", Comput. conf., AFIPS conf. proc., vol. 37, pp 281-285, 1970

18) ----, " Packet switching with satellites,", Nat. comput. conf., AFIPS conf. proc., vol. 42, 1973, pp 695-702,

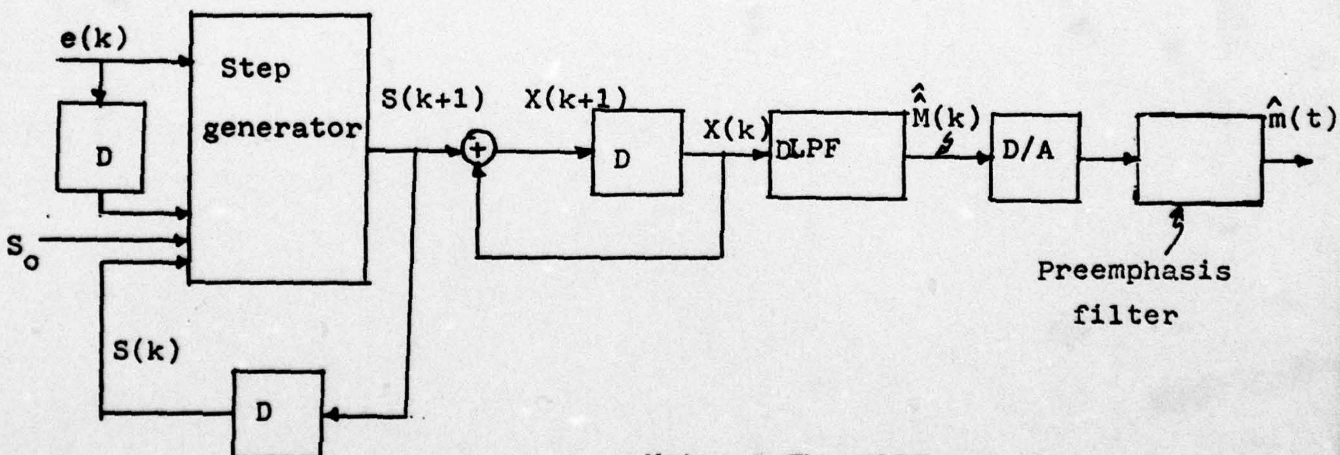
19) L. G. Roberts, " Dynamic allocation of satellite capacity through packet reservation", AFIPS conf. proc., vol 42, pp 711-716, 1973.

20) L. Kleinrock, F. Tobagi, " Packet switching in radio channels, Part I- carrier sense multiple acces modes and their throughput delay characteristics ", IEEE Trans. COM-TECH vol 23, pp 1400-1416, Dec. 1975.

21) ----, " 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.19775.



Encoder



Decoder

<u>Note</u>: The DLPF and the preemphasis filters are used when f_s = 10 Kb/s and below.

Fig.2.1 Block diagram of SVADM

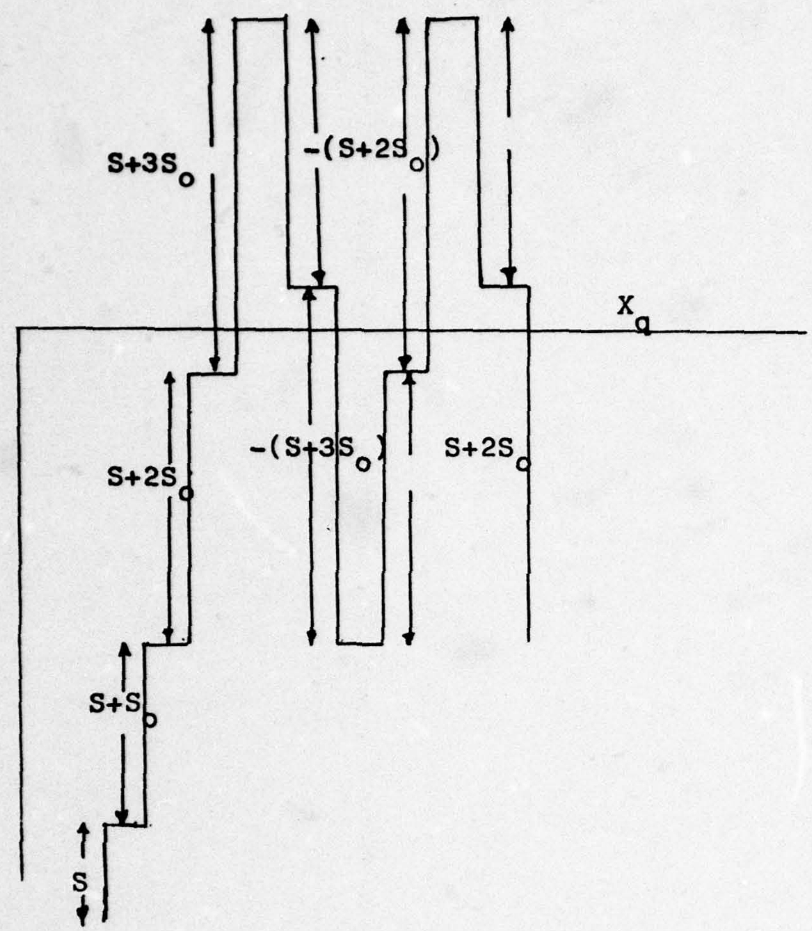


Fig. 2.2 Response of the SVADM to a STEP INPUT

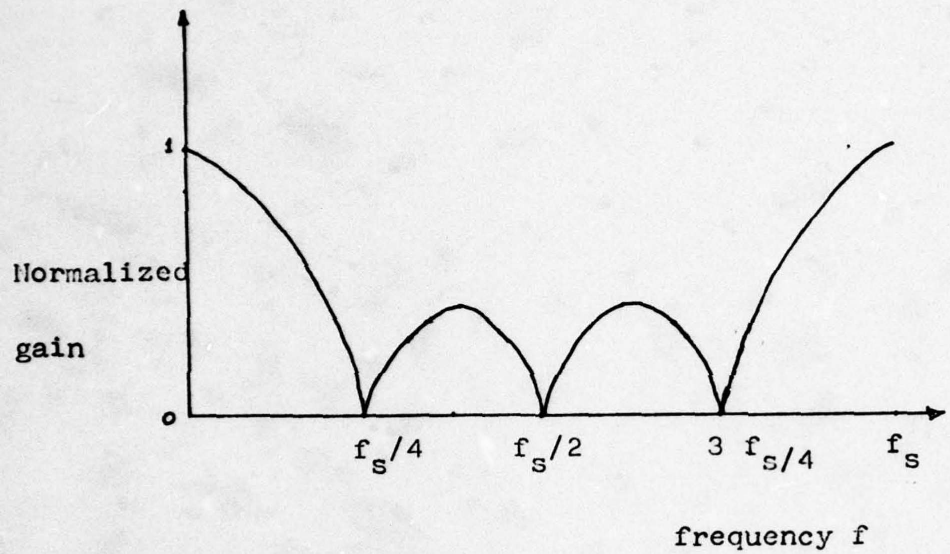


Fig. 2.3 Frequency response of the DLPF

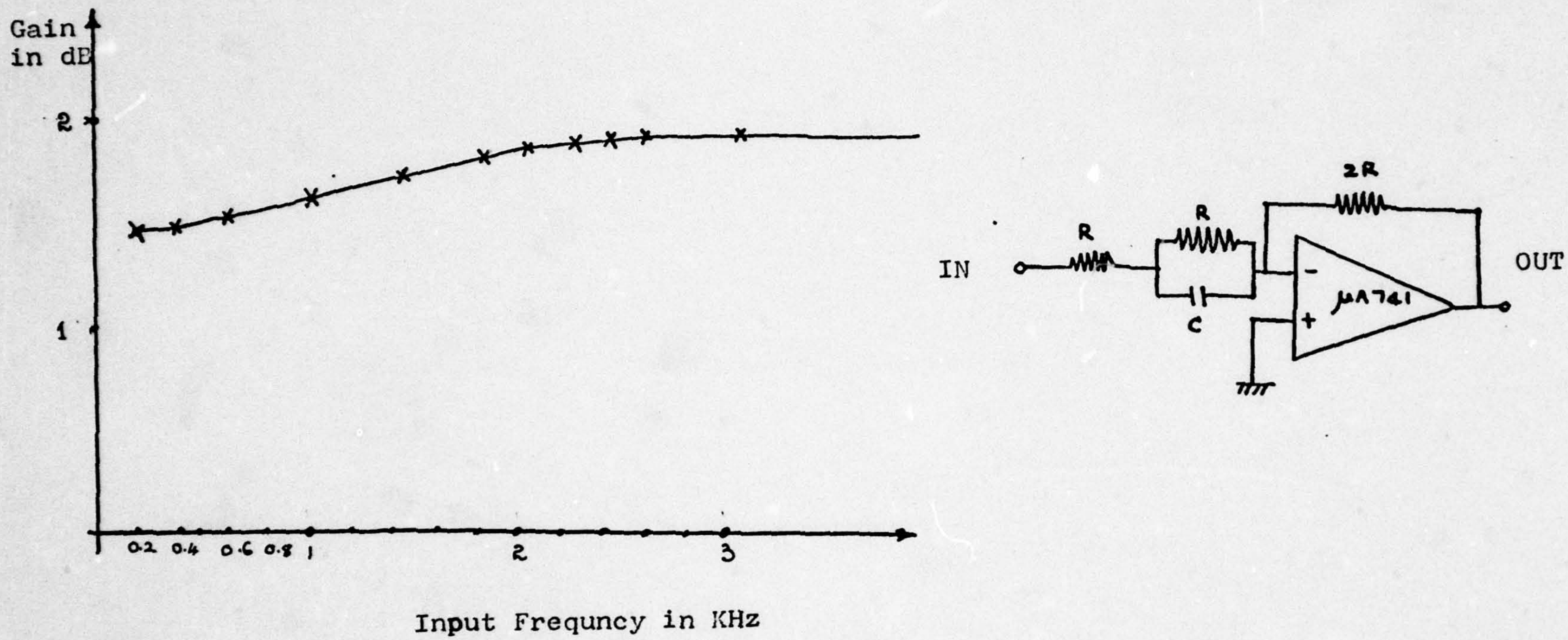


Fig.2.4 Preemphasis Filter and its Frequency Response

Sampling frequency 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)

Table 2.1. Leak factor

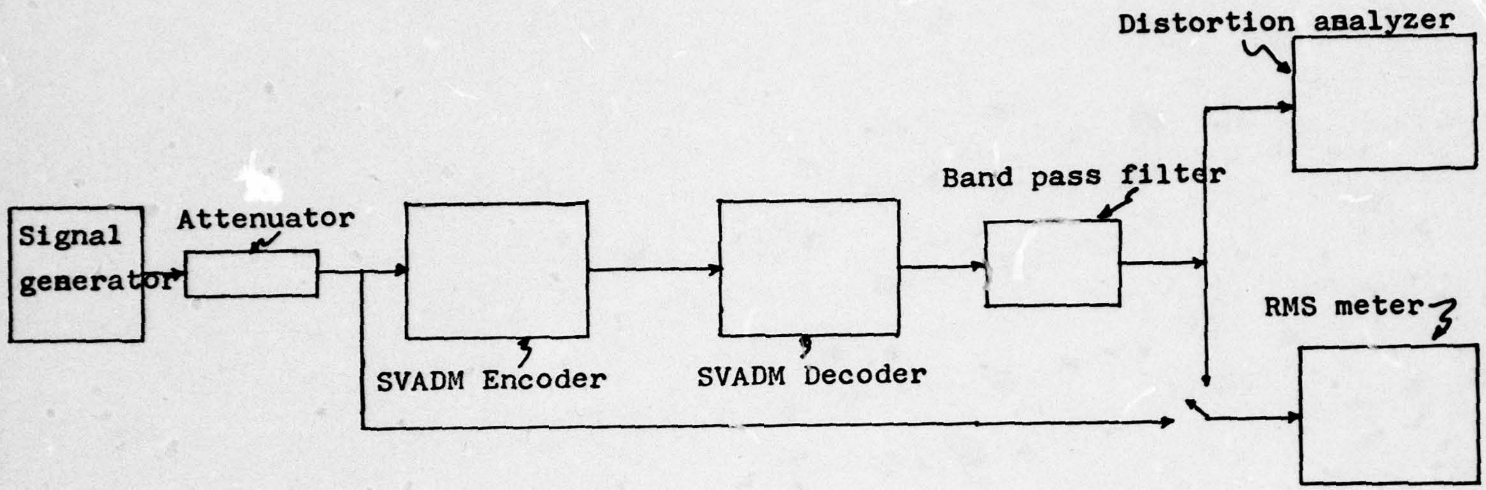


Fig.2.5 Test set up to evaluate the SVADM

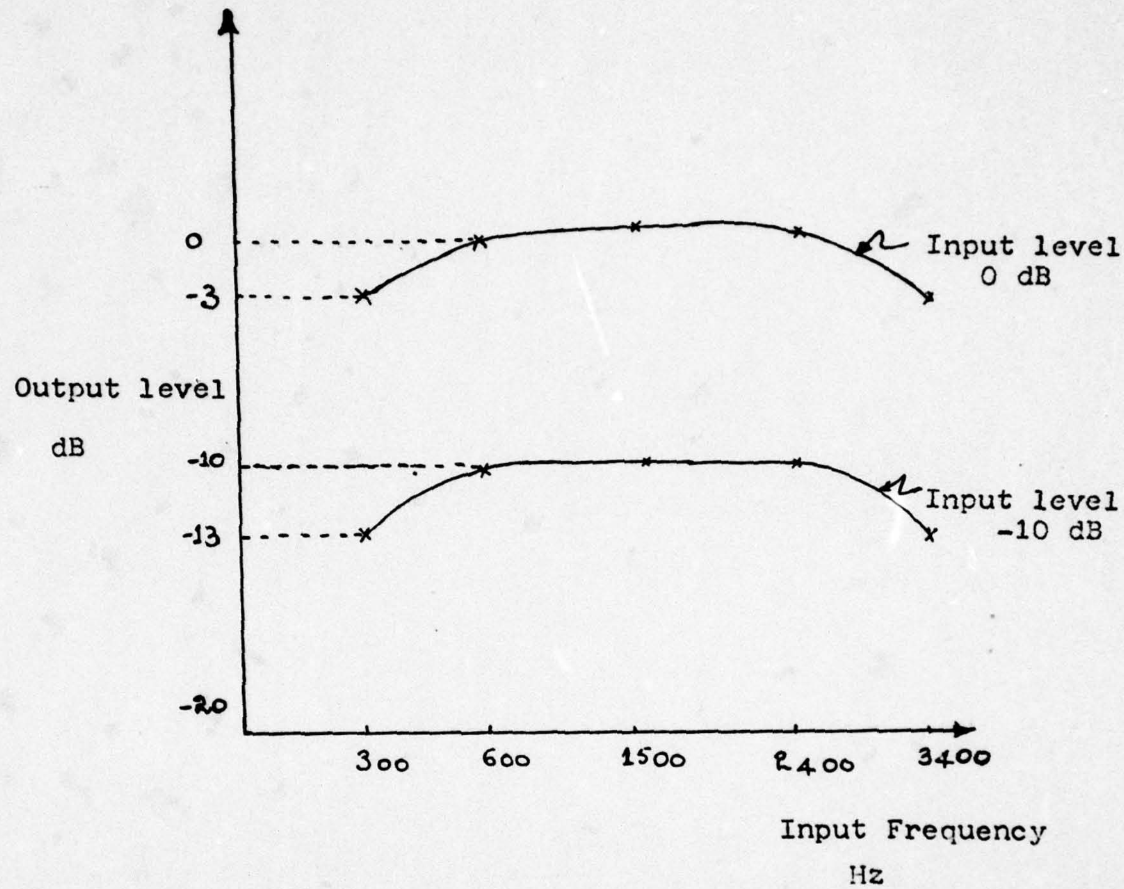


Fig.2.6 Bandwidth Test

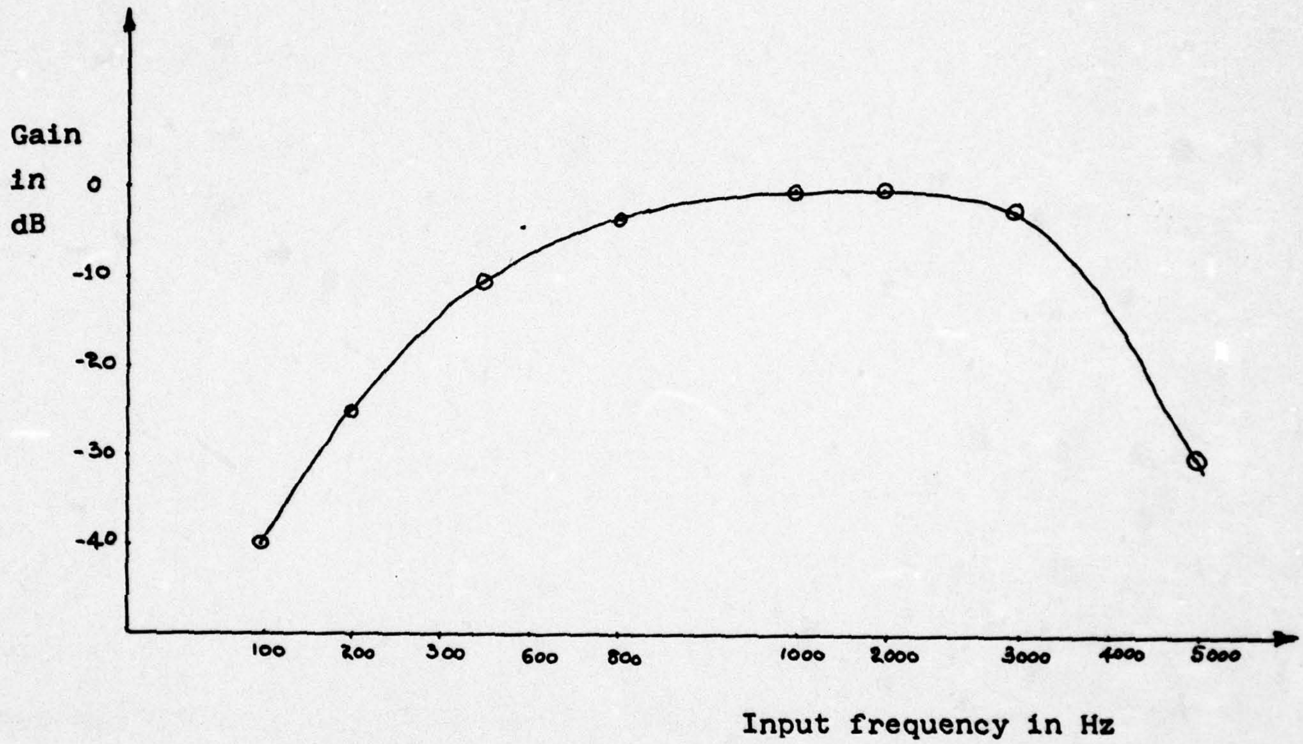


Fig.2.7 C Message weighted filter

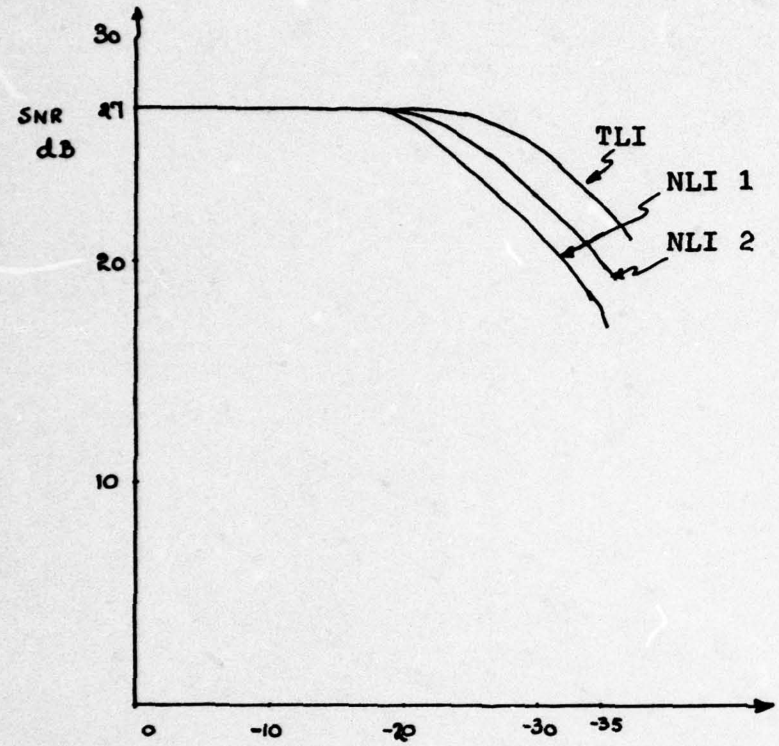


Fig.2.8 (a) Dynamic range of the SVADMs

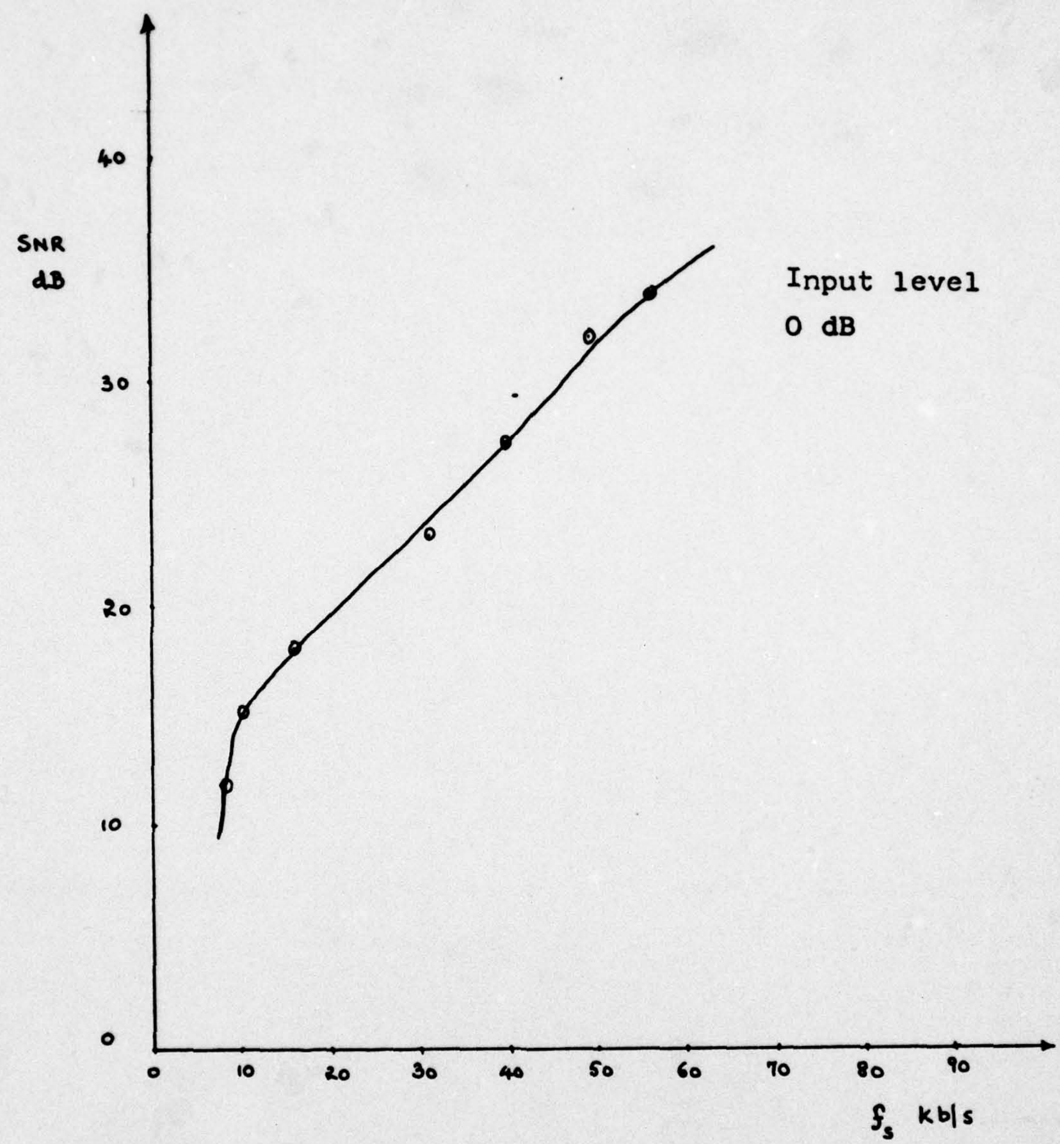


Fig.2.8 (b) SNR vs f_s

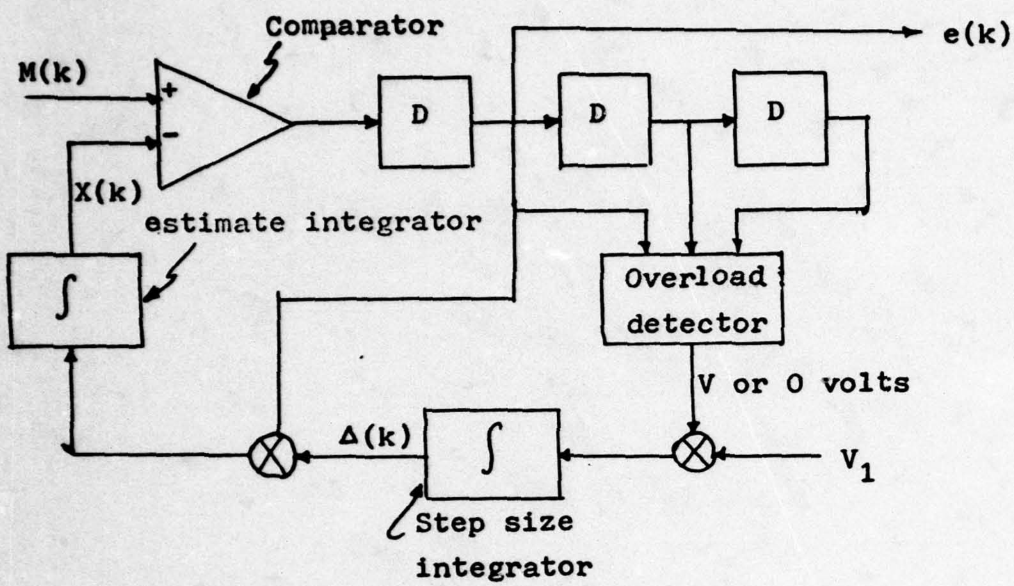


Fig.3.1 Block diagram of the CVSD

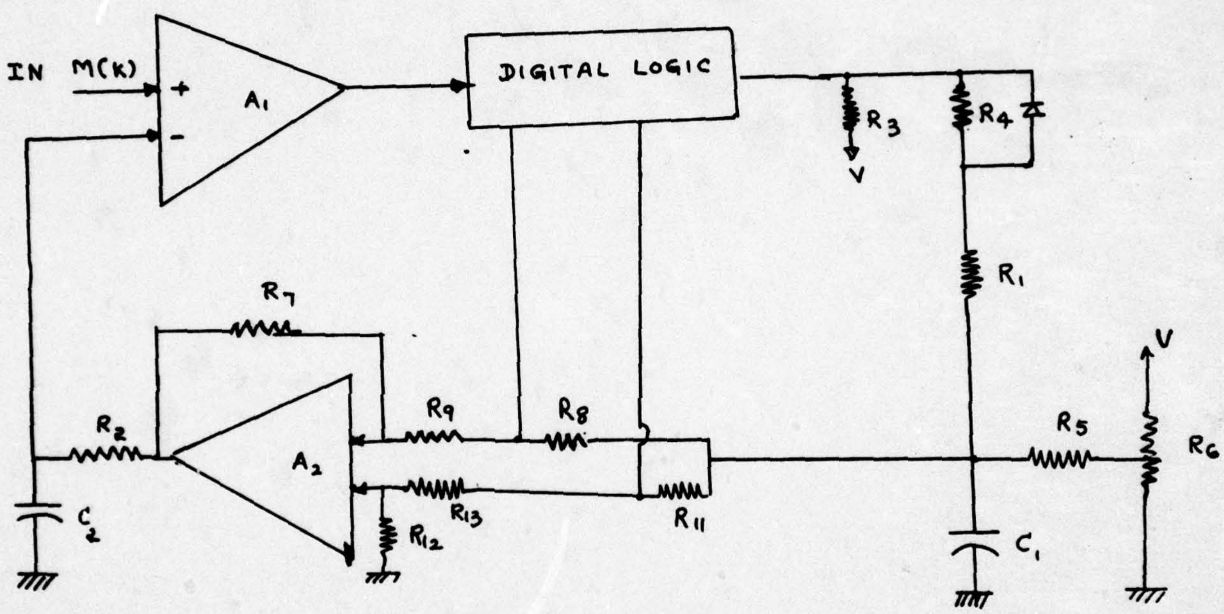


Fig. 3.2 Circuit implementation of the CVSD

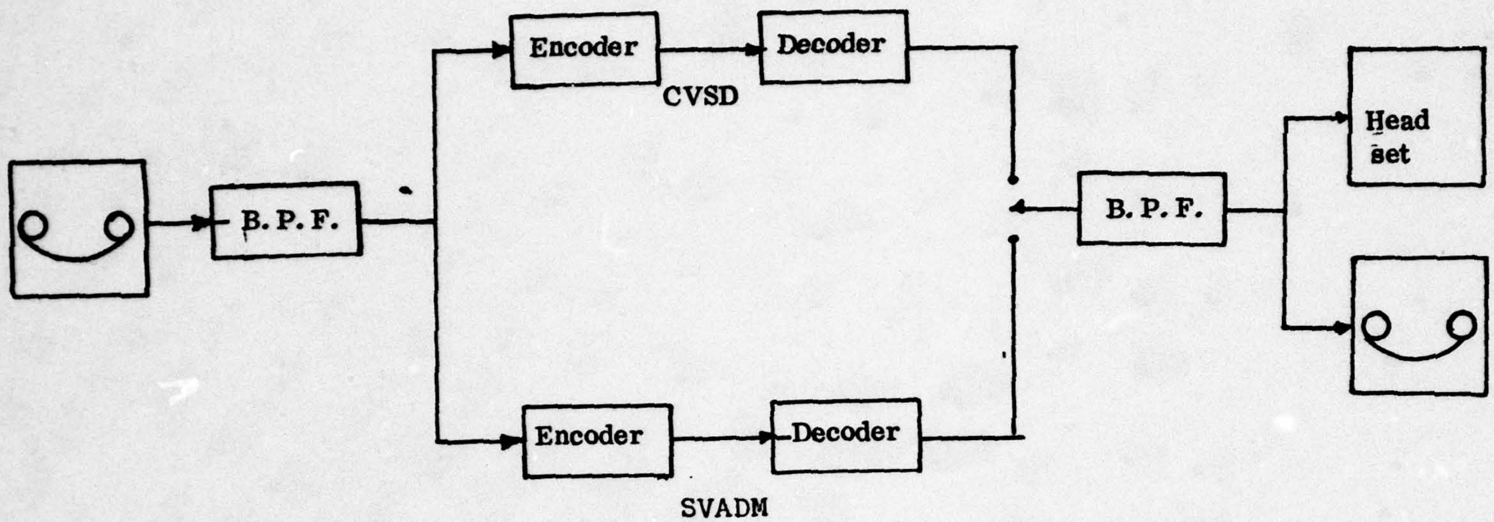


Fig. 4.1 TEST SET UP FOR SUBJECTIVE COMPARISON OF THE CVSD AND THE SVADM

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

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 preffered
8	Noisy	less noisy	SVADM preferred

TABLE 4.1 Subjective comparison of the CVSD and the SVADM as a function of bit rate

f_s Kb/s	Input level dB	Motorola CVSD Subjective	SVADM Comparison	
32	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	No difference
	-10	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	No difference
	-20	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	No difference Harris CVSD loses Intelligibility
	-30	a) Buzzy b) Breaking voice c) Understandable	a) Intelligible b) Acceptable	SVADM preferred
	-40	a) Not intelligible	a) Buzzy b) Understandable	SVADM preferred
16	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	No difference
	-10	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	SVADM preferred
	-20	a) Buzzy b) Intelligible c) Acceptable	a) Intelligible b) Acceptable	SVADM preferred
	-30	a) Not intelligible	a) Noisy b) Intelligible	SVADM preferred
	-40	a) Not intelligible	a) Not intelligible	

TABLE 4.2 Comparison of dynamic ranges of CVSD and SVADM

f_s Kb/s Kb/s	Input level dB	Motorola CVSD Subjective	SVADM comparison	
9.6	0	a) Intelligible b) Acceptable	a) Intellgible b) Acceptable	No difference Granularity due to low f_s exists
	-10	a) Not intelligible	a) Intelligible	SVADM preferred
	-20	a) Not intelligible	a) Intelligible b) Noisy	SVADM preferred
	-30, -40	a) Not intelligible	a) Not intelligible	

TABLE 4.2. Contd...

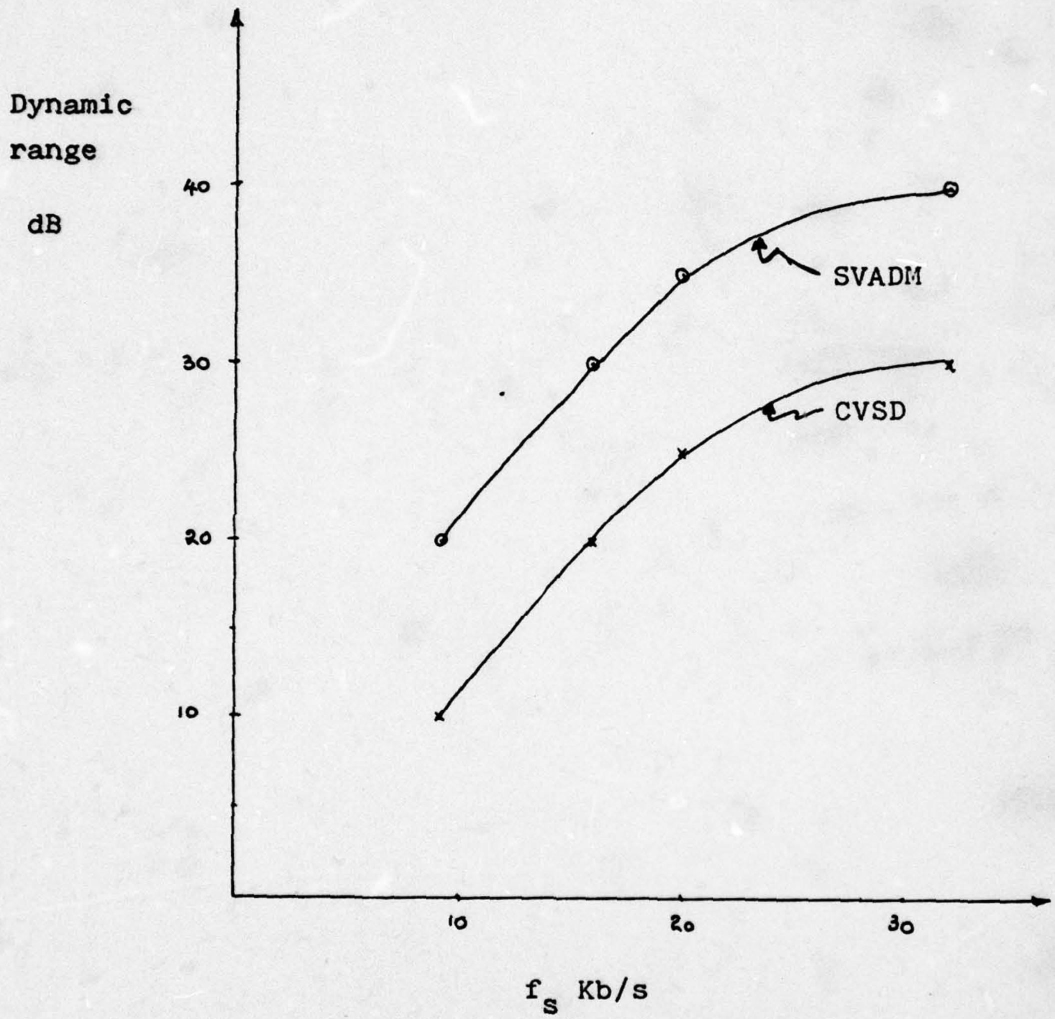


Fig. 4.2 Dynamic range as a function of bit rate

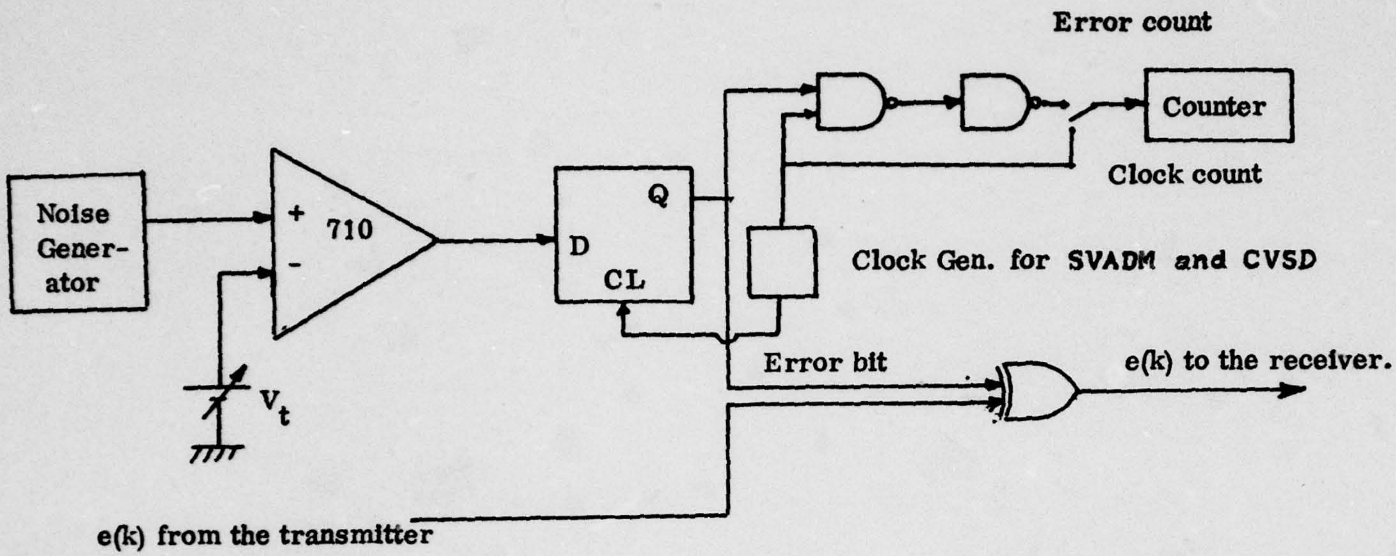


FIG 4.3 GENERATION OF SINGLE RANDOM ERRORS.

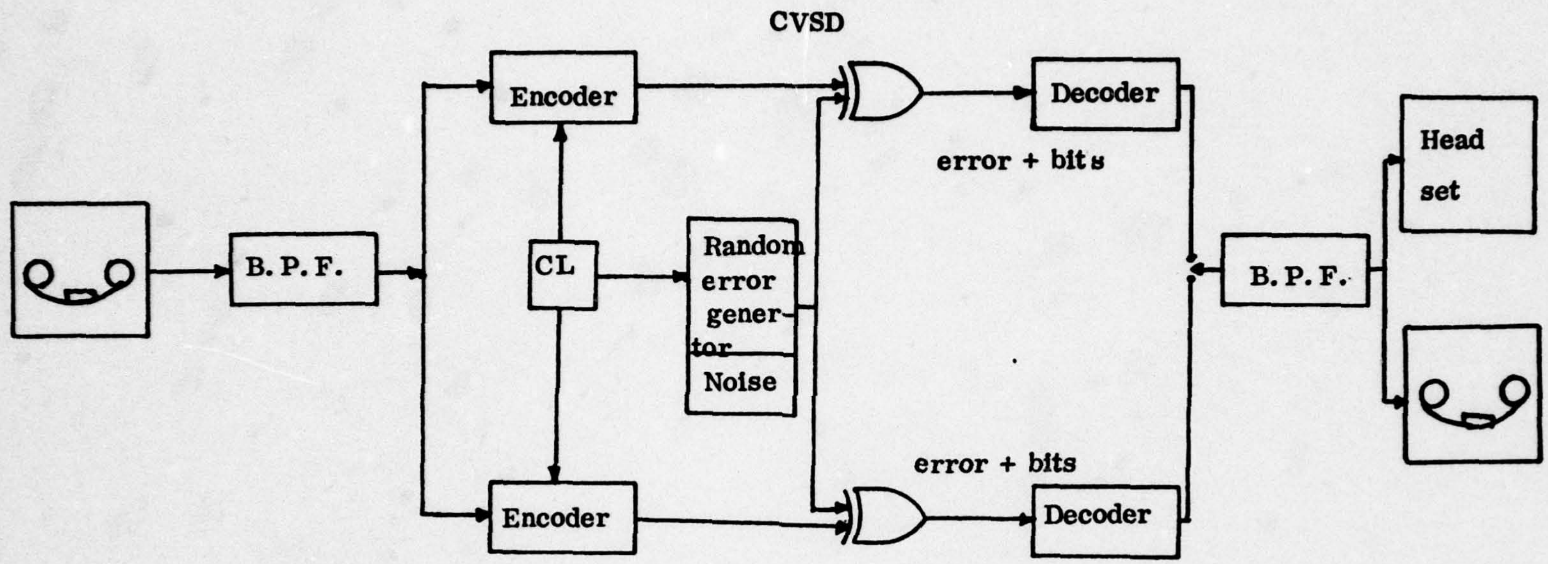


FIG. 4.4 TEST SET UP FOR COMPARISON OF THE CVSD AND THE SVADM IN THE PRESENCE OF ERRORS.

f_s Kb/s	Input llevel dB	Error rate	Motorola CVSD	SVADM	
32	0	10^{-4}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-3}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-2}	a) Intelligible b) Back ground noise with smearing	a) Intelligible b) Back ground noise	No preference
		10^{-1}	a) Intelligible b) More noise	a) Intelligible More noise	CVSD preferred
32	-20	10^{-4}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-3}	a) Barely intelligible	a) Intelligible	SVADM preferred
		10^{-2}	a) Barely intelligible	a) Intelligible	SVADM preferred
		10^{-1}	a) Not intelligible	a) Not intelligible	-
16	0	10^{-4}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-3}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-2}	a) Intelligible	a) Intelligible	SVADM preferred
		10^{-1}	a) Not intelligible	a) Not intelligible	-

TABLE 4.3 Subjective comparison of the CVSD and the SVADM at different error rates

f_s Kb/s	Input level dB	Error rate	Motorola CVSD	SVADM	
16	-20	10^{-4}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-3}	a) Barely intelligible b) Not acceptable	a) Intelligible b) Acceptable	SVADM preferred
		10^{-2}	a) Barely intelligible b) Not acceptable c) Heavy back ground noise	a) Intelligible b) fluttering noise	SVADM preferred
		10^{-1}	a) Not intelligible	a) Not intelligible	-
9.6	0	10^{-4}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-3}	a) Same as at no errors	a) Same as at no errors	No preference
		10^{-2}	a) Intelligible b) Noisy	a) Intelligible b) Noisy	No preference
		10^{-1}	a) Not intelligible	a) Not intelligible	-
9.6	-20	10^{-4}	a) Same as at no errors	a) same as at no errors	No preference
		10^{-3}	a) Not intelligible b) Words clipped	a) Intelligible	SVADM preferred
		$10^{-2}, 10^{-1}$	a) Not intelligible	a) Not intelligible	-

TABLE 4.3. Contd...

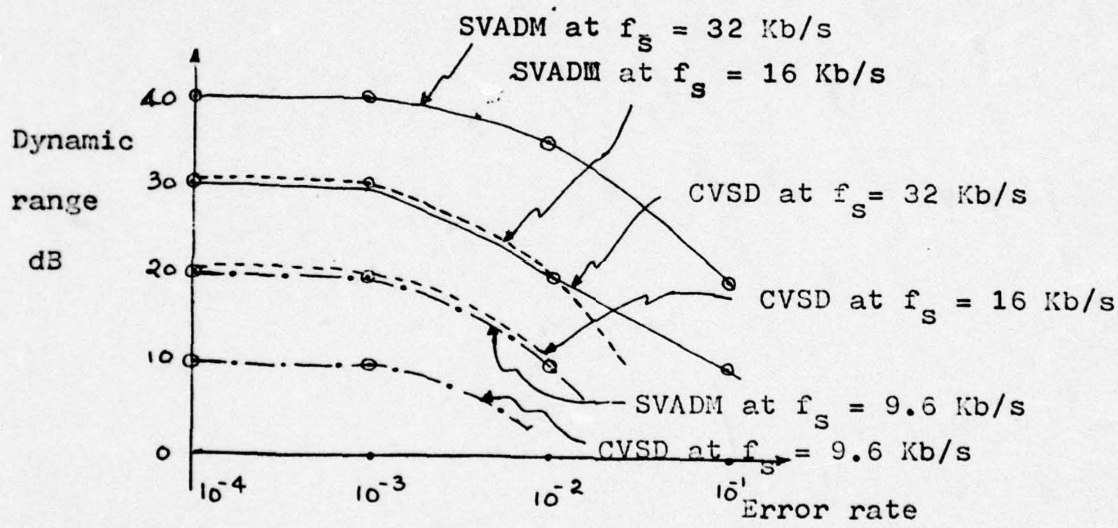


Fig. 4.5 Dynamic range vs error rate

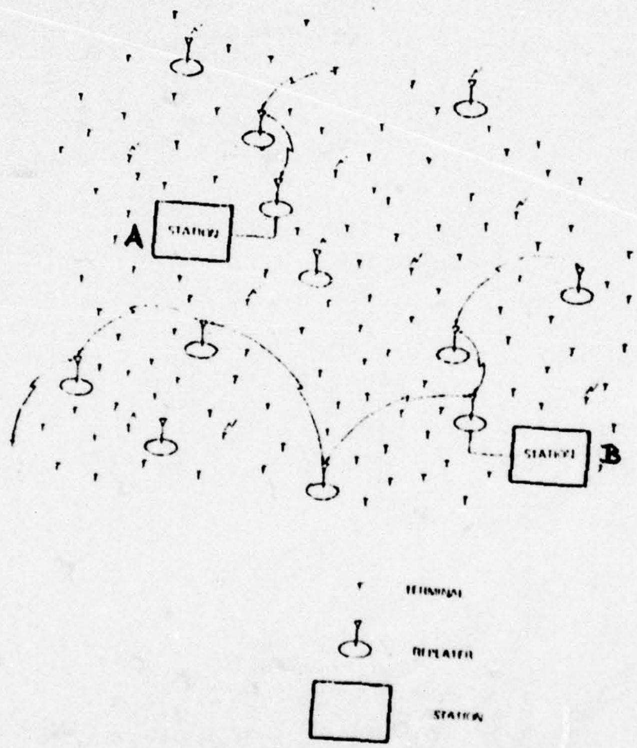


Fig. 5.1 A Packet Radio Network

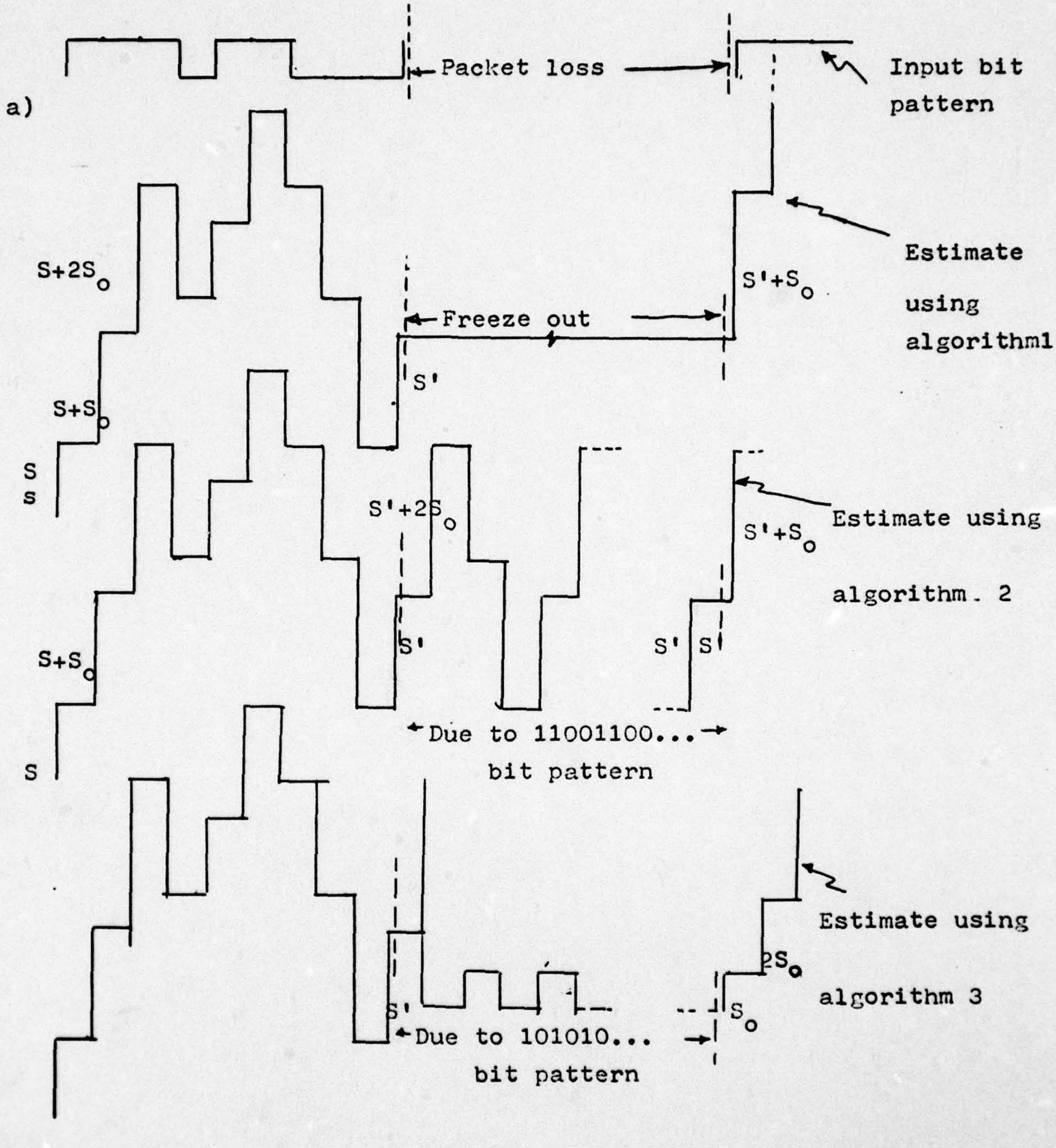
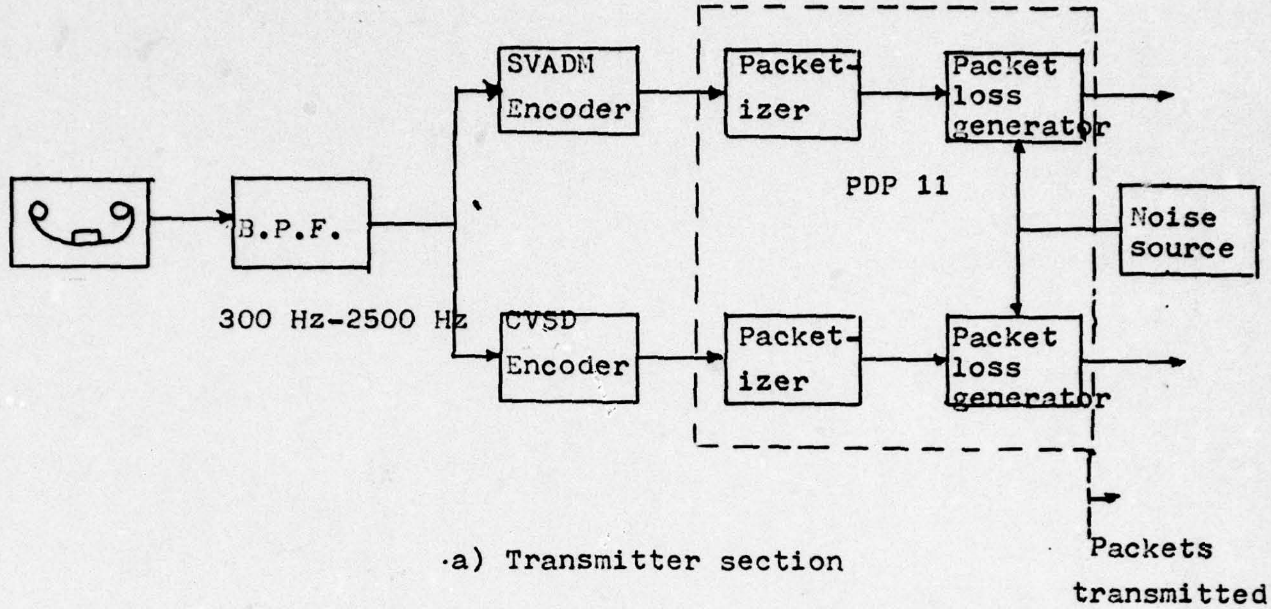
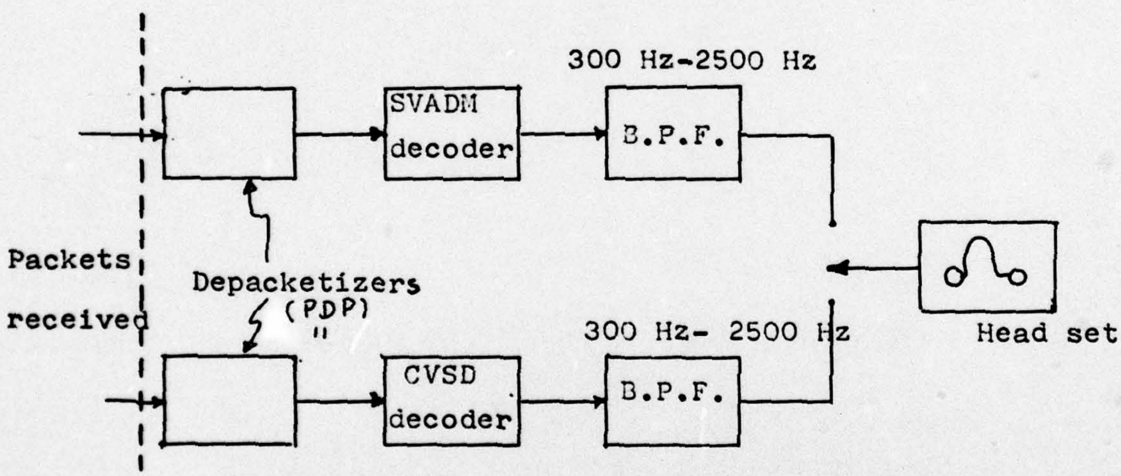


Fig. 6.1 Estimates using algorithm 1,2,3.



a) Transmitter section



b) Receiver section

Fig. 6.2 Test set up for packet loss studies

Packet size P bits	f_s Kb/s	Amount of speech lost due to a packet loss q msec.
2048	16	128
1024	16	64
512	16	32
256	16	16
2048	9.6	212
1024	9.6	106
512	9.6	53
256	9.6	26.5

TABLE 6.1. Speech lost due to a packet loss

Packet size P bits	f_s Kb/s	loss rate r	CVSD performance (subjective)	SVADM performance (subjective)	
2048	16	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	performances are similar to when r=0
		10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Intelligible b) Not acceptable	a) Intelligible b) Not acceptable	Breaks in speech are noticed
	9.6	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Granularity is heard
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0

TABLE 6.2 Subjective comparison of the CVSD and the SVADM in terms of P, f_s, and r.

Packet size P bits	f_s Kb/s	loss rate r	CVSD Performance (subjective)	SVADM Performance (subjective)	
2048	9.6	10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Not intel-ligible	a) Not intel-ligible	Breaks in speech are noticed
1024	16	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Intelligible b) Not acceptable	a) Intelligible b) Not acceptable	Breaks in speech are noticed

TABLE 6.2 Contd...

Packet size P bits	f_s Kb/s	loss rate r	CVSD Performance (subjective)	SVADM Performance (subjective)	
1024	9.6	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Granularity is heard
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Not intel-ligible	a) Not intel-ligible	Breaks in speech are noticed
512	16	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0

TABLE 6.2 Contd...

Packet size size P bits	f_s Kb/s	loss rate r	CVSD Performance (subjective)	SVADM Performance (subjective)	
512	16	10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Intelligible b) Not acceptable	a) Intelligible b) Not acceptable	Breaks in speech are noticed
	9.6	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Granularity is heard
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Not intel-ligible	a) Not intel-ligible	
256	16	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0

TABLE 6.2 Contd...

Packet size P bits	f_s Kb/s	loss rate r	CVSD Performance (subjective)	SVADM Performance (subjective)	
256	16	10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Intelligible b) Not acceptable	a) Intelligible b) Not acceptable	Breaks in speech are noticed
	9.6	0	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Granularity is heard
		10^{-4}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	performances are similar to when r=0
		10^{-3}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-2}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Performances are similar to when r=0
		10^{-1}	a) Intelligible b) Acceptable	a) Intelligible b) Acceptable	Breaks in speech are noticed
		$2(10^{-1})$	a) Not intel-ligible	a) Not intel-ligible	Breaks in speech are noticed

TABLE 6.2 Contd...

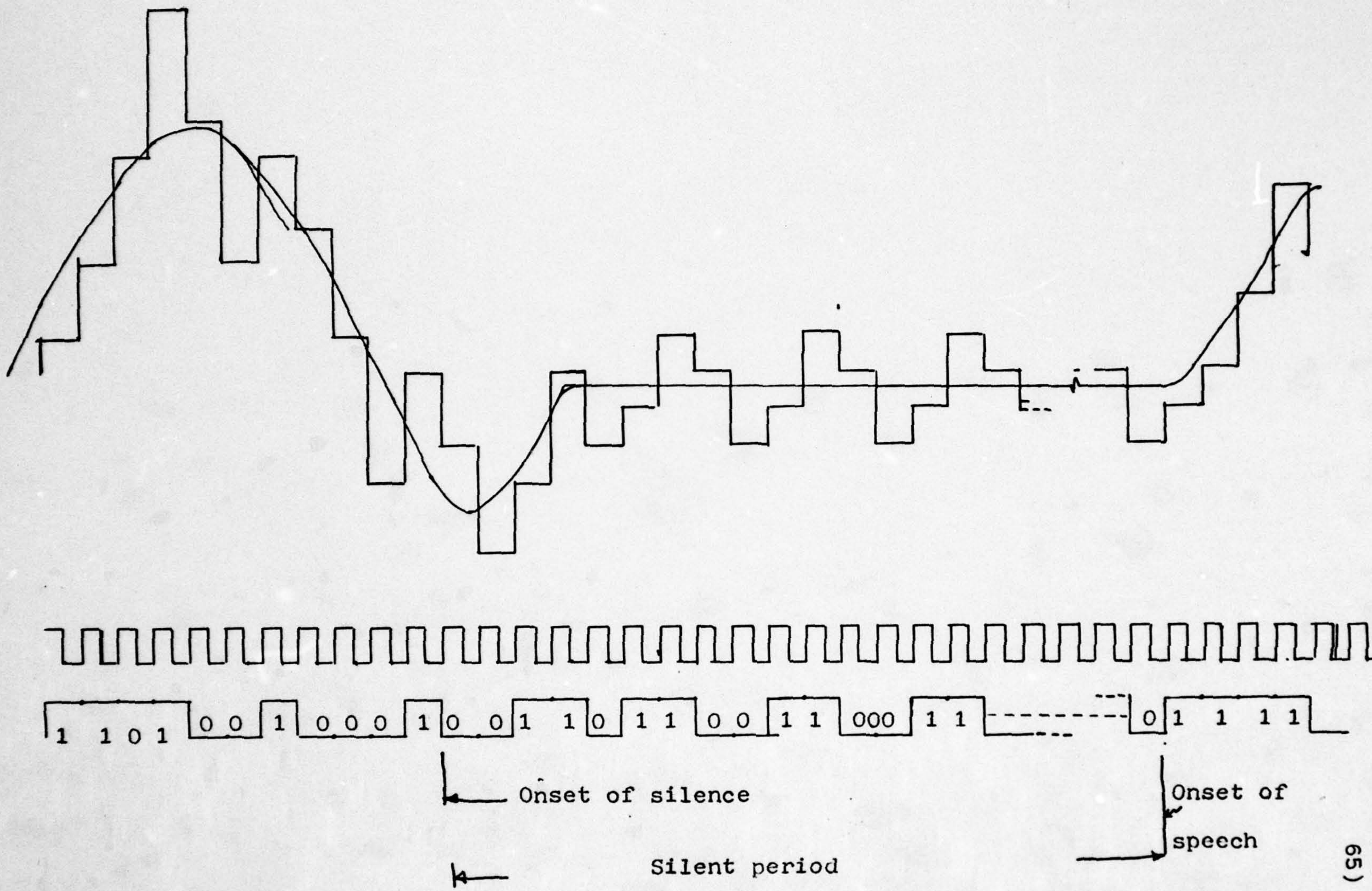
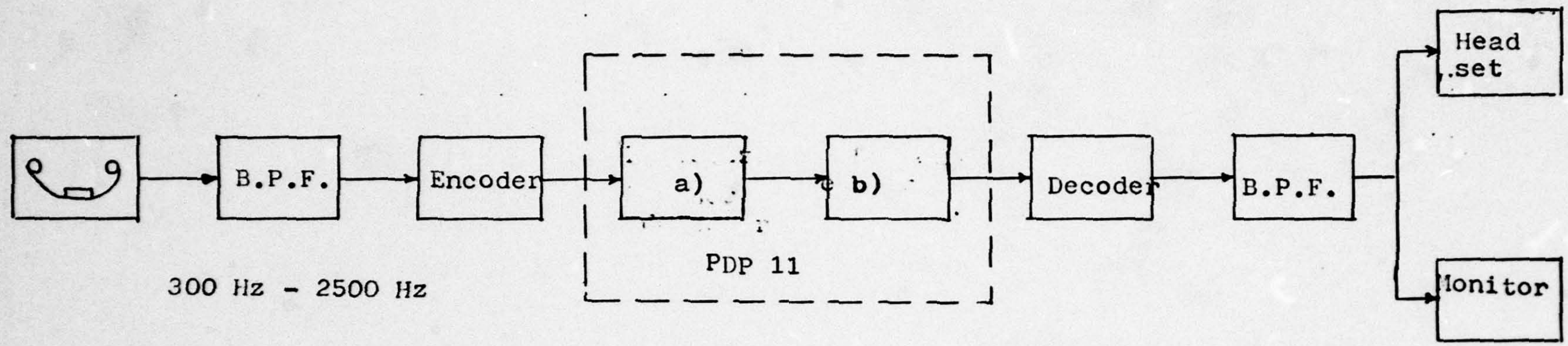


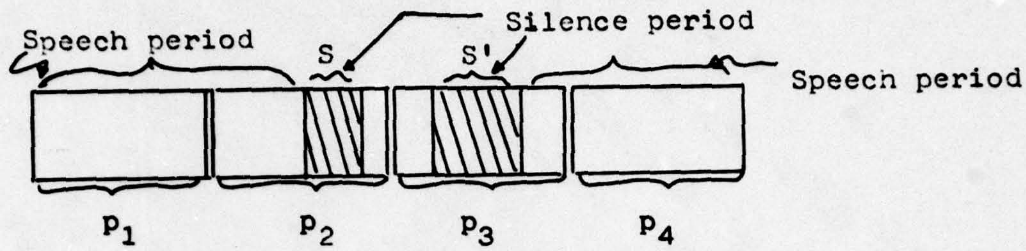
Fig 7.1 Timing diagram for the onset of speech and the onset of silence



a) Packetizer and silence detector

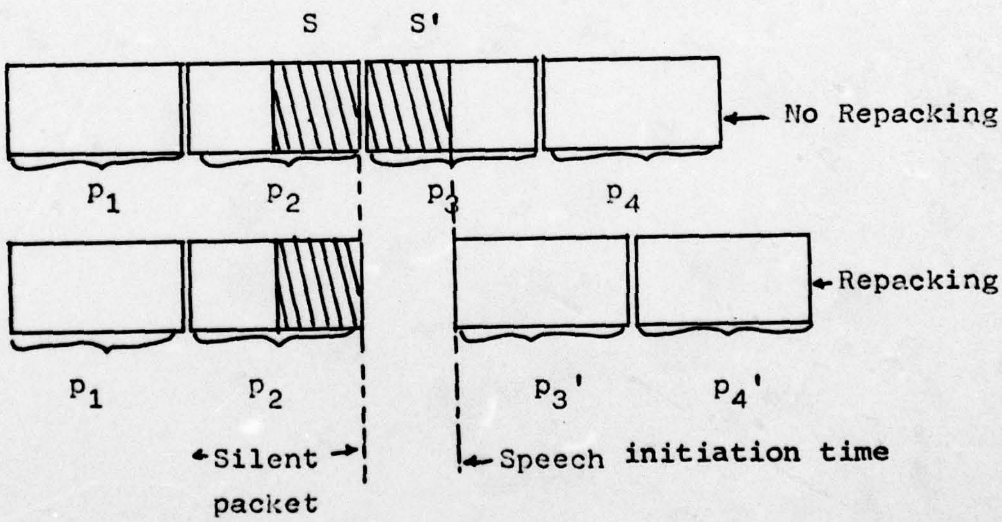
b) depacketizer and steady state generator

Fig. 7.2. Silence detector and speech onset detector



Size of p_1, p_2, p_3 and p_4 = P; S and S' represent amount of silence detected and p_2 and p_3 are silent packets and will not be transmitted if $S/P \geq T_p$ and $S'/P \geq T_p$ respectively. No repacking is possible in this situation to save the information of speech in p_2 and p_3.

(a)



Repacking will detect the speech initiation in packet p_3.

(b)

Fig. 7.3 Comparison of No-Repacking and Repacking Schemes

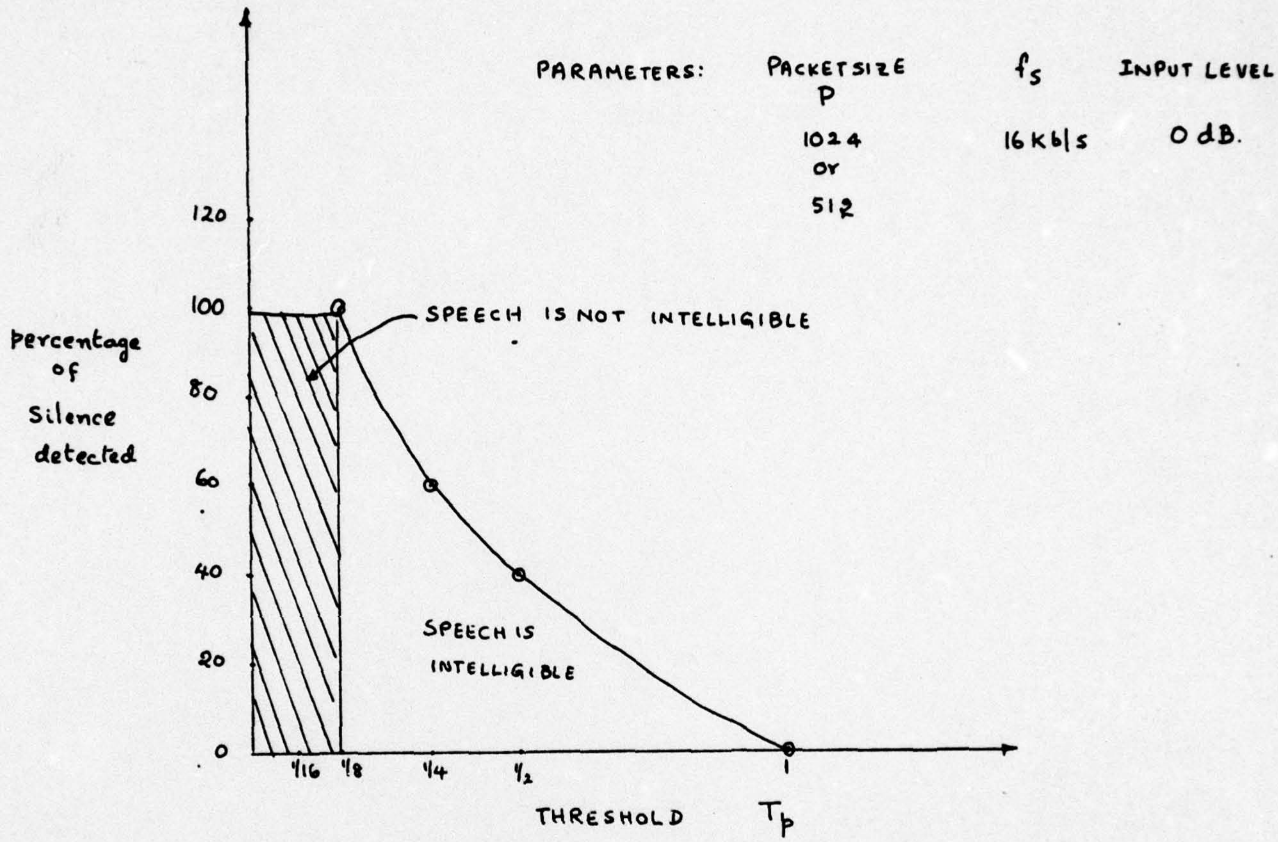


Fig. 7.4 Percentage of silence detected vs T_p

Input level dB	Sampling rate f_s Kb/s	Packet size P bits	Thresh-old T_p ratio	Processed speech quality No Repacking	Repacking	
0	16	1024	1/2	a) acceptable b) intelligible	a) acceptable b) intelligible	No difference
			1/4	a) acceptable b) intelligible	a) acceptable b) intelligible	No difference
			1/8	a) not accept-able b) speech is clipped	a) acceptable b) breaks are not noticeable c) intelligible	Repacking is significantly better than no repacking
			1/16	a) not accept-able b) not intelli-gible	a) not accept-able b) not intelli-gible	Words are missing

TABLE 7.1 Subjective comparison of No-Repacking and Repacking schemes using delta modulator as a source in packet voice transmission .

Input level dB	Sampling rate f_s Kb/s	Packet size P bits	Threshold T_p ratio	Processed speech quality		
				No Repacking	Repacking	
-10	16	1024	1/2	a) acceptable b) intelligible	a) acceptable b) intelligible	No difference
			1/4	a) acceptable b) intelligible	a) acceptable b) intelligible	No difference
			1/8	a) not acceptable b) not intelligible	a) acceptable b) intelligible with noticeable breaks	Repacking preferred

TABLE 7.1. Contd...

Input level OdB	Sampling rate f_s Kb/s	Packet size P bits	Threshold T_p ratio	Processed speech quality: No Repacking	Processed speech quality: Repacking	
0	9.6	1024	1/2	a) acceptable- b) intelligible	a) acceptable b) intelligible	No difference
			1/4	a) not acceptable b) not intelligible	a) not accepta able b) not intelligible	Words are clipped
-10	9.6	1024	1/2	a) acceptable b) breaks are noticeable c) intelligible	a) acceptable b) breaks are less than than that of Repacking c) intelligible	Repacking preferred
			1/4	a) not acceptable b) not intelligible	a) not acceptable b) not intelligible	

TABLE 7.1. Contd...