DESIGN CONSIDERATIONS FOR THE IMPLEMENTATION OF A MOBILE PACKET DATA NETWORK

1./ **DESIGN CONSIDERATIONS FOR THE IMPLEMENTATION**

OF A MOBILE PACKET DATA NETWORK /

Submitted according to the terms
of Contract OSU81-00148

(RN: 36100-1-0151)
(FC: 4113-16350-2202)

Prepared by: Dr. J.S. Riordon (Principal Investigator)
Dr. S.A. Mahmoud
Dr. S.E. Aidarous
Dr. D. Falconer
M. El-Tanany
G. Chan

Carleton University
March 1982

ACKNOWLEDGEMENT

The authors wish to acknowledge the valuable encouragement and guidance provided by Dr. John DaSilva who acted as the contract technical authority on behalf of the Department of Communications. His technical advice has contributed substantially to the research work reported here.

The authors are also grateful to Dr. H.M. Hafez, Department of Systems and Computer Engineering, Carleton University, for many constructive suggestions during the course of this research.

ABSTRACT

This report presents the results of research work aimed at the development and implementation of an integrated voice/data communications system over mobile radio channels. Three problems are investigated: (1) Analysis of the downlink loop in a two hop mobile UHF transmission system in which units communite with each other through a fixed base station, (2) the optimization of a quadrature coherent mean square error detector and its application to Tamed Frequency Modulation systems to achieve excellent spectral properties and good bit error rate performance, and (3) the design and implementation of a speech detector capable of detecting speech clauses in mobile radio channels. Computer simulation and laboratory test results are reported for the three problems under investigation.

Contents

1. Introduction

The demand for channels in the UHF radio spectrum, both for data and for voice, requires that the limited spectrum available be used as efficiently as possible. One way in which this can be done is the combination of both data and voice on a single channel, on a voice priority basis. A second method is the dynamic sharing of a group of channels amongst a larger group of users. When speech and data are transmitted on the same channel as packets, advantage can be taken of the bursty nature of each of these sources to combine the two approaches. Moreover, developments in VLSI have made the widespread use of digitally controlled radio dispatch systems technically feasible and economically attractive. Research carried out under this contract has focused on the design, analysis, simulation, and prototype development of a packet mobile radio system combining data and voice in the 800 MHz band.

1.1 Summary of Phase A

The first stage of this research, reported by Riordan, Mahmoud and Aidarous*, examined the basic structure of a two-hop mobile transmission system in which units communicate

---

* J.S. Riordan, S.A. Mahmoud and S. Aidarous, "Design Considerations in Packet Mobile Radio Data Networks", Report No. SC81-1, Department of Systems and Computer Engineering, Carleton University, March 1981.

with each other through a fixed base station. Analysis in Phase A concentrated on the uplink model. User group members are assumed to gain access to uplink channels by a variant of carrier sense multiple access. One of the essential parameters of the uplink model is the size of channel group to which a caller is assigned. At one extreme, the group of all channels could be regarded as commonly available to each user on demand, so that all channels constitute a single group. At the other extreme, a specific channel could be allocated to each of n users, so that n groups are formed. It was hypothesized that some intermediate condition, in which N_g channels are assigned to each group, may be a more effective allocation of communications resources. The relationship between group size, delay, blocking probability, and traffic intensity was examined. It was shown that under certain circumstances that optimum size of N_g was quite low, perhaps as small as 8 channels. These results were verified through simulation. The downlink hop, access to which is controlled by the base station, was also considered briefly.

A second part of Phase A was an examination of the relationship between packet loss and speech intelligibility. Subjective tests indicate that the optimum packet length is a function of loss rate. With transmission at 16 kilobits per second, best intelligibility was associated with packets of

200 bytes and greater in length. It is suspected that this phenomenon may be, in part, an artifact associated with the particular codec chip used for the experiments, however.

1.2 Phase B: Current Research

Research in Phase B reflects a greater emphasis on prototype design and development. Complementary to Phase A, analysis has concentrated on the downlink hop. Two feasible methods of shared channel access have been identified, and suitable protocols have been specified. The use of a separate downlink signalling channel, foreshadowed in Phase A, is considered in some detail, with analysis accompanied by corresponding systems simulation.

On the implementation side, two major problems have been addressed. The first of these is the design of a novel modulation system. Continuous Phase Modulation (CPM), which uses partial response signalling, has the advantage of excellent spectral properties and good error performance. To achieve these results, research in Phase B focused on the optimization of a quadrature coherent detector, in the mean square error sense, for a wide class of CPM with modulation index $h = 1/2$. The application of this detector to Tamed Frequency Modulation (TFM) has been investigated in detail.

The second problem investigated in this research is the design and implementation of a speech detector capable of detecting pauses in conversations held on mobile radio channels so that they can be allocated to the transmission of short data packets. A speech detection algorithm is proposed and implemented in real time on a digital signal processor. The detector is tested with various combinations of speech and noise samples.

1.3 Report Outline

Chapter 2 of this report examines three downlink channel models. One of these, while potentially efficient, is judged to be somewhat unreliable, and so is not considered further. Also considered in Chapter 2 are the protocols associated with the feasible downlink methods and an analysis of the delay associated with the use of a dedicated downlink signalling channel. Simulation in Chapter 2 is presented which verifies the analysis, and indicates that relatively small delay should be observed in channel access on the downlink hop.

Chapter 3 of this report investigates the optimization of a quadrature coherent detector, in the mean square error sense, for a wide class of Continuous Phase Modulation (CPM) schemes with modulation index $h = 1/2$. The application of this detector to Tamed Frequency Modems (TFM) is then investigated in detail. The system error rate performance was found to be very close to optimum for relatively low SNR while the asymptotic degradation is less than 0.6 dB.

Chapter 4 presents the main concepts and implementation details of a real time speech detector for the mobile radio environment. Various aspects of performance evaluation are examined: test bed, sample signals, subjective tests, objective results from the "activity monitor", and a microcomputer-based statistics collector.

Chapter 5 presents a summary of the results of the research completed in Phase B and provides suggestions and recommendations for further research.

2. Modelling and Simulation of the Two Hop System

2.1 Introduction

Following the development in Phase A of the contract, this chapter considers the development of an analytic model for traffic in the two hop system (see Fig. 2.1), and simulation results related to that model. Complimenting the emphasis in Phase A on the uplink channel, this chapter will concentrate on the handling of downlink traffic. Section 2.2 considers three methods of sharing downlink channels controlled from the base station. Each method requires the use of special signalling packets in addition to the voice packets; the form of these signalling packets and the associated protocols are described in section 2.3. One of the methods involves the use of a dedicated signalling channel, and the capacity of such a channel is analyzed in section 2.4. Section 2.5 presents a simulation of the overall model, and a comparison of the performance predicted through analysis and simulation is made. Section 2.6 summarizes results.

2.2 Downlink channel models

Voice packets arriving from transmitters over the uplink channel are buffered by the base station prior to retransmission. Temporary storage is necessary since, in the absence of dedicated downlink channels, dynamic allocaion of channels to receivers must be carried out. Three variations of the downlink allocation precedure will be described.

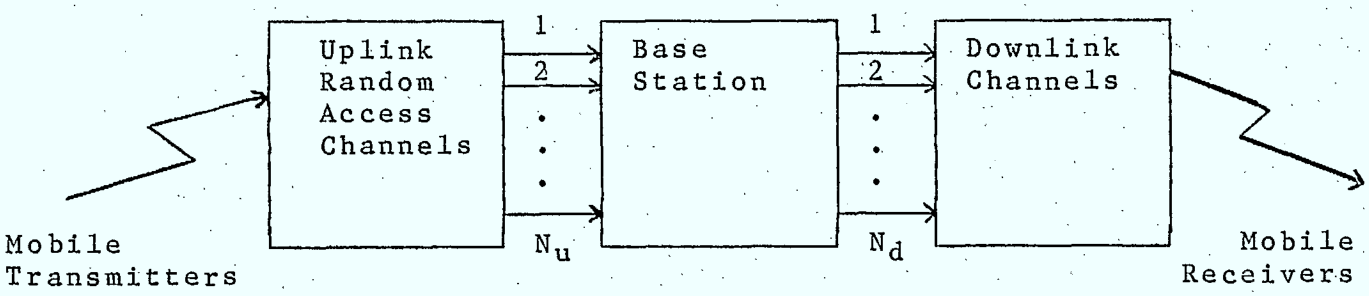


Fig. 2.1 Packet Mobile Network

2.2.1 Dedicated Signalling Channel Model

The basic dynamic allocation scheme is dedicated signalling channel model, so named because a single dedicated signalling channel is used to advise receivers of a message at the base station. All receivers in the quiescent state tune to the signalling channel. The base station then notifies receivers of an incoming talkspurt by the transmission of a signalling packet which indicates both the addressee and the channel upon which the talkspurt will be transmitted. The addressee receiver must then tune to the requisite channel and receive the talkspurt. At the completion of the talkspurt, the receiver returns again to listen on the signalling channel.

2.2.2 Seasonal Migration Model

The need for a dedicated signalling channel can be avoided by the allocation of two receivers, say A and B, to each channel. If a message appears for A, receiver B continues to listen passively. If a message appears for receiver B during transmission of the talkspurt A, an interrupt packet is generated directing receiver B to another channel. The latter may be a quiescent channel to which other receivers are tuned, or it may be one of a pool of additional overflow channels designated for this purpose. Upon completion of the talkspurt reception, channel B will again tune to its designated channel. This technique, described briefly in the initial contract report, may be termed a <u>seasonal migration model</u>. In effect, the receiver migrates to a different channel if its nominal channel is busy, but returns to it at the completion of each talkspurt.

2.2.3 Gypsy Model

A variant on this scheme might be termed the <u>gypsy model</u>. In this scheme, receiver B listens passively on the channel upon which it last received a talkspurt; other receivers, say A and C, may also be listening on the same channel. If a second talkspurt arrives for receiver B and the channel is available, then it is transmitted directly. If the channel is busy with a talkspurt transmission to either A or C, then an interrupt packet is generated as in the previous one, and receiver B is directed to another channel. The difference between this strategy and the migration model is that receiver B would not return to the channel to which it was directed, but would remain tuned to the last channel upon which a talkspurt was received.

2.2.4 Evaluation

There is probably little difference in efficiency between the gypsy model and the migration model. The gypsy model would be more effective in the case in which the time between the end of one talkspurt and the beginning of the next (for a particular receiver) was less than the channel tuning time of about 30 ms. The gypsy model has the disadvantage, however, that errors in the signalling packet could cause the base station to "lose" a receiver. In other words, a receiver might be listening on one channel while the base station expects it to be on another. Subsequent downlink messages could not be transmitted until the proper channel had been located. The effort of locating this channel would probably nullify any other advantage of the gypsy method. Therefore, it is assumed that downlink transmission will take place using either dynamic downlink allocation with a dedicated signalling channel, or by means of the seasonal migration model.

2.3 Downlink Channels Protocol

Packets may arrive at random times to the base station. Often they are not processed immediately because the message processor may be occupied. Thus they are buffered in an input buffer pool as shown in Fig. 2.2. Each packet will be processed, using its header, and directed to the appropriate output buffer before transmission to the appropriate receiver. The base station will monitor the status of both channels and receivers.

2.3.1 Dedicated Signalling Channel Protocol

1. Identify receiver through the header included in the packet.
2. If the receiver is silent (i.e., tuned to the signalling channel), search for the earliest available channel and send a signalling packet to direct the receiver to the designated channel.

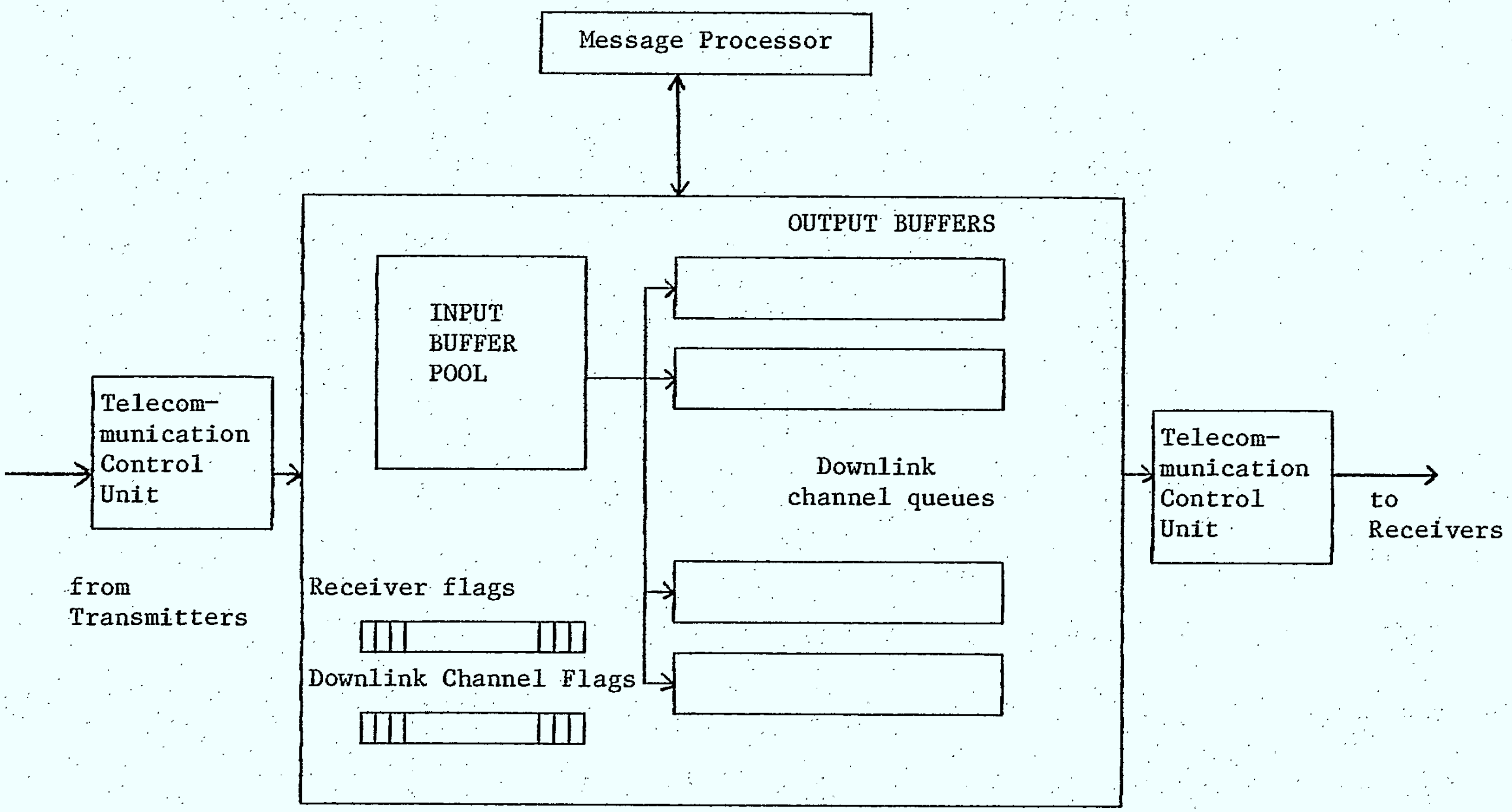


Figure 2.2 Buffering and Queuing in Base Station

3. If the receiver is busy (i.e., receiving a talkspurt on a specific channel), transmission will be delayed until the receiver returns to its silent state. A search for the earliest available channel is started and a signalling packet will be sent to direct the receiver.

The protocol is depicted in Fig. 2.3. It must be noted that there is a possibility of having a talkspurt arrived for a receiver while it is receiving the previous one. This is due to both search delays included in the CSMA technique used for the uplink channels and the search technique used for downlink channels.

2.3.2 Seasonal Migration Protocol

1. Identify receiver through the header in the packet.
2. Check if the receiver is silent (tuned to its designated channel) or busy (i.e., receiving a string of packets on a specific channel).
3. If the receiver is silent, a check will be made if there is a scheduled hold on its nominal channel.
   If there is a hold, search for the earliest available free channel, hold the channel, and schedule transmission.
   If there is no hold, hold its nominal channel and schedule transmission.
4. If the receiver is busy, check if there is a hold on his designated channel.

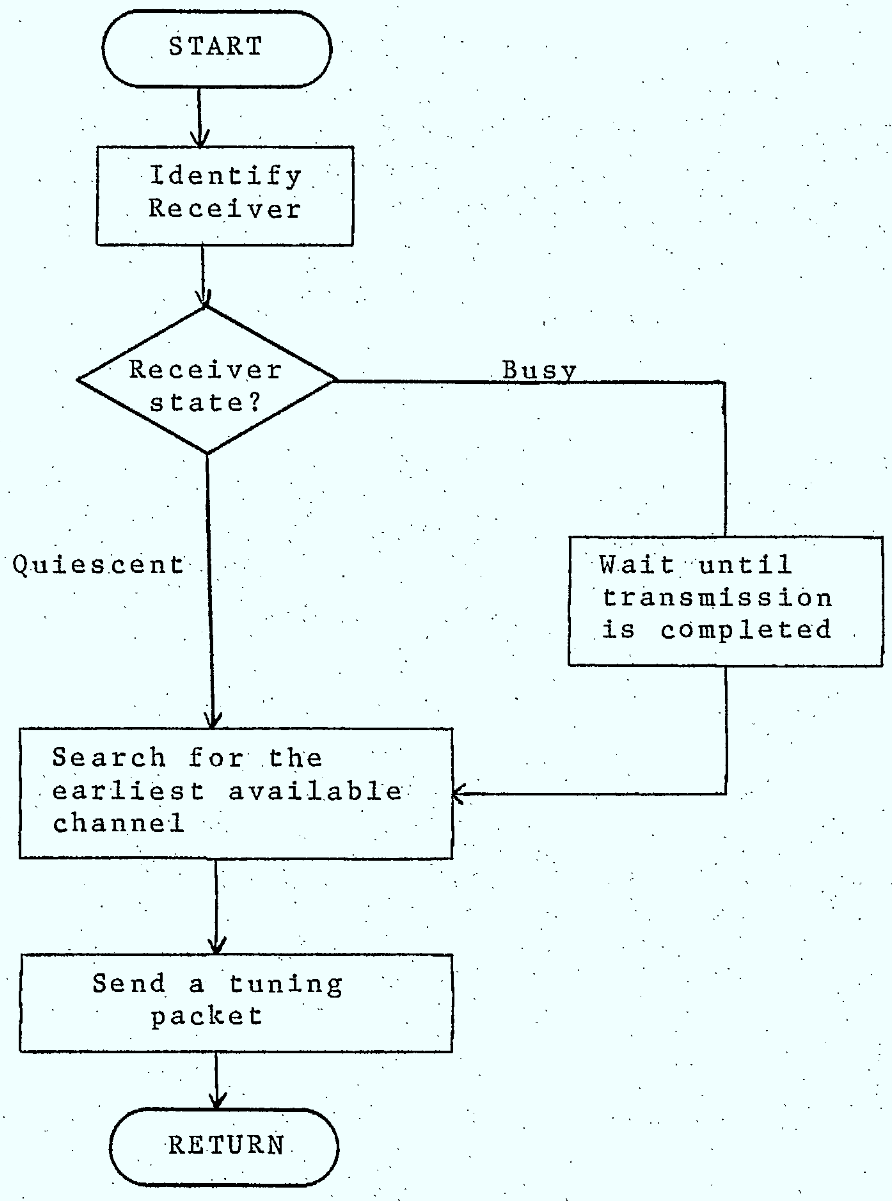


Fig. 2.3 Dedicated Signalling Channel Protocol

If there is no hold, send a "stay tuned" packet, hold the channel and schedule transmission.

If there is a hold on the channel, search for the earliest available free channel, hold the channel, schedule transmission, and send a tuning packet if it is not its nominal channel.

N.B.: Since there may be two talkspurts searching for the same downlink channel at the same moment, thus a hold of this channel will be needed by only one of the two talkspurts (see Fig. 2.4).

2.3.3 Evaluation

The use of a dedicated signalling channel will facilitate messages and downlink channel allocation in the base station. However, efficient use of the additional signalling channel must be investigated. The efficient use of downlink channels through the seasonal migration protocol is associated with a complex processing in the base station. Inspite of the fact that additional processing in the base station has slight effect on the delay, the seasonal migration protocol is more valuable to higher delay variance. On the other hand, the additional tuning delay associated with the dedicated signalling channel protocol will increase the average delay of packets.

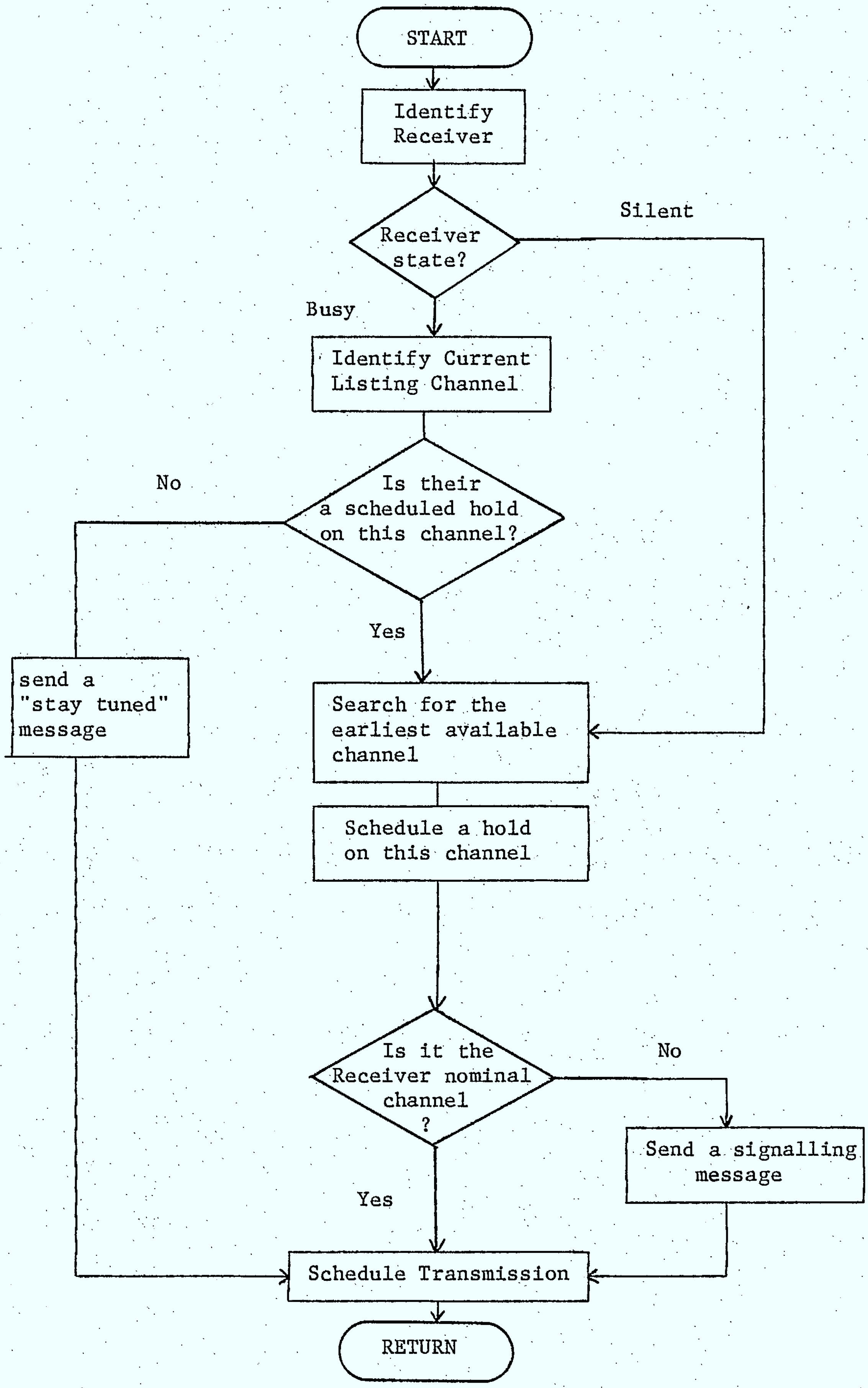


Fig. 2.4   Seasonal Migration Protocol

The gypsy protocol has not been considered due to its unreliability since there is no apriori knowledge, available to the base station, in order to monitor the receivers.

2.4 Signalling channel capacity

A dedicated downlink signalling channel is devoted entirely to communications overhead. It is, therefore, important to ascertain the number of active downlink channels which can be supported by one such overhead channel.

Assume a dedicated downlink signalling channel to which receivers tune in the quiescent state. A short packet (see Fig. 2.5) is sent indicating the presence of a message (talkspurt) for a particular receiver, and indicating the channel upon which the message will be sent. To combat fading, the signalling packet is transmitted twice, with a 10 ms gap between transmissions.

With a 16 kb/s signalling rate, this 50 bit packet is 3.125 ms long. The typical talkspurt is about 1.3 seconds in duration. If two signalling packets are sent for each talkspurt, then the ratio of their channel usage is:

$$\frac{t_u}{2t_p} = \frac{1.3}{6.25\text{x } 10^{-3}} = 208$$

	Flag	Address	Rpt no.	Channel	FEC	Flag
No. of bits	8	8	2	8	16	8

Fig. 2.5 Signalling Packet Format

For rapid signalling, the traffic intensity on the signalling channel should be less than that on the message channels. The above ratio indicates that one signalling channel could handle about 100 message channels.

This assumption can be examined more accurately through the use of a simple M/D/1 queueing model. The expected delay, under the assumption of Markovian arrivals and deterministic service, is

$$E(t) = \frac{1}{u\ (1-p)} \left[1 - \frac{p}{2}\right]$$

where = service rate = $\dfrac{1}{mt_p} = \dfrac{1}{2 \times 3.125 \times 10^{-3}}$

m = number of packet repeats, chosen as 2 here.

p = signalling channel traffic intensity

As an example, select p = 0.4, a fairly low traffic intensity. The delay is then

$$E\ (t) = \frac{6.25 \times 10^{-3}}{0.6}\ (1 - 0.2) = \underline{8.33\ \text{ms}}$$

which is only 33% greater than the signalling packet duration itself.

This delay figure is to some extent misleading, because of the 10 ms gap between repeated signalling packets. It would be more accurate to say that the throughput of the signalling channel under the stated load is

$$\frac{1}{8.33 \times 10^{-3}} = 120 \text{ packets/second}$$

Because the signalling packets are short relative to talkspurt time, the signalling channel is capable of handling a large number of normal communications channels. With a duplicated signalling packet occupying less than 10 ms, inclusive of queuing time, more than 100 downlink channels could be accommodated with little delay. It is, therefore, concluded that for heavy traffic, the use of dedicated signalling channel is justified.

2.5 Simulation of Performance

Simulation of the two-hop system is considered. For the uplink channel model, the CSMA protocol described in Phase A is used*. Uplink channels are grouped into 8 channels/12 transmitters. Talkspurts are generated with average length

---

* J.S. Riordan, S.A. Mahmoud, and S.E. Aidarous, "Design Considerations in Packet Mobile Radio Date Networks", Report No. SC81-1, Dept. of Systems and Computer Engineering, Carleton University, March 1981.

of 1.3 sec. and silent gaps average length of 1.95 sec. both exponentially distributed. Maximum number of channels searched is 8, and the channel sensing delay is 30 m.s. The one way propagation delay is 50 u. sec.

A processing delay of 500 u. sec. is considered in the base station. This delay accounts for checking and classification of different packets received and allocating them to the appropriate buffers for retransmission to receivers. It includes also overhead needed to generate the associated signalling packets. Buffering, classification and queuing in the base station is shown in Figure 2.2.

2.5.1 Performance of Dedicated Signalling Channel Model

Dynamic allocation of downlink channels using a dedicated signalling channel will be considered. The protocol described in Section 2.3.1 is used. For the downlink model, 6-8 channels/12 receivers are used. In most runs, 7 downlink channels/12 receivers were used since no additional capacity is required, as in the case of the uplink channels, for retransmission and the CSMA strategy. A signalling packet length of 3.125 m.s. is considered.

Figure 2.6 shows the simulation model used and simulation results are shown in Table 2.1 for different numbers of transmitters. Figure 2.7 shows the downlink delay drops below the 10 m.s. level for groups of transmitter greater than 36. However, the signalling channel delay is kept below 3.2 m.s. It must be noticed that an additional 30 m.s. has to be added to the downlink delay to account for the receiver tuning delay.

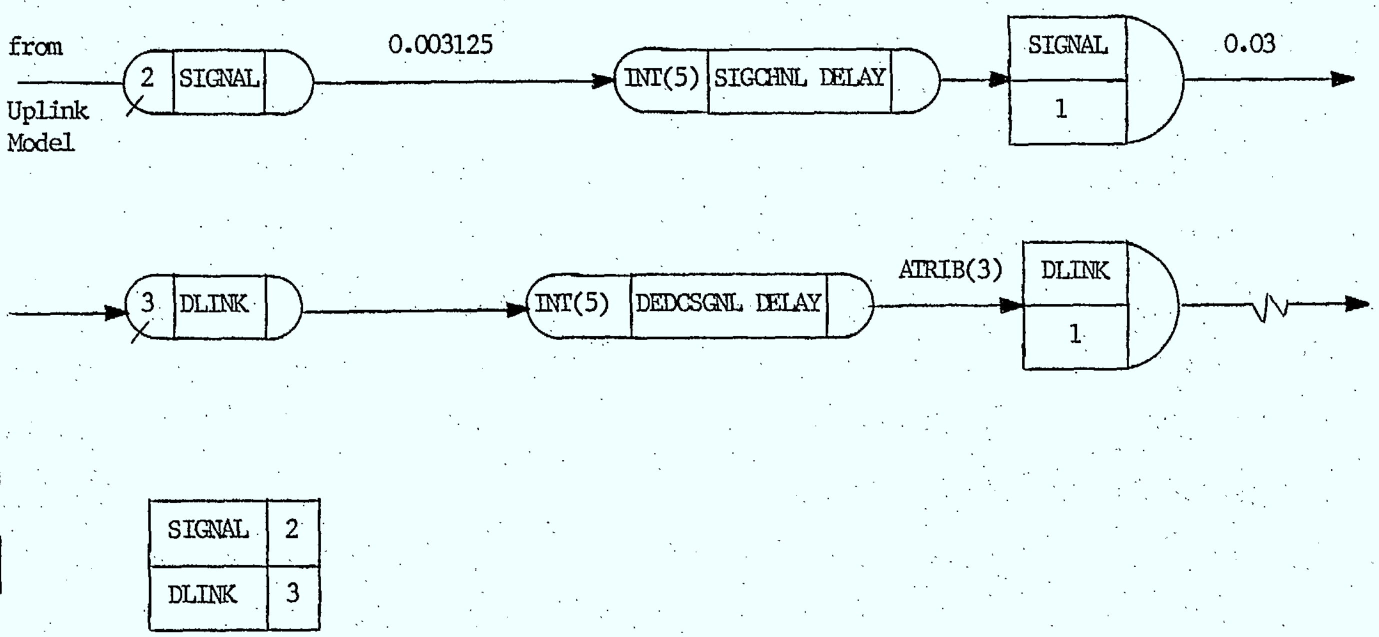


Fig. 2.6 Dedicated Signalling Channel Model

Table 2.1 Dedicated Signalling Channel Model Performance

No. of Channels Transmitters (Receivers)	No. of Uplink Channels	No. of Down-link Channels	Signalling Channel		Downlink Channel	
			av. delay (m.s.)	traffic	av. delay (m.s.)	traffic
12	8	8	3.141	0.0103	31.76	0.5831
		7	3.141	0.0108	109.0	0.5831
		6	3.141	0.0108	392.6	0.5831
18	12	10	3.141	0.0166	68.05	0.70391
24	16	14	3.151	0.0221	16.31	0.6675
30	20	17	3.155	0.0276	19.28	0.6782
36	24	21	3.165	0.0328	11.04	0.6592
42	28	24	3.185	0.0387	6.444	0.6741
48	32	28	3.191	0.0437	3.291	0.65106

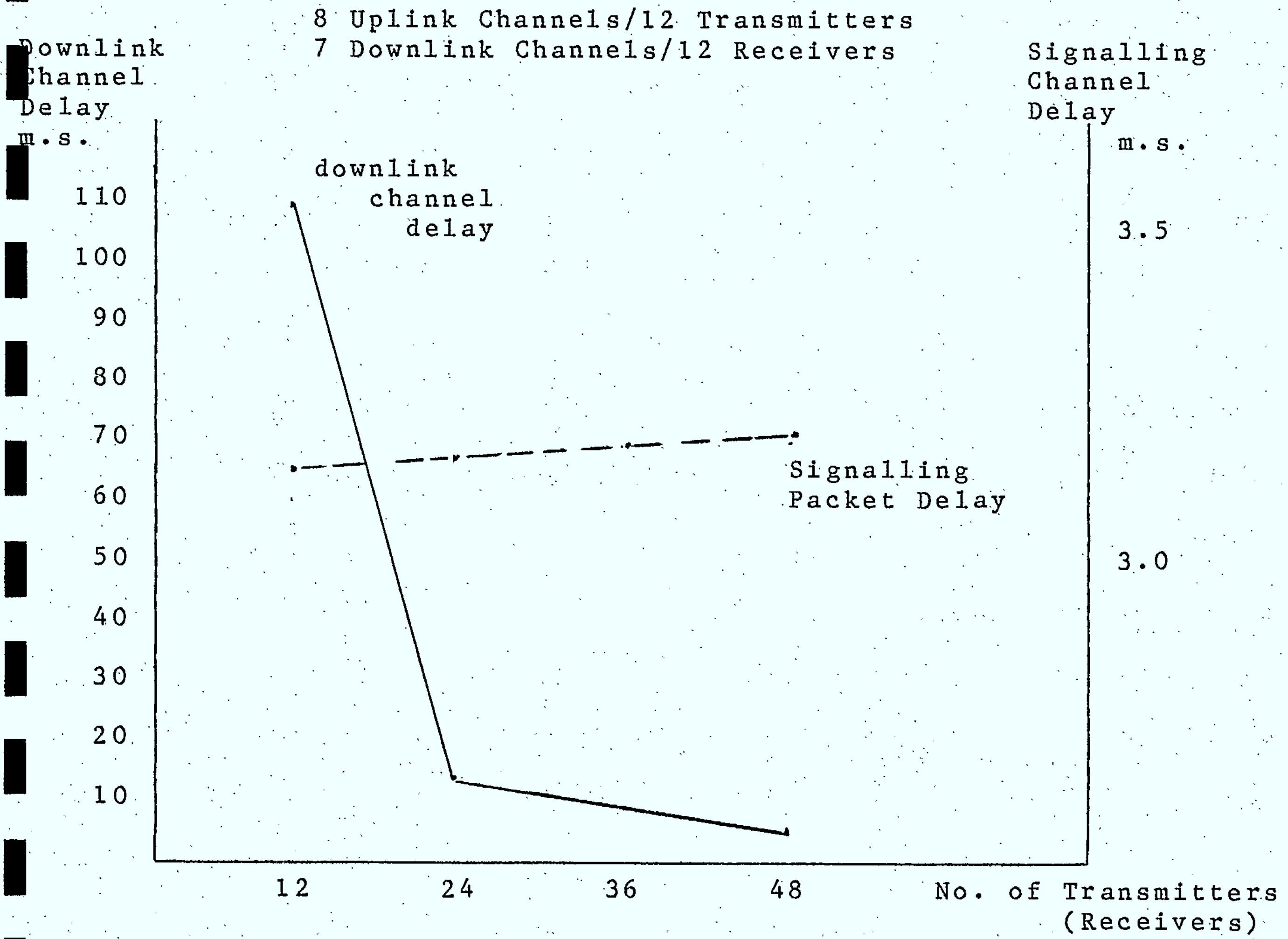


Figure 2.7 Dedicated Signalling Channel Model Delay

2.5.2 Performance of Seasonal Migration Model

In this model, two receivers will be allocated to one downlink channel. In other words, in the case of 7 downlink channels/12 receivers, six pairs will be assigned to six channels unless a signalling packet direct them to another channel. The other non-assigned channel will serve as additional overflow channel.

The seasonal migration model is shown in Fig. 2.8 and the delay is shown in Table 2.2 for different values of additional pool channels. It is noticed that a remarkable reduction of delay is associated with the increase of the additional pool channels. This results from the fact that the tuning delay (30 m.s.), which represents a high percent of the total delay, is not added to all packets as the case of the dedicated signalling channel model.

Table 2.2 Seasonal Migration Model Delay (24 Receivers)

Number of Additional Pool Channels	0	1	2	3	4
Average Delay m.s.	69.21	45.7	32.07	21.15	12.1

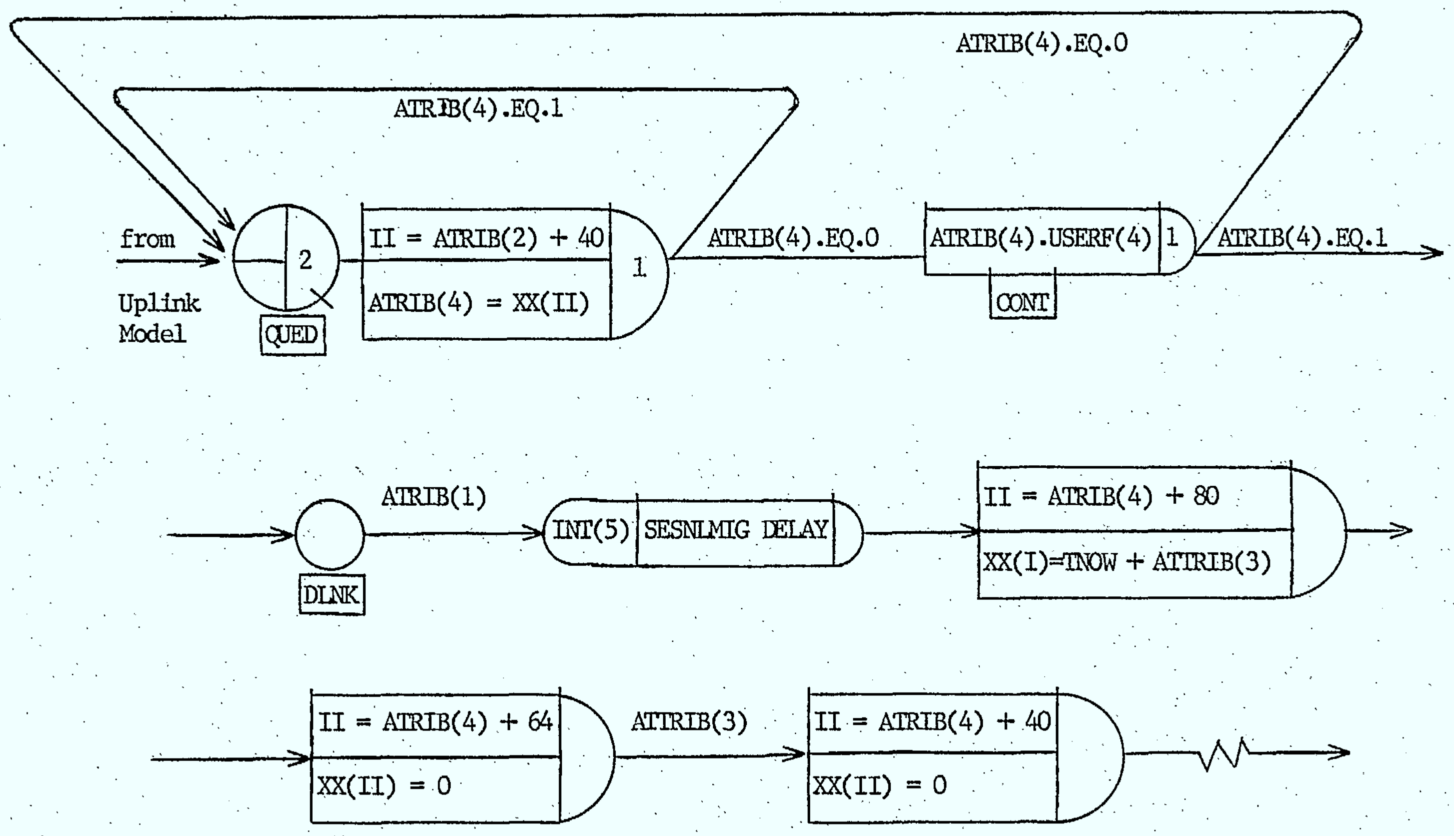


Fig. 2.8 Seasonal Migration Simulation Model

2.5.3 Evaluation

Comparing the results of the above two cases, we can clearly notice the difference in the delay characteristics of the two proposed models. The dedicated signalling channel is characterized by a relatively high average delay and low variance, while the seasonal migration model is characterized by a relatively low delay and high variance (see Fig. 2.9). However, considering the voice intelligibility tests performed in Phase A of this work, the delay variance in the seasonal migration model can be remarkably reduced by discarding high delay packets in the base station. From the delay characteristics shown in Fig. 2.9, the number of these packets is very small and is within the intelligibility limits specified by the experimental work in Phase A.

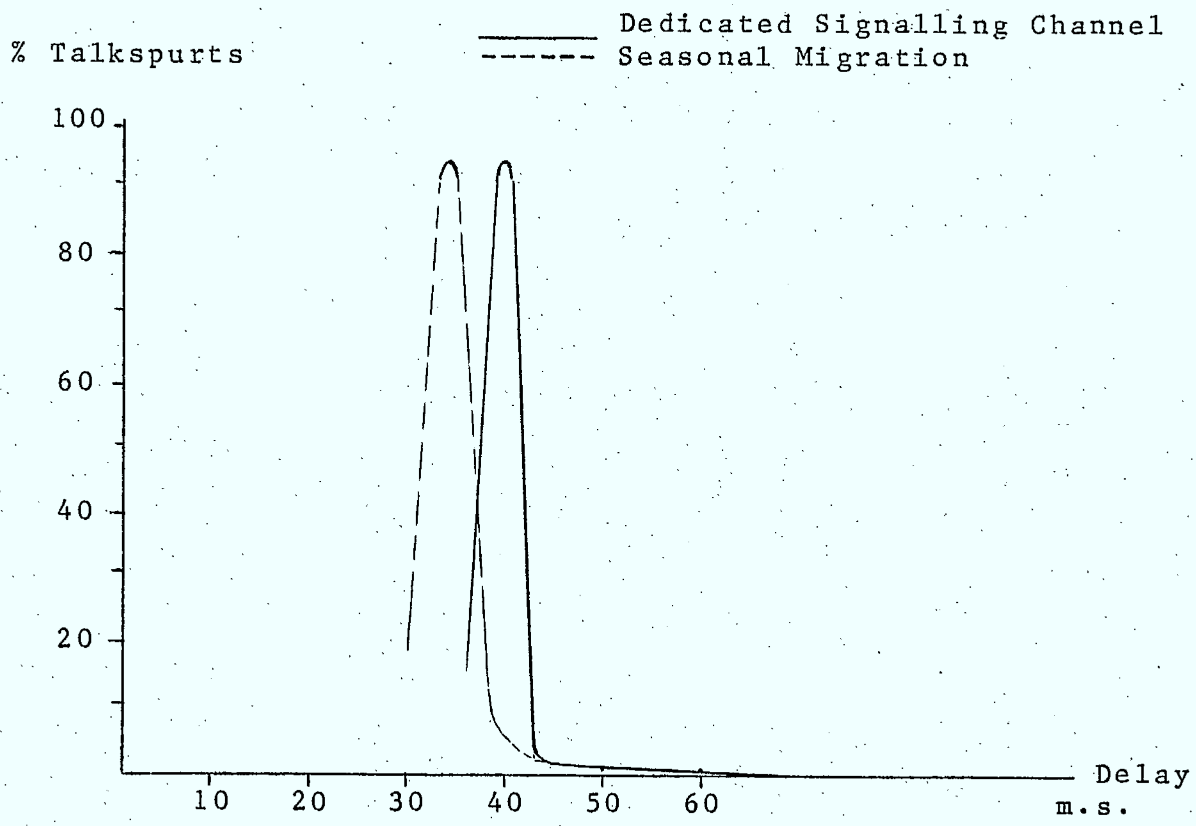


Fig. 2.9 Delay Distribution

2.6 Summary

Both the dedicated signalling channel model and the seasonal migration model appear to be suitable as downlink channel allocation procedures. Analysis and simulation indicate that the delays associated with downlink transmission will be caused by queuing within the base station buffers and by receiver channel switching time. the signalling time, to alert the receiver to the presence of a talkspurt, is typically less than 10 milliseconds. Thus, the choice between the dedicated signalling channel model and the seasonal migration model should be made on grounds other than delay. The former method has the advantage of relative simplicity, but does require an additional channel. The latter makes more efficient use of downlink channels, but requires a somewhat more complex protocol. Dedicated downlink signalling would be justified when a large number of channels, say fifty or more, is involved.

It is recommended that further work be carried out to specify more precisely procedures for handling packets in the base station buffer, and to analyze the resultant delays. Operationally, tests of the two downlink protocols should be made. Consideration should be given to the possibility of a hybrid protocol which is capable of operating in either the seasonal migration mode or the dedicated signalling channel mode, as conditions dictate.

3.0 Optimal Design of a Quadrature Coherent Detector for a Tamed Frequency Modem

3.1 Introduction

It has recently been demonstrated that a wide class of modulation methods in the CPM family possess excellent spectral properties. This is achieved by using partial response signalling, and keeping the signal phase continuous at the data switching instants. Some schemes have been shown to combine excellent spectral properties and good error performance by comparison to QPSK [1], [2], ([3], [4] and [5]).

The optimum receiver, for most of these schemes, is a Maximum Likelihood Estimator (MLE), [4], [6]. The MLE consists basically of a bank of matched filters followed by a Viterbi processor, [4]. The number of those filters increases rapidly with the receiver observation interval, LT, as well as the order of the partial response system polynomial. In practice, the implementation of such a receiver could be a heavy burden, especially for relatively high speed data systems.

It is also known that [2,7,8] CPM systems, with modulation index h=1/2, bear some similarities to MSK; i.e., under certain conditions, the phase of the CPM signal could be decomposed into a component identical to the MSK's plus an "intersymbol interference" component which is strictly less than $\pi/2$ in absolute value. In these cases, the CPM signal will show an open eye which is very similar to the MSK's. Therefore, the demodulation strategy suggested by de Buda [1] could be used if two arm filters are included to reduce the effect of the deliberately introduced intersymbol interference (ISI) component.

The objective of this research is the optimization of the receiver predetection filters. This problem has been approached in [2] and [7] for the case of TFM. The philosophy in [2] and [7] is approximating the TFM system by an "equivalent" baseband system, in order to make use of the linear theory in [9] for deriving the optimum arm filters. The performance of some selected filters, employed in the receivers of several CPM(h=1/2) systems, was investigated in [8]. It was found that some selected filters performed better with some CPM schemes than others. In this section, we approach the problem in some generality. This approach lead to a general form of a suboptimum filter, which could be useful for a wide class of CPM with h=1/2.

3.2 Theory and Main Concepts

A block diagram of a CPM modulator is shown in Figure 3.1. The input data bit, α_i, is binary and represented by an impulse $\delta(t-iT)$, where 1/T is the input data rate.

The input bit α_i is first passed through a premodulation filter with impulse response g(t), g(t) is assumed to be symmetric around the data impulse, and extends over a time interval MT.

The total change in the modulated signal phase, $\phi(t)$, due to one bit of information is restricted to $\mp\frac{\pi}{2}$. Therefore, g(t) has to satisfy;

$$\int_{-\infty}^{\infty} g(\tau)\, d\tau = 1/2 \qquad 3.1$$

Due to the overlapping nature of the premodulation filter output pulses, the total change in $\phi(t)$ during one signalling interval, T, is affected by M adjacent bits. This could be written as;

$$\phi[(n+1)T] - \phi[nT] = \pi \sum_{i=-\frac{M-1}{2}}^{M-1/2} C_i\, \alpha_{n+i} \qquad 3.2$$

where,

$$C_i = \int_{-(i+\frac{1}{2})T}^{-(i-\frac{1}{2})T} g(\tau)\, d\tau \qquad 3.3$$

and,

$$\sum_{i=-\frac{M-1}{2}}^{\frac{M-1}{2}} C_i = 1/2$$

It could be easily shown that the signal absolute phase at time t=nT is given by;

$$\phi(nT) = \frac{\pi}{2} \sum_{i=-\infty}^{n} \alpha_i + \pi \sum_{i=1}^{M} \sum_{j=i}^{M} (C_j \alpha_{n+i} - C_{-j} \alpha_{n-i+1})$$

$$= \phi_{MSK}(nT) + \phi_i(nT) \qquad 3.4$$

where $\phi_{MSK}(nT)$ is the phase, at t=nT, of an MSK signal using the same input data as the CPM signal. ϕ_i is due to the introduced intersymbol interference. In order for the CPM signal to have an open eye, $\phi_i(nT)$ must be less than $\pi/2$ in absolute value. This leads to the following constraint on $\{C_j\}$:

$$\sum_{i=1}^{M} \left(\left| \sum_{j=i}^{M} C_j \right| + \left| \sum_{j=i}^{M} C_{-j} \right| \right) < \frac{1}{2} \qquad 3.5$$

Equation 3.5 is satisfied for several CPM systems including TFM[2], Doubinary MSK[4], Gaussian MSK[10] and Raised Cosine CPM with h=1/2 [4]. Therefore, the receiver structure shown in Figure 3.2 could be used for these systems, as well as any CPM system satisfying the constraints given by equations 3.1 and 3.5.

3.2.1 Prediction Filter Optimization

A schematic diagram for the receiver under consideration is given in Figure 3.2. Perfect synchronization is assumed for the carrier and bit timing. The channel is assumed additive white Gaussian with spectral density N_o. Using the same notation as in [4], the input signal to the receiver is given by,

$$r(t) = S(t, \underline{\alpha}) + n(t) \tag{3.6}$$

where

$$S(t, \underline{\alpha}) = \sqrt{\frac{2E_B}{T}}\, \mathrm{Cos}\,[\phi(t, \underline{\alpha}) + \omega t]$$

$$\triangleq I(t, \underline{\alpha}) \cos \omega t - Q(t, \underline{\alpha}) \sin \omega t \qquad 3.7(a)$$

where

$$I(t, \underline{\alpha}) = \sqrt{\frac{2E_B}{T}} \cos \phi(t, \underline{\alpha})$$

$$Q(t, \underline{\alpha}) = \sqrt{\frac{2E_B}{T}} \sin \phi(t, \underline{\alpha}) \qquad 3.7(b)$$

and

$$n(t) = n_c(t) \cos \omega t - n_s(t) \sin \omega t \qquad 3.8$$

where E_B is the signal energy per bit, the data rate is 1/T, is the information-carrying phase, and w is the carrier radian frequency.

The input signal to the arm filters in Figure 3.2, after dropping the components centered around 2w, is given by;

$$r_i(t, \underline{\alpha}) = \frac{1}{2} I(t, \underline{\alpha}) + \frac{1}{2} n_c(t)$$

$$r_q(t, \underline{\alpha}) = \frac{1}{2} Q(t, \underline{\alpha}) + \frac{1}{2} n_s(t) \qquad 3.9$$

Therefore, the signal presented to the I-channel sampler at time t=nT is given by;

$$D(nT) = \int_{-\infty}^{\infty} [I((nT-\tau),\underline{\alpha}) + n_c(nT-\tau) h(\tau)\, d\tau \tag{3.10}$$

where h(t) is the impulse response of the predetection filter. In the absence of noise, and the deliberately introduced ISI,D(nT) is given by;

$$D(nT) = D_{MSK}(nT) = \sqrt{\frac{E_B}{2T}}\, \mathrm{Cos}\, \phi_{MSK}(nT) \tag{3.11}$$

Let us define the error function e(nt) as,

$$e(nT) = D(nT) - D_{MSK}(nT)$$

and, therefore, the mean square error becomes;

$$E\left\{e^2(nT)\right\} = E\left\{\left[\frac{1}{2}\int_{-\infty}^{\infty} [I((nT-\tau),\underline{\alpha}) + n_c(nT-\tau) h(\tau)\, d\tau - \sqrt{\frac{E_B}{2T}}\, \mathrm{Cos}\, \phi_{MSK}(nT)\right]^2\right\} \tag{3.12}$$

Assuming that $n_c(t)$ and $I(t,\underline{\alpha})$ are statistically independent, equation 3.12 could be rewritten as,

$$E\left\{e^2(nT)\right\} = \frac{E_B}{2T} + \frac{1}{4}\iint_{-\infty}^{\infty} [R_i(\tau_1-\tau_2) + R_n(\tau_1-\tau_2)]\, h(\tau_1) h(\tau_2)\, d\tau_1\, d\tau_2$$

$$-\sqrt{\frac{E_B}{2T}} \int_{-\infty}^{\infty} \left[E\left\{ I(nT-\tau,\underline{\alpha}) \,\middle|\, \mathrm{Cos}\, \phi_{MSK}(nT) > 0 \right\} \right. \tag{3.13}$$

$$\left. - E\left\{ I((nT-\tau),\underline{\alpha}) \,\middle|\, \mathrm{Cos}\, \phi(nT) < 0 \right\} \right] h(\tau)\, d\tau$$

$R_i(\tau)$ and $R_n(\tau)$ are the autocorrelation functions of $I(t,\underline{\alpha})$ and $n_c(t)$ respectively.

Minimizing the mean square error function as given by equation 3.13 could lead to improving the system error performance under the condition that the eye diagram, following the arm filter, is not completely closed for any permutation of the data vector, $\underline{\alpha}$. Therefore, we will constrain the expectation of the eye opening associated with the data vector $\underline{\beta}$ to a constant; i.e.,

$$\int_{-\infty}^{\infty} [E\{I((nT-\tau),\underline{\beta}) \mid \cos\phi_{MSK}(nT)>0\} - E\{I(nT-\tau,\underline{\beta}) \mid \cos\phi_{MSK}(nT)<0\}]\, h(\tau)\, d\tau = f \qquad 3.14$$

where $\underline{\beta}$ is the set of data vectors which determine the inner boundary of the eye diagram in the time interval $(n-1)T \leq t \leq (n+1)T$, and f is a constant.

The minimization of equation 3.13, under the constraint of equation 3.14 is a standard problem in the calculus of variations. Following the procedure outlined in [11], we arrive at the following result for the frequency characteristic of the premodulation filter,

$$H(\omega) = \frac{I(\omega) + \lambda I_s(\omega)}{F(\omega) + N(\omega)}\, e^{j\omega nT} \qquad 3.15$$

where,

$$F(\omega) = \mathcal{F}[R_i(\tau)]$$

$$N(\omega) = \mathcal{F}[R_n(\tau)]$$

$$I(\omega) = \mathcal{F}[E\{I(nT-\tau,\underline{\alpha}) \mid \cos\phi_{MSK}(nT)>0\} - E\{I(nT-\tau,\underline{\alpha}) \mid \cos\phi_{MSK}(nT)<0\}]$$

$$I_s(\omega) = \mathcal{F}[E\{I(nT-\tau,\underline{\beta}) \mid \cos\phi_{MSK}(nT)>0\} - E\{I(nT-\tau,\underline{\beta}) \mid \cos\phi_{MSK}(nT)<0\}]$$

$\mathcal{F}$ denotes the Fourier transform, λ is a Lagrange multiplier under low SNR conditions, equation 3.15 is simplified to:

$$H(\omega) \approx [I(\omega) + \lambda I_s(\omega)]\, e^{j\omega nT}$$

and, therefore, h(t) is approximated by;

$$h(t) \approx h_{amf}(t) + \lambda h_{min}(t) \qquad 3.16$$

where

$$h_{amf}(t) = E\{I(nT-\tau,\underline{\alpha}) \mid \cos\phi_{MSK}(t)>0\} - E\{I(nT-\tau,\underline{\alpha}) \mid \cos\phi_{MSK}(nT)<0\}$$

$$3.17$$

$$h_{min}(t) = E\{I(nT-\tau,\underline{\beta}) \mid \cos\phi_{MSK}(t)>0\} - E\{I(nT-\tau,\underline{\beta}) \mid \cos\phi_{MSK}(nT)<0\}$$

$h_{amf}(t)$ is recognized as the impulse response of an Average Matched Filter (AMF). h(t) as defined equation 3.16 by is matched to a weighted average of the baseband signal. Hence, the arm filter is a Weighted Average Matched (WAMF). h(t) is, strictly speaking, nonrealizable. However, a digital implementation becomes possible if $h(t)$ is restricted to a time interval, LT, and delayed by LT/2. We define h (t) as follows

$$h_L(t) = \begin{cases} h(t-LT/2) & o \leq t \leq LT \\ o & \text{otherwise} \end{cases} \qquad 3.18$$

We shall denote a filter with impulse response $h_L(t)$ by WAMFL.

The Lagrangian multiplier, in equation (16), will be assigned the value of $\lambda = \lambda_{opt}$ which minimizes the system symbol error rate as given by equation 3.22.

3.3 Performance Analysis and Application to TFM

The signal presented to the I-channel sampler, at time t=nT, is given by equation 3.10, with h(t) given by equation 3.18 for an observation interval LT. For a given data sequence $\underline{\alpha} = \underline{\alpha}^{i}$, D(nt) could be split into a signal component;

$$D_S(nT) = \int_0^{LT} \sqrt{\frac{E_B}{2T}}\ \mathrm{Cos}\ \phi\,(nT-\tau, \underline{\alpha}^{i})\ h_L(\tau)\,d\tau \qquad 3.19$$

and a zero mean Gaussian noise component with variance,

$$\sigma^2 = \frac{N_0}{4} \int_0^{LT} h_L^2(\tau)\,d\tau \qquad 3.20$$

Therefore, the conditional probability of symbol error is given by,

$$P_I\,(\mathcal{E} \mid \underline{\alpha}^{i}) = Q\left(\sqrt{d_i \frac{E_B}{N_0}}\right) \qquad 3.21$$

where

$$d_i = \frac{2}{T}\ \frac{\left[\int_0^{LT} \mathrm{Cos}\ \phi((nT-\tau), \underline{\alpha}^{i})\ h_L(\tau)\,d\tau\right]^2}{\int_0^{LT} h_L^2(\tau)\,d\tau} \qquad 3.22(a)$$

d_i is referred to as the squared Euclidean distance [8]. The average probability of symbol error is,

$$P_I(\mathcal{E}) = \sum_{i=1}^{2^L} P(\underline{\alpha}^i)\, Q\left(\sqrt{d_i \frac{E_B}{N_0}}\right) \qquad 3.22(b)$$

and the average probability of bit error would be,

$$P(\mathcal{E}) = \sum_{i=1}^{2^L} P(\underline{\alpha}^i)\left[P_I(\mathcal{E}|\underline{\alpha}^i)(1-P_Q(\mathcal{E}|\underline{\alpha}^i)) + P_Q(\mathcal{E}|\underline{\alpha}^i)(1-P_I(\mathcal{E}|\underline{\alpha}^i))\right]$$

$$\simeq 2\,P_I(\mathcal{E}) \qquad 3.23$$

For a given system, and for a small observation interval, one could write $P_I(\mathcal{E})$ in terms of λ, and proceed to minimize $P_I(\mathcal{E})$ with respect to λ. This should lead to the optimum value of $\lambda = \lambda_{OPT}$. However, the problem was found difficult to be dealt with analytically. Therefore, an exhaustive search for λ is preferred, and could be easily done on a digital computer.

3.3.1 Application to TFM

Tamed Frequency Modulation is known as one of the most efficient modulation techniques in the CPM family. It has been shown that TFM, if optimally detected, is only 1dB worse than the optimally detected QPSK [12]. The impulse response of the TFM premodulation filter is given by [2];

$$g(t) = \mathcal{F}^{-1}\{G(\omega)\}$$

where

$$G(\omega) = \cos^2\left(\frac{\omega T}{2}\right) \frac{\omega T/2}{\sin(\omega T/2)} \qquad 3.24$$

g(t) is basically infinite, and must be truncated for practical realizations; it is truncated to 5T for our application, and Figure 3.6(a) shows the corresponding eye diagram.

A data vector 15-bits long was found practically sufficient for the computation of $h_{amf}(t)$ and $h_{min}(t)$ [equn. 3.16]. The symbol error rate, as a function of E_B/N_o, was computed according to Equation 3.22 for different values of LT, with λ as a parameter.

The values of $\lambda \approx$ 0.225, 0.125, 0.115, 0.113 yielded minimal error rates for the observation intervals LT=4T, 6T, 8T and 10T respectively. The resulting impulse responses and the corresponding frequency responses are given in Figure 3.3 and Figure 3.4 respectively. The AMF is also shown for comparison. The symbol error rate curves, obtained by the impulse responses of Figure 3.3, are given in Figure 3.5(a) for low SNRs, and Figure 3.5(b) for high SNRs. The error rate obtained by a MSK filter is also shown.

It can be seen that WAMF6 and WAMF8 yielded almost identical error performance. This could be explained by the fact that $h_{amf}(t)$ and $h_{min}(t)$ (equation 3.16) are practically limited to 6T. Therefore, increasing the observation interval beyond 8T is not expected to yield any improvement in the system error performance. The system performance with a WAMF6 is about 1dB worse than QPSK for SNR $\leq$ 10dB, while the asymptotic degradation is $\leq$ 1.6dB; i.e., the WAMF6 TFM receiver performance is very close to optimum (MLE) for low SNR, and the asymptotic degradation, compared to a MLE receiver, is approximately 0.6dB. The performance of the WAMF4 is about 0.06dB worse than that of the WAMF6 for all SNRs, while the MSK filter is about 1dB worse. It can also be seen that although the performance of the AMF is good for very low SNRs, it is not acceptable for the practical range of SNRs.

The minimum Squared Euclidean Distance (as defined in [8]) given by equation 3.21(a), is hown in Table (I) for different predetection filters. The SPAM filter [8] is also shown for comparison.

The eye diagrams, generated by the impulse responses of Figure 3.3 are shown in Figure 3.6. It can be seen that the eye patterns, associated with WAMF8-WAMF4, have essentially the same level at the sampling moment t=nT. However, the zero crossings in the case of the WAMF4 are less jittery; a fact which could be important in systems where the symbol timing is extracted from the I and Q channel signal´s zero crossings.

3.4 Summary

We have seen that de Buda's FFSK demodulator could be used for a wide class of CPM with h=1/2. The optimization of this receiver, in the mean square error sense, has been shown to be realizable. A general formula for a suboptimum predetection filter has been presented. The filter is matched to a weighted average of the baseband signal, and is easy to compute and design.

The theoretical results have been applied to a TFM receiver, and the system error performance was found very close to that of an MLE receiver for low SNR (less than 10dB). The assymptotic degradation is $\leq$ 0.6dB, compared to the optimum (MLE) receiver. It was also found that, by restricting the impulse response of the arm filters to 4T, a near-optimum performance is still possible. This could be an important consideration in the design of relatively high speed data systems, whenever a digital implementation of the ARM filters is required.

MLE	SPAM	MSKF	AMF	WAMF4	WAMF6	WAMF8
1.57	0.95	0.96	0.784	1.251	1.279	1.31

Table 3.1 Minimum Squared Euclidean Distance for Different TFM Receivers

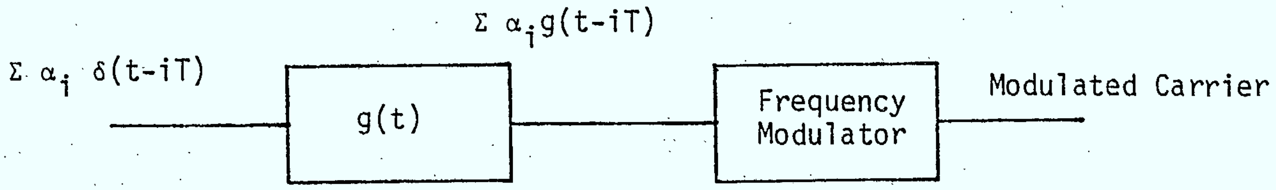


Figure 3.1 CPM Modulator

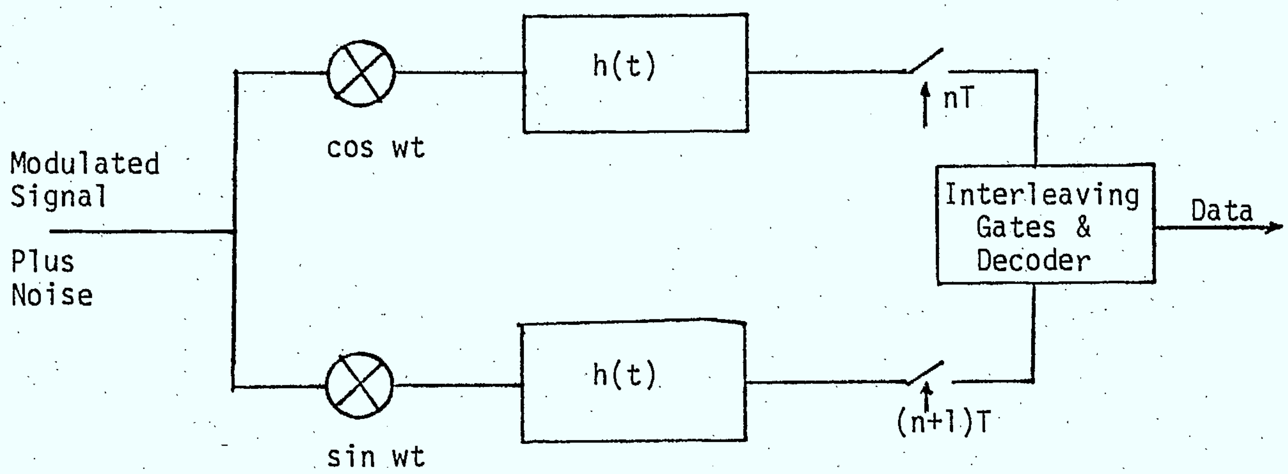


Figure 3.2 Receiver Structure

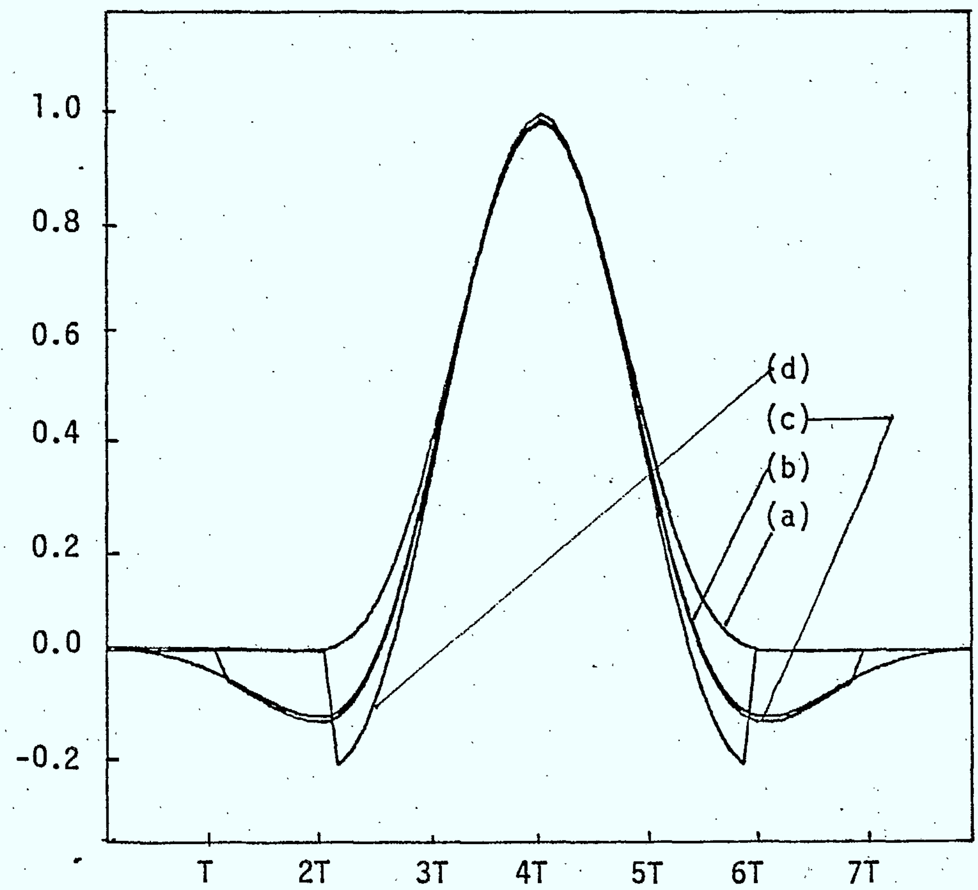


Figure 3.3 Impulse Response of the Predetection Filter

(a) AMF (b) WAMF8 (c) WAMF6 (d) WAMF4

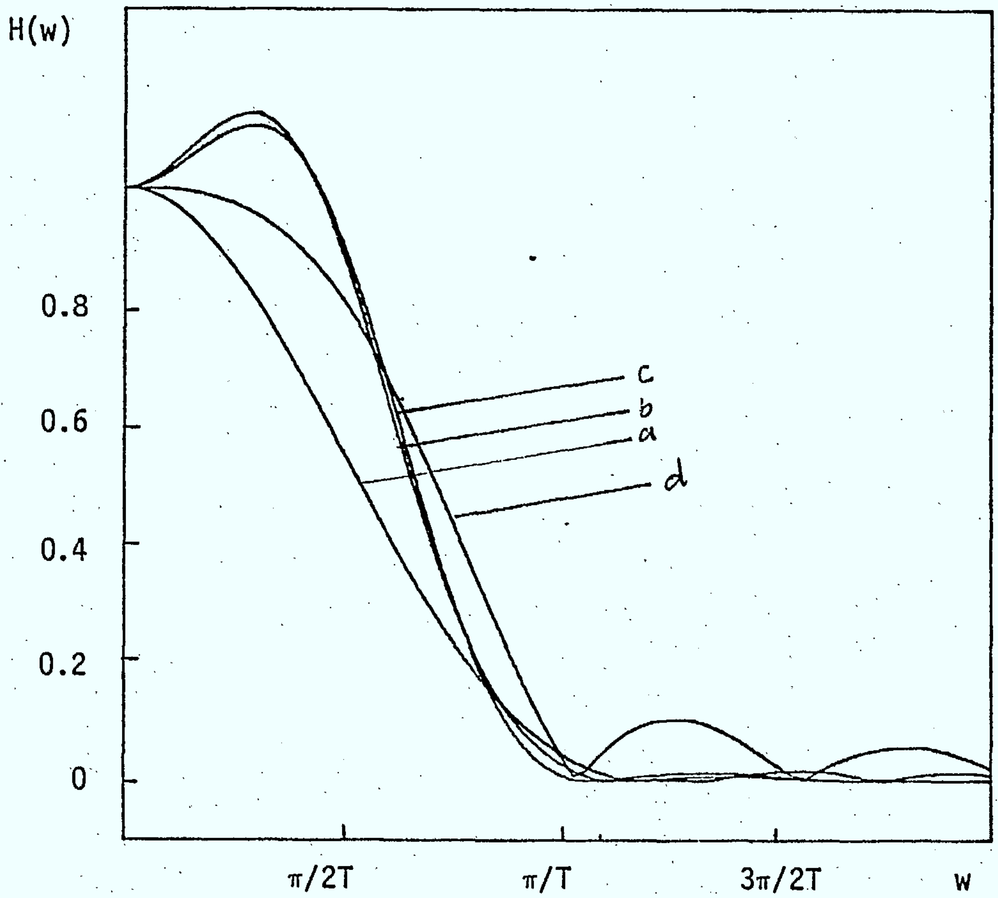


Figure 3.4 Frequency Characteristics of the Predetection Filters

(a) AMF (b) WAMF8 (c) WAMF6 (d) WAMF4

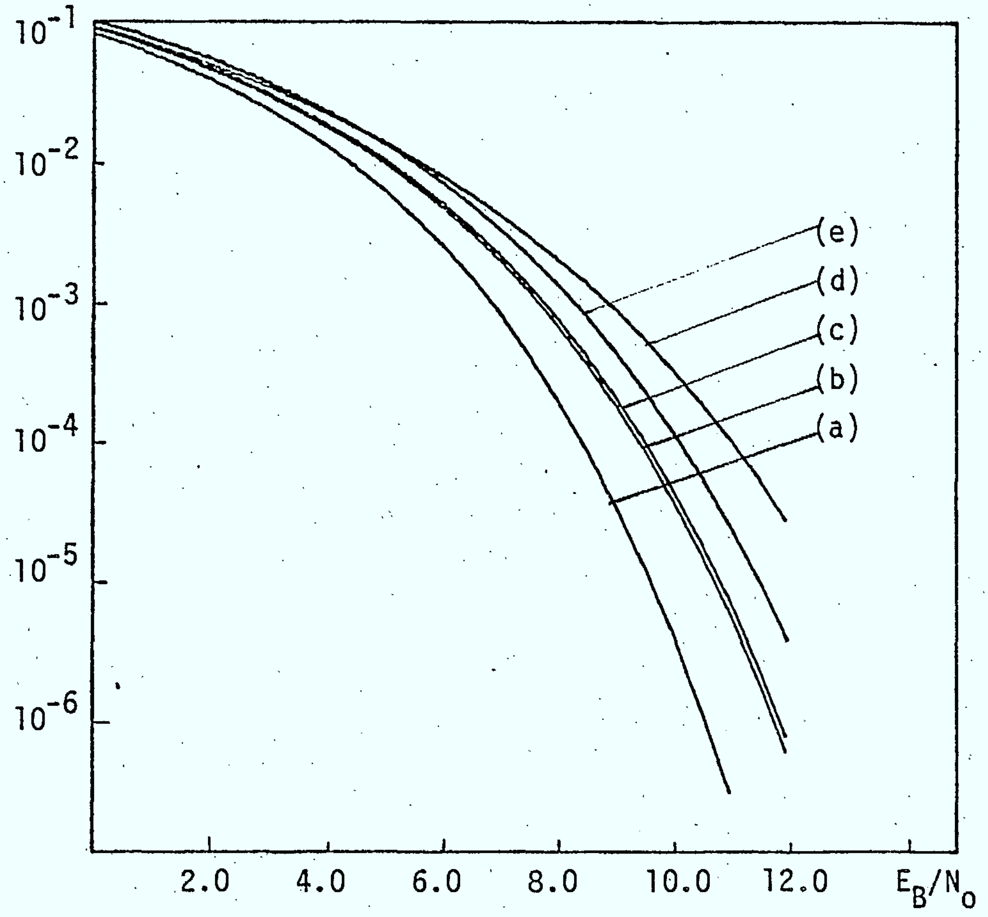


Figure 3.5a Symbol Error Rate for QPSK(a), and TFM With WAMF6(b), WAMF4(c), AMF(d), and A MSK Filter (e)

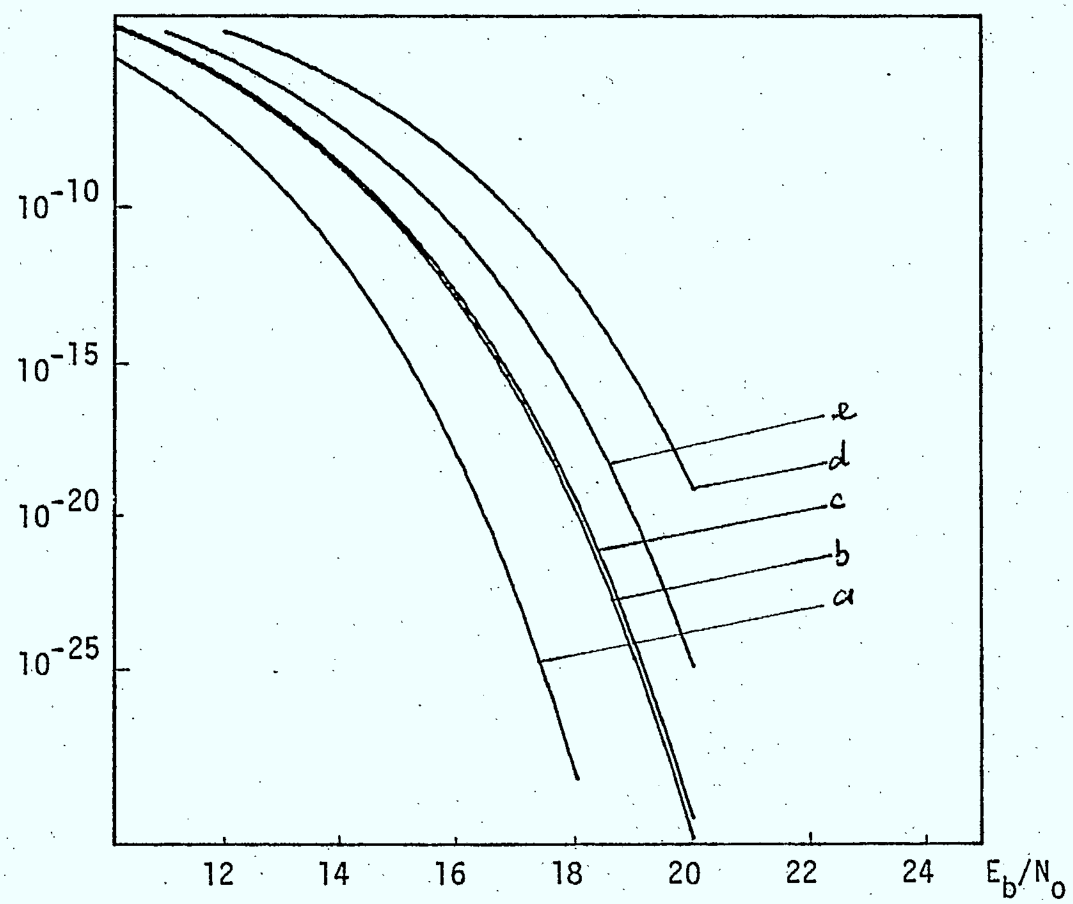
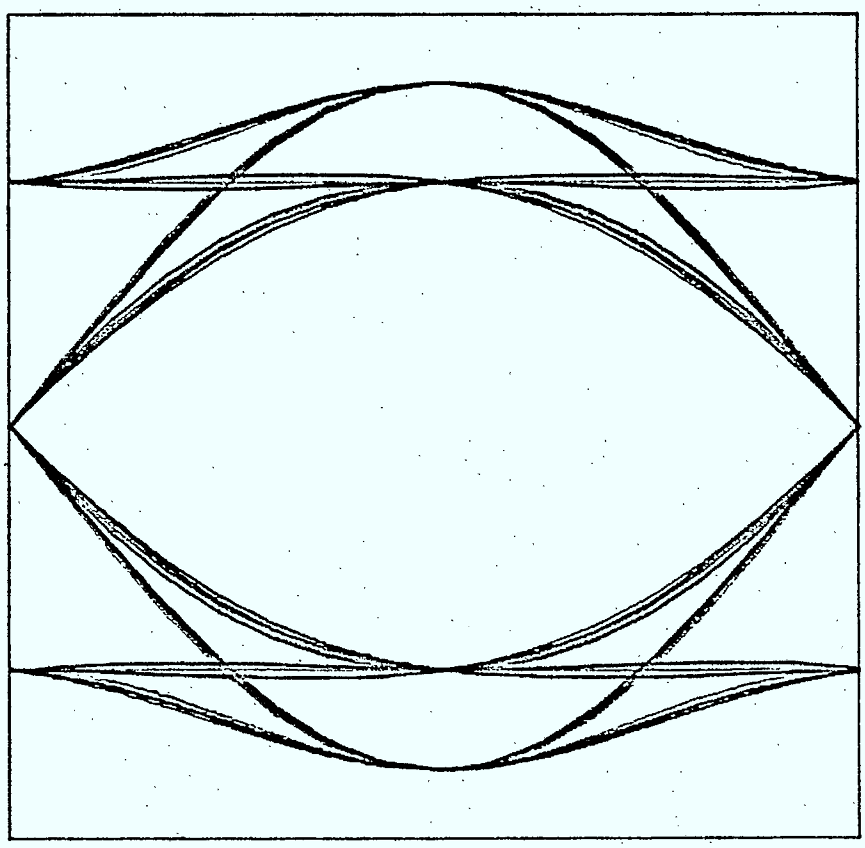


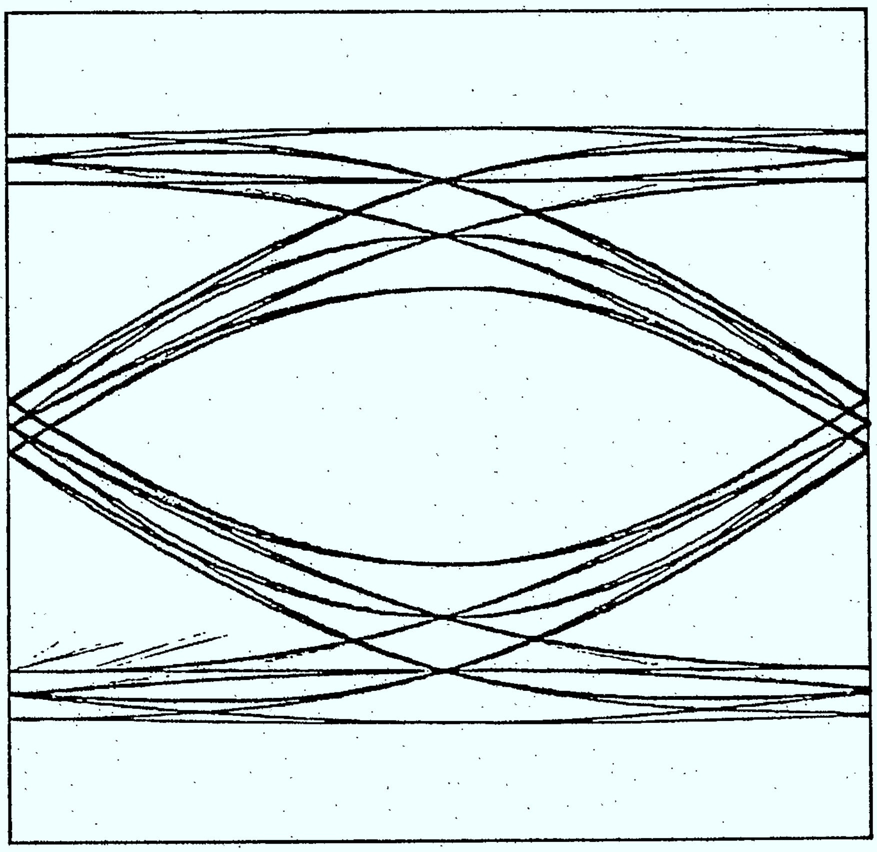
Figure 3.5b Symbol Error Rate for QPSK(a) and TFM With WAMF6(b), WAMF4(c), AMF(d) and A MSK Filter (e)



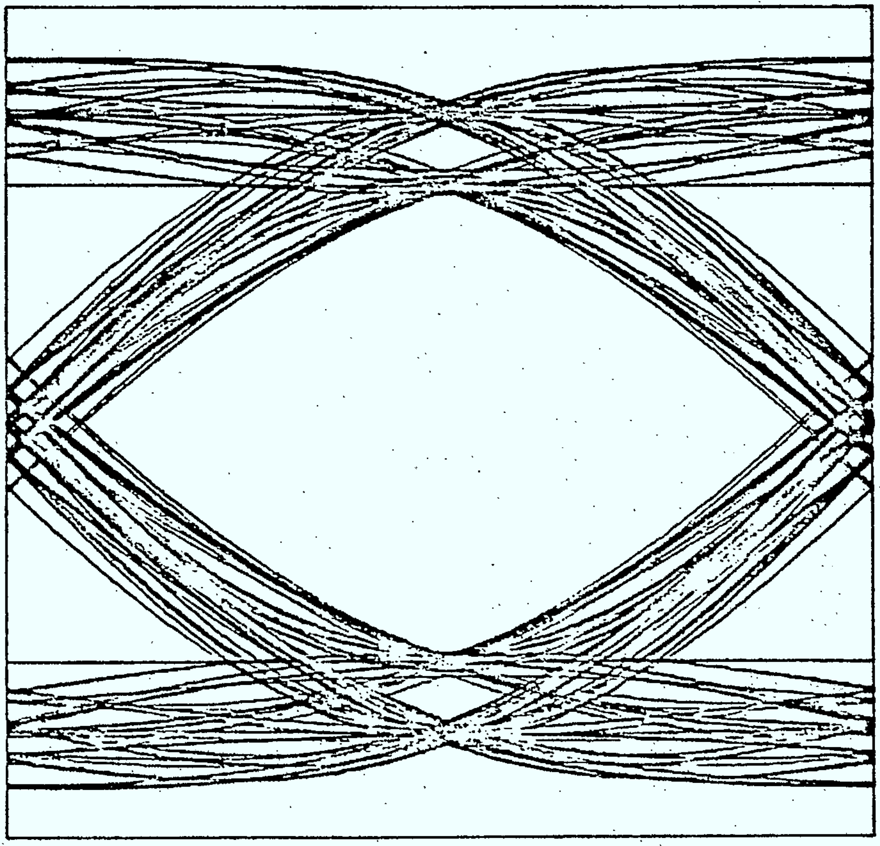
(a)

Figure 3.6 Eye Diagrams Obtained by

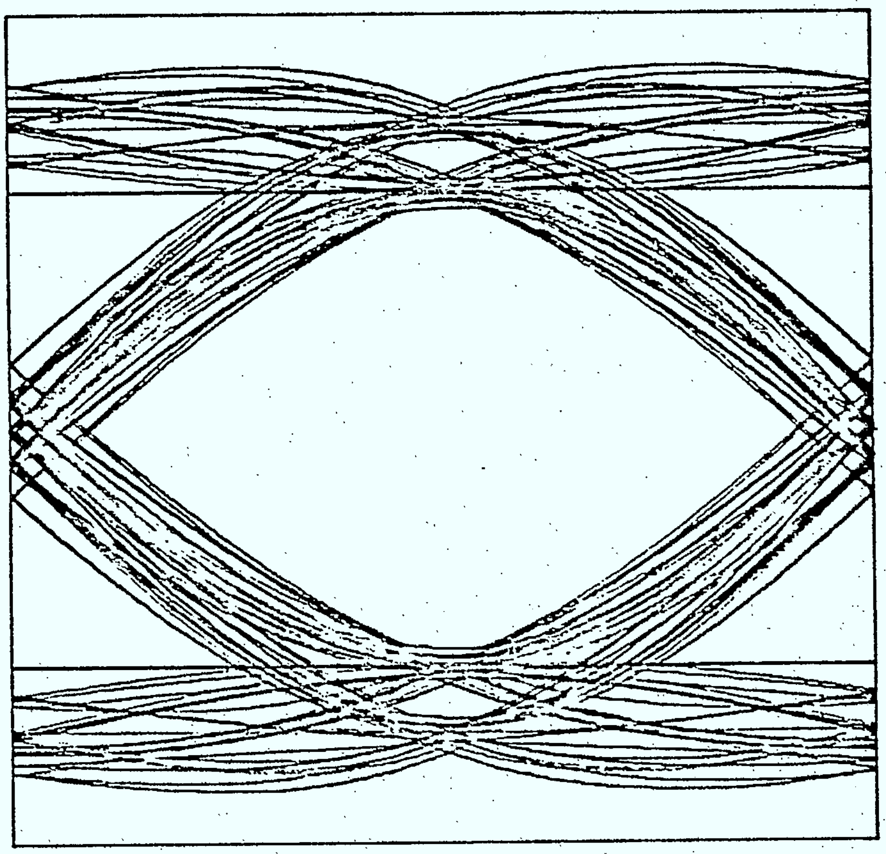
(a) A Very Wide Band Filter
(b) AMF
(c) WAMF8
(d) WAMF6
(e) WAMF4



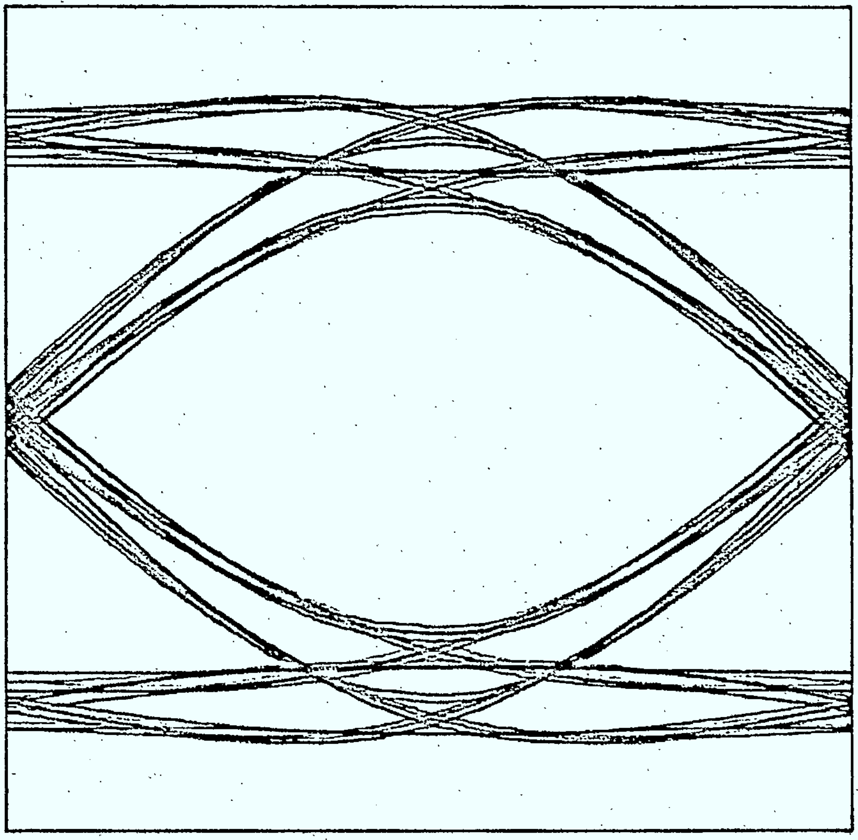
(b)



(c)



(d)



(e)

REFERENCES

1. R. de Buda "Coherent Demodulation of Frequency Shift Keying with Low Deviation Ratio" IEEE Transactions on Communications, June 1977, pp. 429-435.

2. F. de Jager and C.B. Dekrer "Tamed Frequency Modulation; A Novel Method to Achieve Spectrum Economy in Digital Transmission" IEEE Transactions on Communications, Vol. COM-26, No. 5, May 1978, pp. 534-542.

3. D. Muiljik "Correlative Phase Shift Keying - A Class of Constant Envelope Modulation Techniques" IEEE Transactions on Communications, Vol. COM-29, No. 3, March 1981, pp. 226-237.

4. T. Aulin, N. Rydbeck, C.E. Sundberg "Continuous Phase Modulation - Part II: Partial Response Signalling" IEEE Transactions on Communications, Vol. COM-29, No. 3, March 1981, pp. 210-225.

5. T. Aulin, N. Rydbeck, C.E. Sundberg "Bandwidth Efficient Digital FM with Coherent Phase Tree Demodulation" ICC 1979, pp. 42.4.1.

6. W.P. Osborne, M.B. Luntz "Coherent and Noncoherent Detection of CPFSK" IEEE Transactions on Communications, Vol. COM-22, No. 8, August 1974, pp. 1023-1034.

7. P. Galko, Pasupathy "On a Class of Generalized MSK" ICC 81, pp. 2.4.1 - 2.4.5

8. T. Aulin, C.E. Sundberg, A. Svenison "Modified Offset Quadrature Receivers for Partial Response Continuous Phase Modulation" TR-152, Telecommunication Theory, University of Lund, Sweden.

9. S. Lucky and Weldon "Principles of Data Communications", McGraw-Hill, 1968.

10. K. Murota, K. Hirade "GMSK for Mobile Radio Telephony" IEEE Transactions on Communications, Vol. COM-29, No. 7, July 1981, pp. 1044-1050.

11. J.B. Thomas "Statistical Communication Theory", John Wily, 1969.

12. D. Muiljik "Tamed Frequency Modulation, A Bandwidth Saving Digital Modulation Method Suited for Mobile Radio", Philips Telecommunication Review, Vol. 37, No. 1, pp. 35-49.

4.0 Mobile Speech Detector

4.1 Introduction

In normal conversations held over a duplex circuit, it has been found that more than half of the channel capacity each way is wasted as silence periods because the talker is either listening or simply pausing [1]. This unused capacity has long been exploited in TASI systems to dynamically concentrate incoming voice channels onto transmission trunks with less than half of the total incoming available-bandwidth [2]. It has been proposed that pauses in conversations held on mobile radio channels be similarly salvaged and allocated to the transmission of short data packets [3]. This requires a speech detector with the ability to measure and track the background noise level so as to provide an adaptable threshold above the background noise level. This is particularly important in the mobile radio environment, where the background noise from the vehicle, other traffic, general urban sounds, etc., varies with time. A speech detection algorithm has been proposed, implemented in real time on a digital signal processor, and tested with various combinations of speech and noise samples.

Section 4.2 reviews some properties of speech, presents assumptions behind our design, and reveals the algorithm proper. Section 4.3 documents the 2920-centered hardware and the software to run on the processor.

Section 4.4 concerns various aspects of performance evaluation: test bed; sample signals; subjective tests; and objective results from the "Activity Monitor", a microcomputer-based statistics collector. In the last section of this Chapter, we offer some comments about the design, suggestions for possible improvements, and possibilities for further research.

4.2 Main Concepts

4.2.1 Premises of the Design amidst existing Speech Detectors

The premises governing the design of the speech detector have been elucidated in a previous report [4]. However, they are re-iterated in the following discussion on the general design philosophy.

The mobile environment is highly variable and its acoustical backgound has not been well characterized. "Noise" to the speech detector is anything other than the desired talker's speech. This could be in the most unfortunate case a third person's speech or speech from a car radio. A passive device, the detector can do nothing about this; so the design of the audio equipment of a mobile transceiver plays an extremely vital role on the effectiveness of the speech detector. The following features are definitely desirable:

(a) The microphone should be insensitive to mechanical vibration. Impulses generated from a microphone in sudden motion are easily mistaken for speech. Furthermore, precautionary measures must be taken against sympathetic vibration both for the microphone's transduction element and the mechanical support for it. Any push-to-talk switch should be decoupled from the audio circuit. An example of a mechanically rugged microphone is of the "dynamic" or "moving coil" variety used in the AMPS control unit [5].

(b) Close-talk noise-cancelling microphone with good directivity is highly desirable. Unfortunately, neither the ubiquitous carbon microphone on most telephones nor the dynamic microphones on AMPS control units belong to the category of "pressure-gradient" or "velocity" microphones [6] that offer such noise rejection capability.

(c) Acoustic feedback on speaker-phones (much like echos in telephony transmission) is considered more a problem of audio system design and out of the context of out present endeavor. Speaker phones are not common in mobile radio anyway.

(d) Proper shielding and grounding is necessary to prevent RFI from being "demodulated" into the audio channel fed to the speech detector by various non-linear circuit elements such as the base-emitter junction of a transistor.

Whereas in our problem the noise background is highly unpredictable, much that is known about speech in general is not particularly sensitive to application specifics. It is known that conversation patterns are functions of the purpose for which it is held and of the conversation environment in which it is held. For instance, talkers in a noisy setting tend to raise their voices and to be more articulative [7]. Also, it would be expected that in situations that demand attention (such as driving), the talkers will speak less and only do so when communication is really essential [28]. Not withstanding all these variants, speech after all must maintain certain characteristics in order to be intelligible. For the speech detector, it suffices to treat speech as a random series of "high" and "low" level segments of length comparable to a syllabic interval (typically 150-350 msec.). This random sequence of high and low level utterances gives speech its characteristically high short-term dynamic range.

Moreover, even though tones that are "buried" in noise could sometimes be audible [9], contextual speech must in the main be above the background noise in order to be intelligible [7]. So one can assume that very little is lost if speech too heavily corrupted by noise to be comprehended is cut off by the speech detector. Of course, in the case where the detector requires more power than the intelligibility-threshold-above-noise, degradation results.

The signal bandwidth has a bearing on the selection of detection mechanisms. In our work, we confined our scope to standard telephony speech, which has a passband from about

200 Hz to 3300 Hz. This implies that a significant amount of spectral energy for unvoiced and whispered speech is cut off.

Prior to presenting the detection algorithm, it would be of interest for the readers to note that references [10] through [20] cover previous work on speech detection, mainly in the fairly quiet non-mobile environment.

4.2.2 The Algorithm

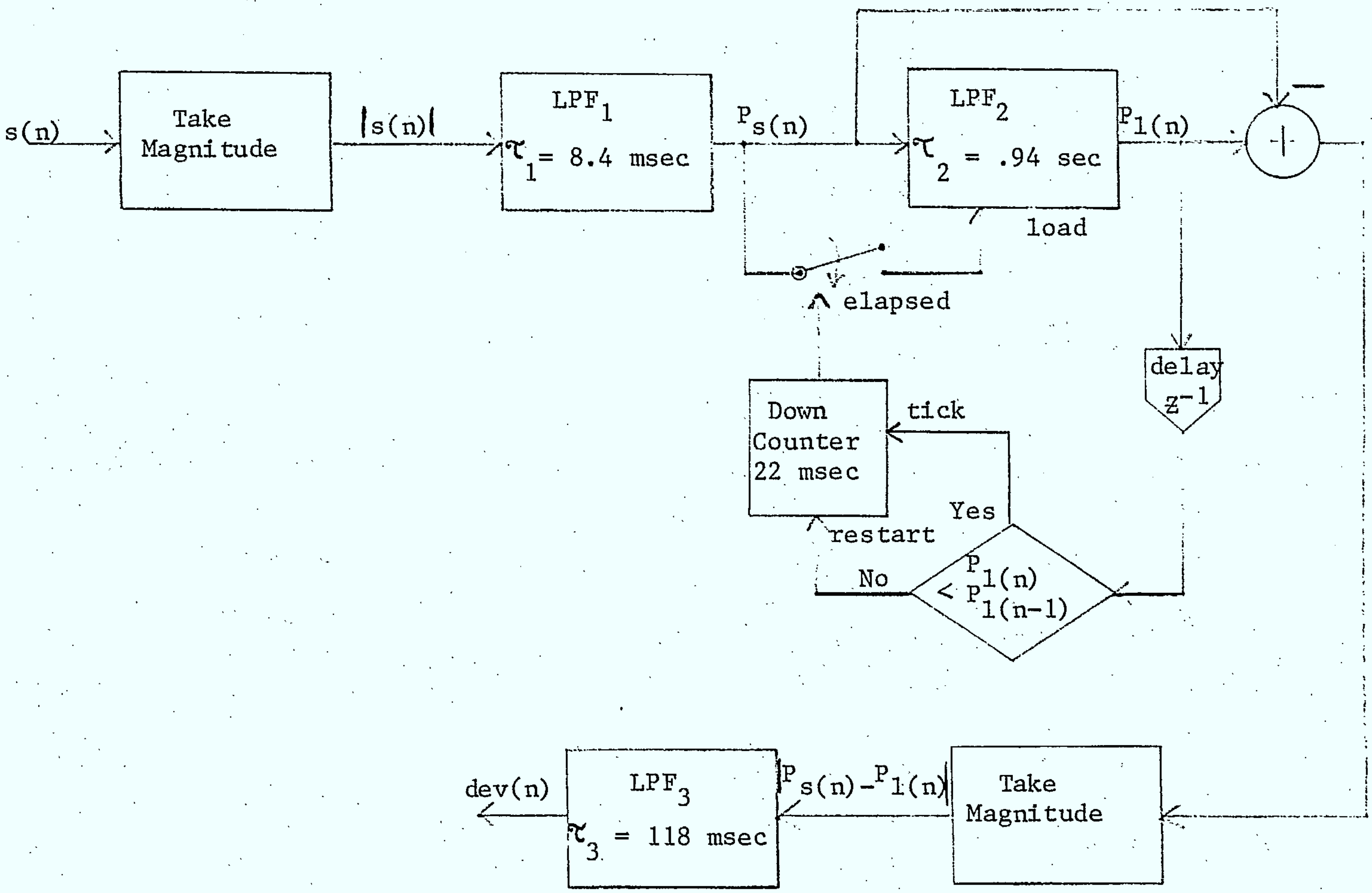
The speech detector declares speech to be present when the magnitudes of two successive speech samples exceed the current detector threshold which is denoted CT. Following any such threshold exceedance, active speech is declared for a minimum of 200 msec., the "hangover" period, so as to bridge momentary gaps or very low-level portions of an utterance. The current threshold CT is updated so as to be sufficiently above the current noise level estimate CNLE. CNLE, in turn, is updated by identifying intervals when the incoming signal is "almost surely noise" (ASN), and measuring the long term signal power during such intervals.

Specific behaviour of low-bit-rate voice coders have been exploited to detect the presence of speech [10, 11]. Detectors of this kind published require hardware in addition to that of the coder; therefore coding-specific detectors are feasible only if the speech coding scheme is fixed. For generality, we decided not to tie our detection scheme to a particular low-bit-rate coding technique. Our detector would simply work on PCM samples. Another disadvantage of coding-specific detection algorithms is that they cannot be easily transported to systems with other coding schemes.

In our speech detection algorithm, a signal sample is processed by a series of three single-pole, unity gain, low-pass filters to obtain three measures: $P_s(n)$, $P_l(n)$, and dev(n) which we called respectively "short-term power", "long-term power", and "deviation", as shown in Figure 4.1. If $P_l(n)$ has been dropping monotonically for 22 msec. it is reset to $P_s(n)$, a "valley seeking" mechanism. This serves two purposes:

(a) $P_l(n)$ rises when there is speech, albeit with a large time constant. If the detector should ever make the mistake of declaring ASN within an utterance,the threshold, being proportional to $P_l(n)$ will be set with a high value; clipping will subsequently follow! "Valley seeking" effectively pulls down $P_l(n)$ in intersyllabic gaps, e.g., during unaccentuated sibilants.

(b) Without the mechanism, it would take on the order of 1 second ($\cong \tau_2$, Figure 4.1) to reach the "possibly noise" (PN) state (described below) at the end of a spurt, lessening the chance of updating the threshold between spurts.



All low-pass filters (LPF) are of the form:

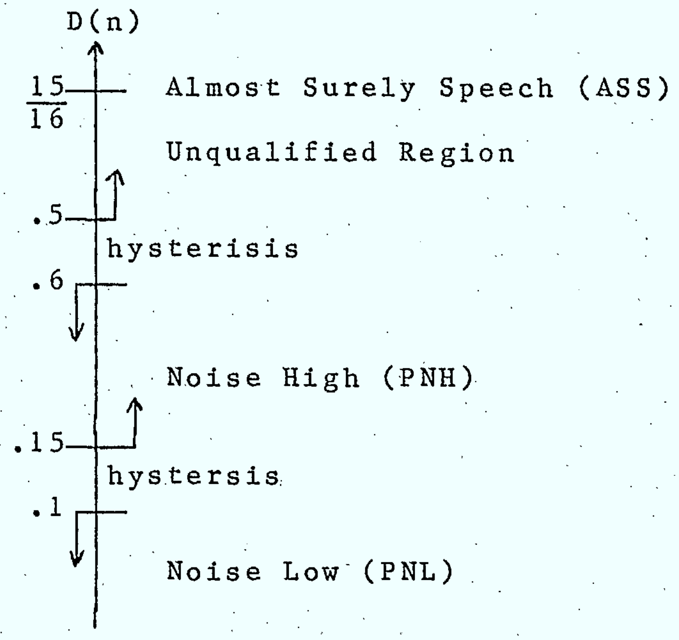
$$\text{output}(n) = \text{output}(n-1) + \alpha \cdot [\text{input}(n) - \text{output}(n-1)]$$

where $\alpha = T \cdot fc$ = (sampling period)(cutoff frequency)

Fig. 4.1 Derivation of $P_s(n)$, $P_1(n)$, and dev(n)

$D(n) \triangleq \dfrac{dev\ (n)}{P_1(n)}$ has been found to be a very effective measure of short-term signal variability. It is the averaged magnitude of $P_s(n)-P_1(n)$, normalized by $P_1(n)$.

Depending on its value, the signal is qualified as follows:



Note: "Possibly Noise" PN = PNH + PNL

Also, we have "Almost surely Noise".

$$ASN = \quad (PN) \text{ and } (P_s(n) \quad CNLE)$$

where CNLE is "Current Noise Level Estimate".

As shown in Figure 4.2, when ASN is true, CNLE is loaded with the current value of $P_1(n)$ and δ is loaded with a fraction, b or $P_1(n)$. If ASN is false, CNLE is incremented by δ. The rationale behind this mechanism is to test the slope of $P_s(n)$

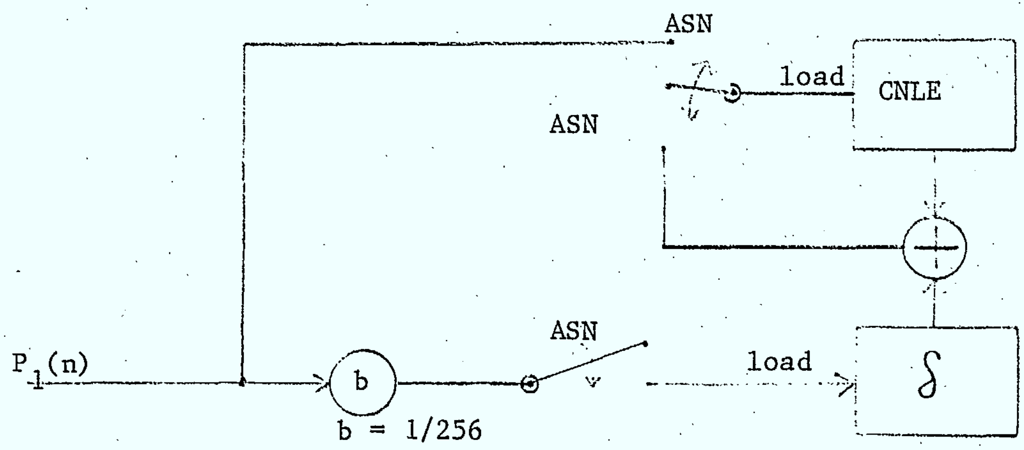


Fig. 4.2 CNLE Mechanism

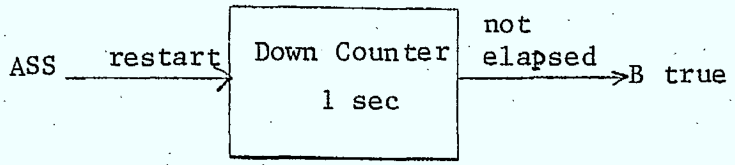


Fig. 4.3 Blanking Interval Generator

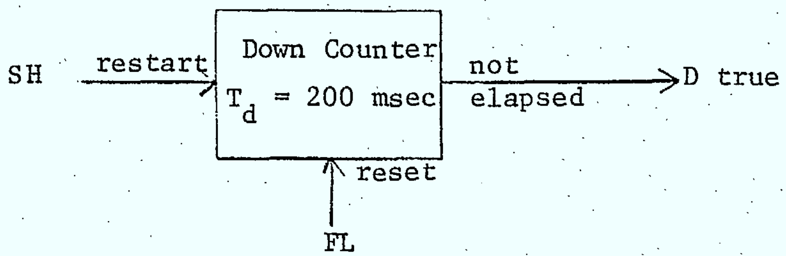


Fig. 4.4 Threshold Leakage Timeout

at a gap-spurt transition. In fact, we could dispense with this mechanism were it not for the long attack time of D(n). This mechanism is sufficiently fast in taking the detector out of the ASN state at the onset of speech to ensure that the threshold is only a function of $P_1(n)$ without speech. When CNLE eventually "catches up" with $P_s(n)$, PN should no longer be true. Since CNLE rises indefinitely, $P_s(n)$ is unlikely to exceed it; therefore the next declaration of ASN relies solely on D(n).

The Current Threshold setting, CT, can assume a number of values, depending on the state of D(n) when the threshold is loaded. Before considering their derivation, let us define several auxiliary functions. Whenever "almost surely speech" (ASS) is true, a blanking interval of 1 second is setup as shown in Figure 4.3, during which a variable B is true.

When B is true. CT can only assume a value T_{min}. Now define the following logic equations:

"Set High" $SH = (ASN)(\overline{B})(PNH)$

"Set Low" $SL = (ASN)(\overline{B})(PNL)$

"Forced Low" $FL = ASS$

Now define a counter to timeout a period (T_d) in which CT, previously set to T_H (to be defined in Fig. 4.5), is decremented until reaching T_h as shown in Figure 4.4.

Finally, CT is derived as shown in Figure 4.5. At every clock cycle, 5 things can happen to CT:

1) Set CT to $T_H \triangleq A_H P_1$ if ASN, D(n) large, and the blanking clock in Figure 4.3 is not ON (B not true).
2) Set CT to $T_L \triangleq A_L P_1$ if ASN, D(n) small, and blanking clock is not ON.
3) Leaks towards $T_h = A_1 P_1$ if the down-counter in Figure 4.4 is still running.
4) Forced to $T_{min} = A_{min}$ (n=latest ASN instant) if ASS.
5) Not affected if none of (1), (2), (3), (4) is true, i.e., $(\overline{ASN})(\overline{ASS})$.

Lastly, the speech detector enters the ON state whenever $|x(n)| >$ CT for N consecutive samples as shown in Figure 4.6 200 msec. for the hangover is believed enough to cover unvoiced consonants, sibilants, and other low-level phonemes.

It was initially thought that with severely limited bandwidth at the high end, zero-crossing counting would not be effective enough to detect low-level unvoiced speech that is most likely to be missed by the theshold but is likely to be covered by the hangover period. The idea of zero-crossing counting is to classify the incoming signal as speech if the speech samples alternate in sign for a number of samples.

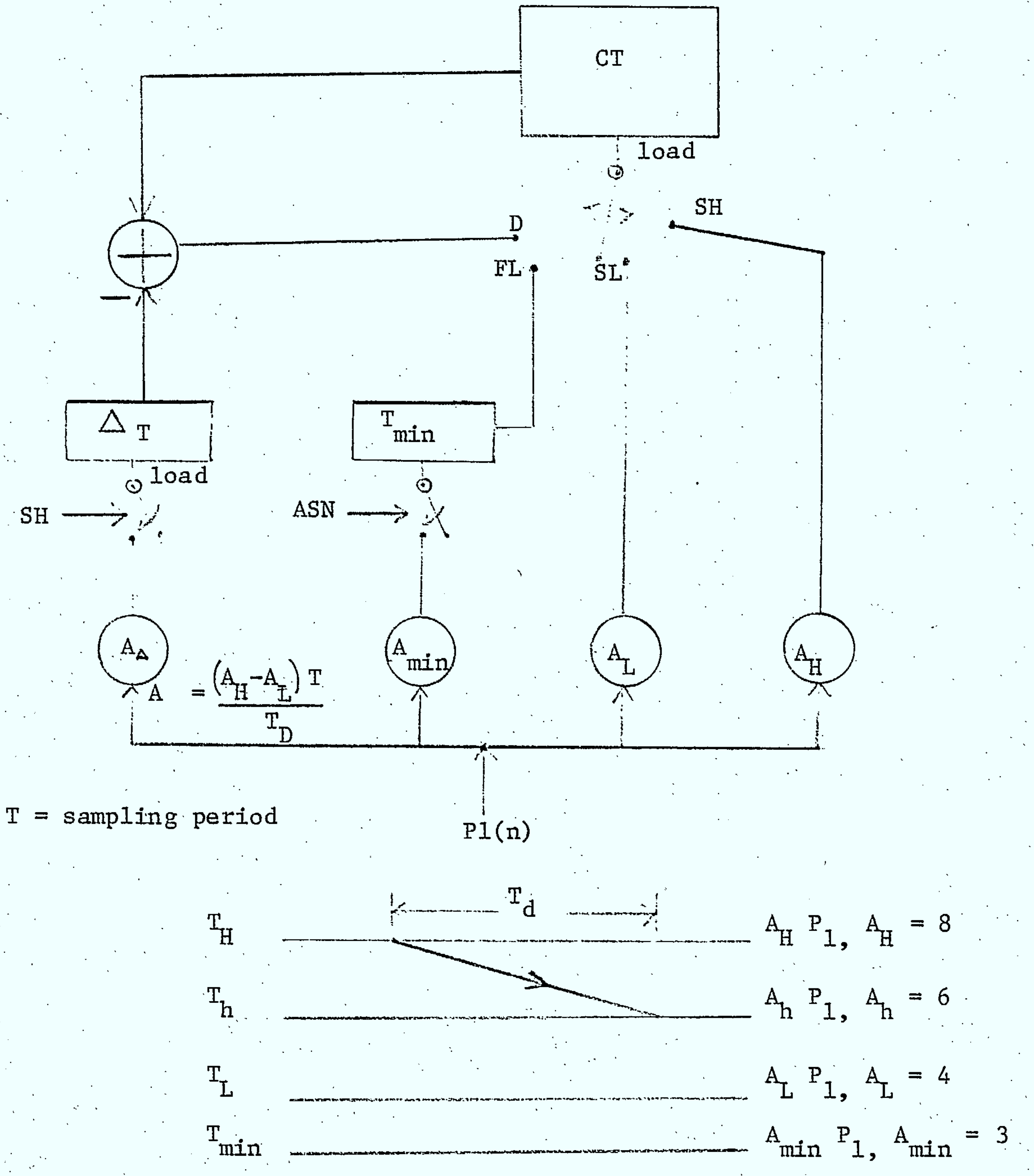


Fig. 4.5 Threshold Setting Logic

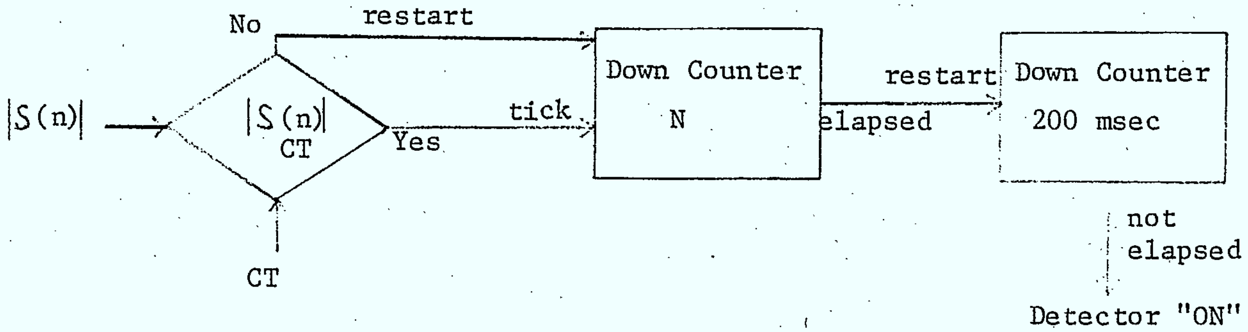


Fig. 4.6 Hangover

greater than or equal to X. As pointed out by Fariello [12], due to the complex interaction between sampling and a stochastic bandpass signal, this X is virtually an experimental number. We found that X is lower bounded by the maximum tolerable false-alarm rate when the input is Gaussian random noise from a noise generator. X would be lower if it were optimized for the recorded mobile noise; however, for worse case design, we elected the generator noise to determine X.

The speech detector normally operates with X=10. When this value is attained, the condition ASS is asserted; the blanking-clock depicted in Figure 4.3 is set to run for a quarter of a second; CT is set under the condition FL; and the hangover counter in Figure 4.6 is activated. Under two different situations X=7 is used: (1) PNH; (2) the blanking-clock is running. The actions following detection under the former condition are the same as when X=10 above. However, under the latter condition, only the hangover counter is affected on detection. This is to avoid positive feedback in which X=10 is first exceeded, turning on the blanking-clock and hangover counter and lowering X to 7, with the entire operation repeated indefinitely now that X is smaller.

4.2.3. Parameter Values

In this section, we attempt to explain the rationale behind the parameter values of the speech detectors, even when these values were emprirically derived. Some feel for these parameters is essential if they are to be modified.

The time constant of LPF_1, is comparable to the longest pitch period; thus this filter smoothes out the fine structure of quasi-periodic voiced speech. The time constant for LPF_2 is a compromise between the contradictory requirements to track the rapidly varying noise level and to maintain the noise-only level in the middle of a speech spurt, a failsafe measure against clipping when the threshold is inadvertently set. Since English speech is composed of consonant-vowel-consonant (CVC) segments, LPF_2's task is mainly to hold the noise only level over a syllabic interval until the valley seeking mechanism comes into play. The 22 msec. setup time for the valley seeker need not be very precise, but it should be longer than the longest possible pitch period so that the decaying envelop of the vocal-tract's impulse response would not be mistaken as a transition to a gap or low-level utterance. One last point from Figure 4.1 is that LPF_3's time constant is of the order of a syllabic interval; thus D(n) does measure dynamic range at a syllabic rate.

The CNLE mechanism (Fig. 4.2) has been discussed in the last section. The parameter b controls the frequency at which the threshold is modified. Ideally, b should be such that CNLE rises no faster than the smallest rate of change of short-term speech level at the onset of a spurt. In the present implementation b = 1/256 corresponds to .28 dB/msec., which is slightly smaller than the smallest value .3dB/msec. quoted in reference [20].

The various classification boundaries for D(n) are given purely experimental values. These decision regions determine the manner by which the threshold is set rather than the ON/OFF state of the detector.

The blanking pulse (Figure 4.3) dictates the amount of time after the latest declaration of ASS that the lowest threshold level must be used. It has several effects:

(1) Low level phonemes within a spurt will likely be detected.

(2) It has a hangover extension effect.

(3) Effects (1) and (2) combined together to smooth out short gaps (greater than a hangover period) in a "continuous" speech segment. The speech detector acts as if it is looking for phrases rather than individual energy packets.

(4) In noisy speech, spurts are lengthened by possibly as much as the residue blanking time after the end of a spurt. Likewise, false activation could last longer if a false alarm sets both the hangover counter and the blanking clock.

With reference to Figure 4.4 and 4.5, both T_d and A_Δ can be adjusted to obtain the desired leakage time T_d and steady state threshold T_h. T_h can be made to be equal to T_L, which appears to be the bottom line for T_h. This arrangement can be likened to minimum mean-squared estimation for a Markov process in which the future estimate given a present sample of the process leaks exponentially toward the mean. In fact, Markovian models for speech patterns have been reported [21-22].

The hangover period grants passage to low-level phonemes. Most non-vowel sounds are usually of shorter duration than vowel sounds, so the hangover does not have to be so long as to cover the longest syllabes in speech. 200 msec. seems to be quite adequate - most existing detectors use smaller hangovers [20].

The sample-above-threshold counting technique can be replaced by comparing the threshold with the output of a LPF that has very small time constant (less than 1 msec.) but the counting technique appears simpler. In general, there is a tradeoff between the threshold level and the number of consecutive samples (or number of samples in a frame) required to exceed the threshold in order to actuate the detectors. The 2 samples used in our detector seems quite marginal intuitively; however, we found that noticeable clipping begins at 4 samples. Thus, unless the thresholds are lowered, no more than 3 samples should be used.

The parameters associated with the zero-crossing detection mechanism have been described in the previous section. As this mechanism was introduced into the detector rather belatedly, we now believe that a plausible improvement is to count the number of zero crossings in a frame of samples. This method does not actually require samples to be stored in frames and it has the virtue of dispensing with the consecutivity requirement, which relies on some obscure relationship between sampling and the frequency content of the incoming signal. The detector will be upgraded with this later on.

4.3 Circuit Details

The majority of custom circuitry in our speech detection test bed, described more fully in Section 4.4, is for monitoring purposes; we shall not describe it here. The amount of hardware devoted to the detector per se is quite minimal and a large part of it is just the "application side" of a SDK2920 development kit. Fig. 4.7 shows a block diagram of the detector hardware. Fig. 4.8 is a circuit diagram for the portion of Fig. 4.7 that belongs to a SDK board; the relevant circuitry is delineated in the diagram.

The input signal to the detector swings between + and - one volt. This signal first passes through the 2912, an anti-aliasing filter that conforms a telephony standard for PCM-codec filter response. A sampled analog filter, the 2912 requires a clock of a few hundred kHz - precise frequency is not important - which is supplied by the 74LS674 oscillator. The output of the filter is fed to channel 0 (pin10) of the 2920. The 2920 requires a capacitor for its sample-and-hold (C34). Its A/D convertor requires a stable voltage reference (R41 pot). Its clock is derived from a 6.67 MHz crystal (1/2), with which the instruction time amounts to 600 ns and the full-program sampling time amounts to 115.1424 micro-sec. The detection algorithm outputs a control signal via output channel 1 to the voice switch (Fig. 4.9), which is nothing more than a buffered 4066 analog switch. The buffer

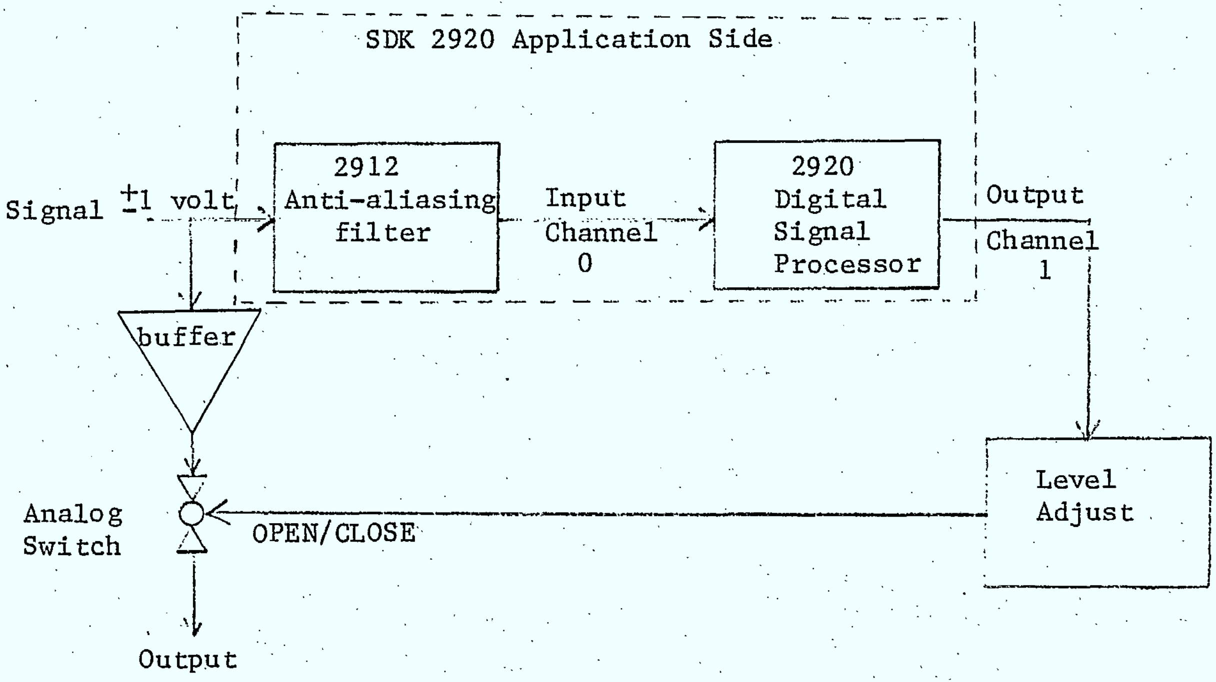


Fig. 4.7 Basic Speech Detection System

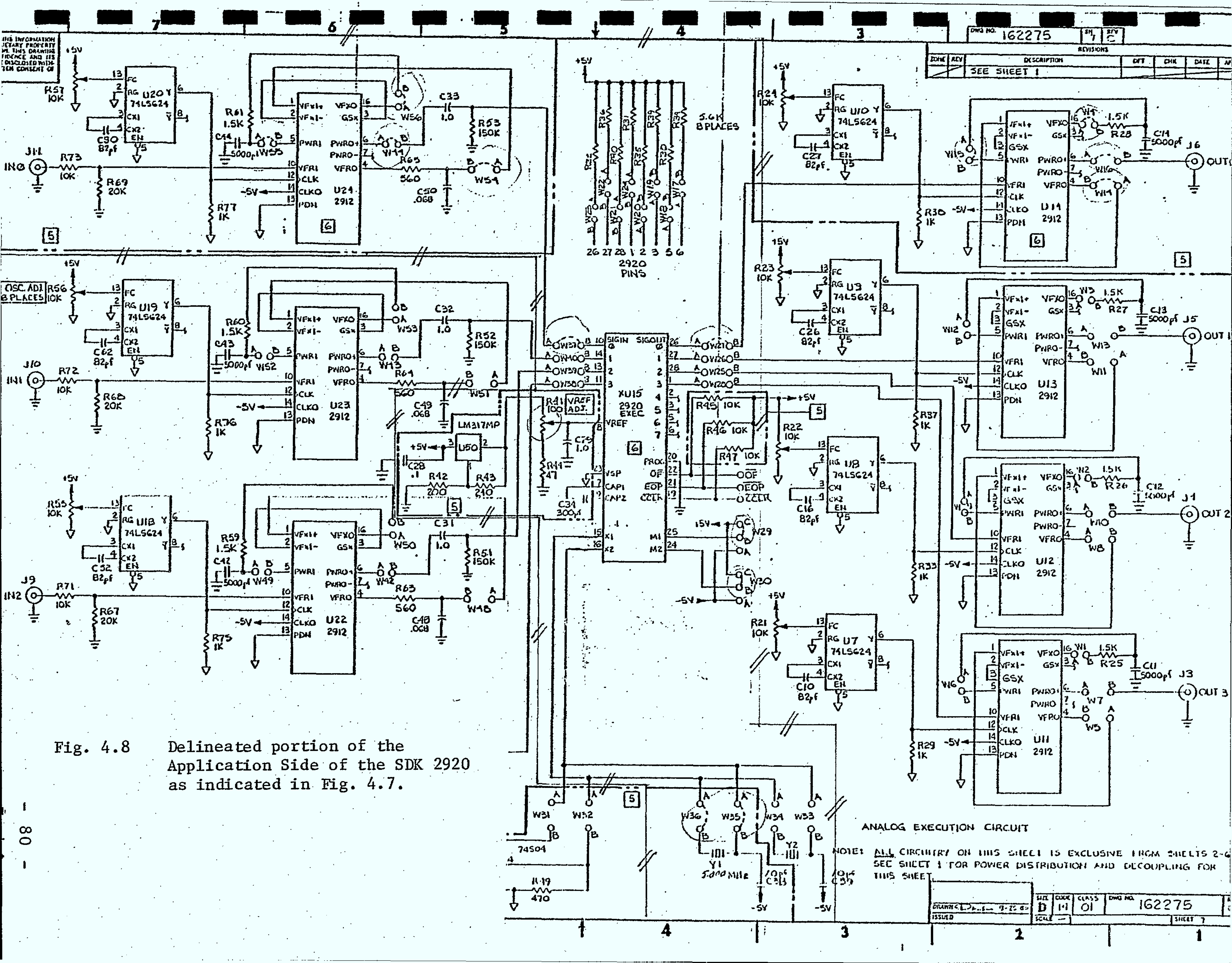


Fig. 4.8 Delineated portion of the Application Side of the SDK 2920 as indicated in Fig. 4.7.

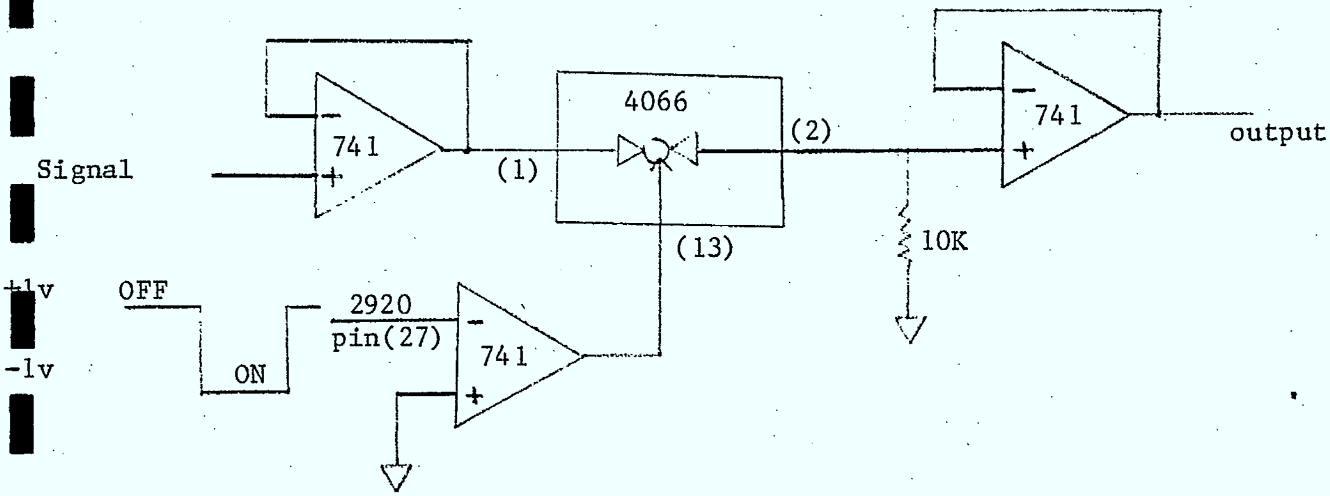


Fig. 4.9 Voice Switch

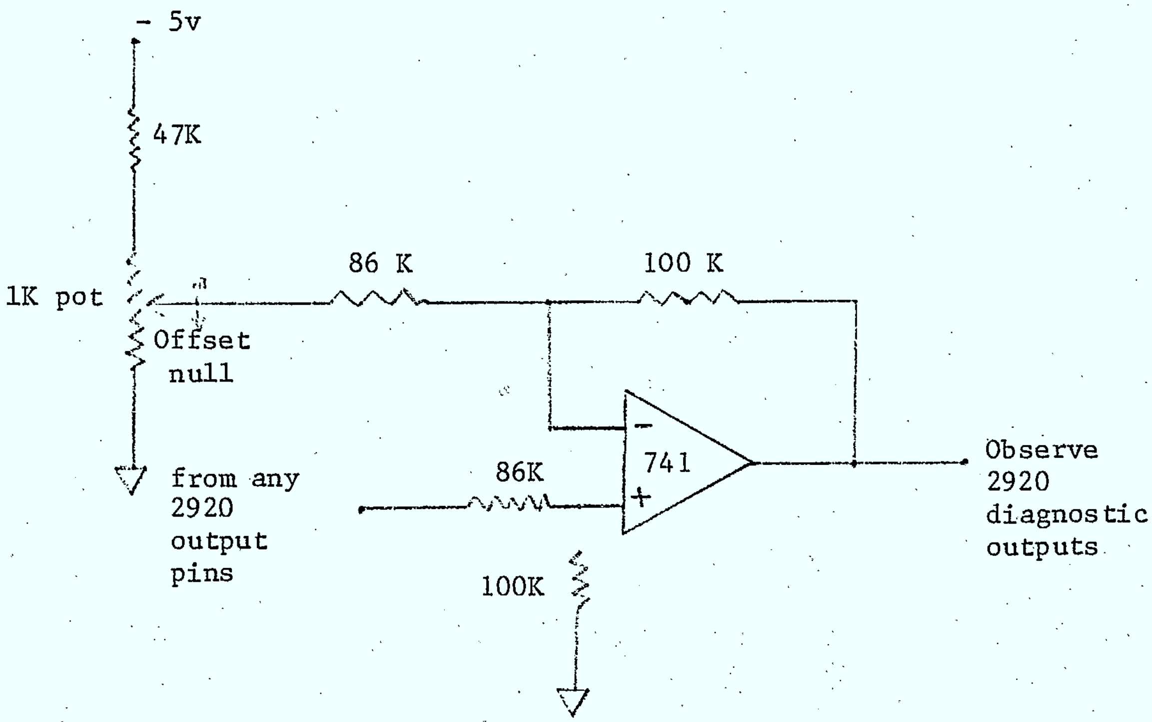


Fig. 4.10 Diagnostic-Output Interface Circuit

isolates the switch from the audio input to the 2912 and is absolutely indispensible in preventing feedback of the switching pulses into the speech detector. The switched audio signal can be optionally picked up from output channel 0, which is connected to a 2912 reconstruction filter (Fig. 4.8). This signal is corrupted by the A/D and D/A conversion processes, thus the analog-switched signal is "cleaner". All the other output channels of the 2920 except channel 7 are used to monitor internal variables, which are either inherent to the algorithm or are created for the monitoring the detector. Channel 7 is wasted as a "cross-talk purging phantom channel"---reference Intel's "bug notes". The output channels are not only offset from the inputs, but they are also scaled down--------an internal zero does not correspond to zero volt and an internal 1 does not yield 1 volt. So to facilitate referencing the output channels with respect to the inputs, the offset is nulled and scale is readjusted by the circuit shown in Fig. 4.10.

Fig. 4.11 summarizes some features of detector's 2920 - code. Note that slightly less than two-thirds of the ROM is used to implement the detector. The remaining space is filled up with digital computation and analog output instructions for monitoring. That the ROM is not fully utilized is due to the simplicity of the filters. If they had been upgraded to Biquad sections, the ROM occupany would

ROM

Program Size:	192 instructions (Full ROM)
Speech Detector Code:	120 instructions
Remaining ROM space:	output internal variables

RAM

27 Out of 40 locations used

I/O

IN0:	Signal Input
OUT0:	Switched Signal Sample
OUT1:	Voice Switch Control Signal
OUT2:	$D(n) = dev(n)/P_1(n)$
OUT3:	Blanking clock
OUT4:	ASN
OUT5:	Current Threshold-gain in effect
OUT6:	Length of contiguous alternating-sign sequence detected

Timing

All filters and counters assume a 6.67 MHz clock or a sample period of 115.1424 micro-sec.

Fig. 4.11 A Summary of Detector Code

at least have surged up by 40 instructions. Thus the chip currently supports 1 input channel and 7 output channels------analog operations determine the size of the program. Also note that only 27 RAM locations out of 40 are consumed. In fact, this figure can be drastically reduced if more of the temporary variables are reused at various parts of the program.

4.4 Performance Evaluation

4.4.1 Test Bed

The speech detector algorithm was developed and tested with a programmable digital signal processing device, the INTEL 2920. It takes analog inputs and outputs and has on-board A/D converter, D/A converter, RAM, EPROM and a capability for arithmetic operations performed under program control. Thus, it provides a flexible vehicle for implementing, testing and modifying speech detector algorithms. Because of its capability for input and output on several analog channels,signals present at several points in the system can be monitored.

Development and testing of programs for the 2920 is facilitated by the SDK 2920 system development kit. In a typical development cycle, a program for the 2920 signal processor is created and edited on an MDS development system. The program is assembled to generate machine codes which are subsequently downloaded onto the development section of the SDK 2920 board. The program can then be loaded into the 2920's EPROM (Figure 4.12).

To execute a program, the chip is transferred to the application section of the SDK 2920, which is connected to signal sources and monitoring equipment as depicted in

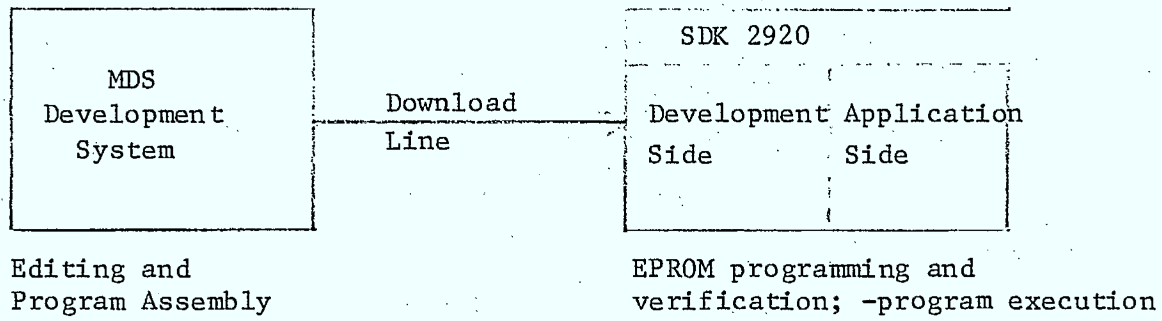


Fig. 4.12 Development Cycle

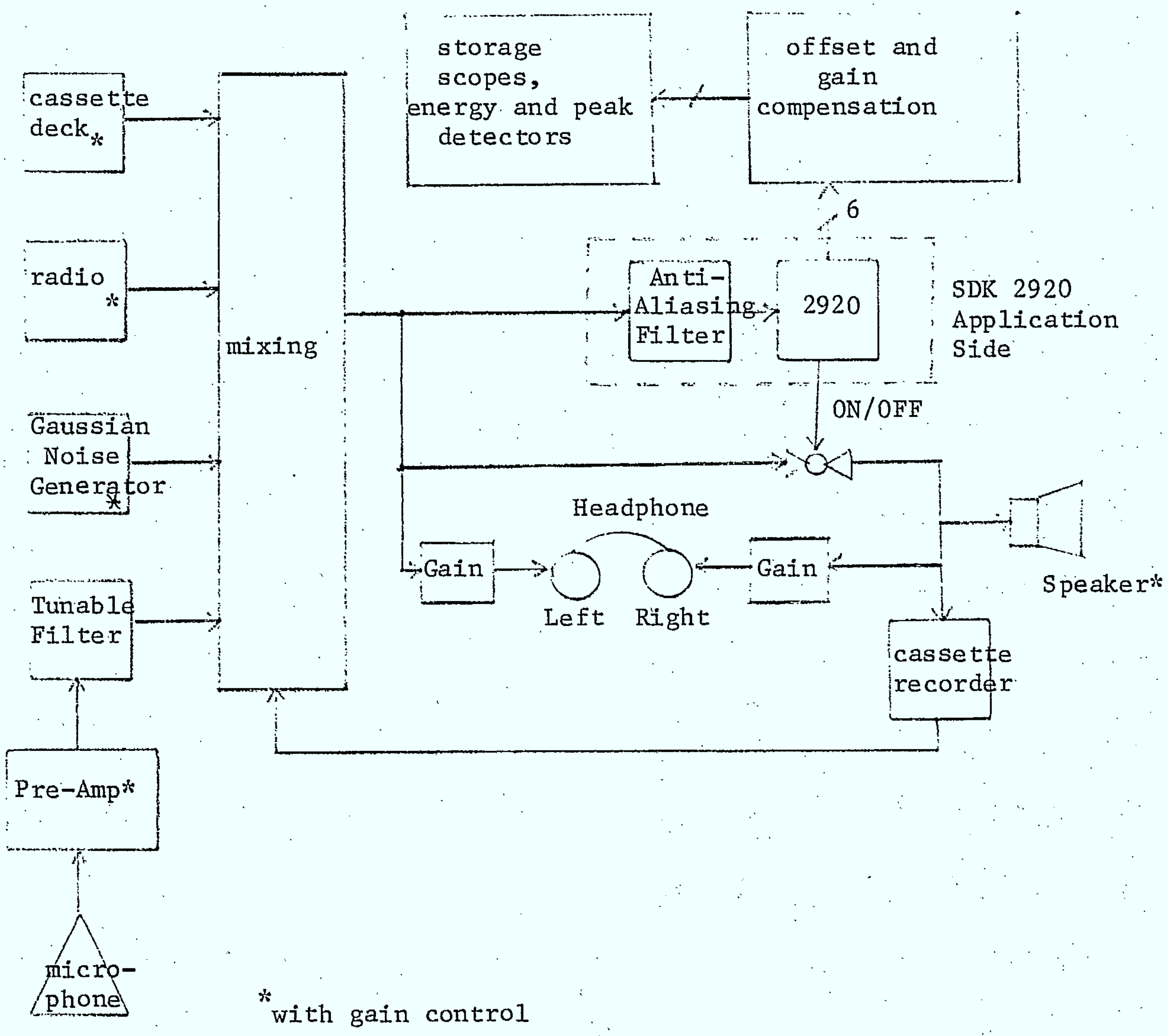


Fig. 4.13 Test Bed

Figure 4.13. It can be seen that the input to the detector is fed to a bidirectional bipolar analog switch controlled by an ON/OFF logic signal from the detector. A stereo headphone driven by the input and output of the switch, separately to the left and right earpieces respectively, enables constant monitoring. ON/OFF effects of the switch can be heightened by carefully adjusting the gains feeding the earpieces. Selected variables in the speech detector algorithm can be examined on the remaining output channels of the 2920. The outputs are monitored with peak detectors and storage scopes.

4.4.2 Desired Sample Signals

Presumably located at the output of the audio front end (microphone, pre-amplifier, and possibly gain control) of a mobile transceiver, the speech detector sees only the electrical transcription of an almost purely acoustical signal. With proper circuit design and shielding, the presence of a RF stage and all of its manifestations such as RFI are irrelevant to the detector. Furthermore, the input signal contains only the mobile talker's monologue, be the channel half-duplex or full-duplex----a speech detector is quite likely of dubious value to a push to-talk half-duplex channel. Echos, an impairment prevalent in long-haul wired telephony, are deemed non-existent as long as speaker-phone-like devices are used for reception----they are, in fact, not used in full-duplex systems for fear of acoustic feedback.

In essence then, the input to a speech detector is the monologue of the near-end talker engaging in a conversation over a full-duplex mobile channel; such monologue is corrupted by the surrounding acoustical noise.

4.4.3 Acquisition of a Sample

We first attempted to record such a monologue at DOC's monitoring station on Albert Street in Ottawa. But we could not obtain realistic samples because:

i) Channel-noise overwhelmed and masked the acoustical background noise component of the input signal.

ii) On simplex paging channels, double tones preceded every short message.

iii) On half-duplex channels, (e.g., taxi dispatch service), two monologues are carried over a frequency channel. Moreover, many of their conversations were extremely short.

iv) On full-duplex telephone channels, repeater outputs carried both monologues (the mobile talker get a sidetone this way) as well as ringing tones. On the other hand, the repeater input could be heavily faded.

v) Impairments due to a local squelch circuit: its effect was minimized by injecting a local carrier but this made tuning unweldy and impaired reception (the incoming signal must be stronger than the local carrier which, in turn, must stay above the squelch circuit's threshold).

More realistic samples of mobile radio speech were obtained with the help of a voluntary amateur-radio operator whom we drove around the city, recording the output of his push-to-talk microphone. In the OFF position, the microphone's push-to-talk switch only suppressed the carrier but did not cut off the aduio input. The operator had an earphone for reception, so that the far-end talker's signal was not recorded. Of course, the operator's talk pattern would have been different if he had been given a full-duplex channel. Initially, we recorded a tape impaired by the AGC of a cheap cassette recorder. In a subsequent trial, we used a recorder with a manual level control. The only impairments in this final version were impulses caused by the push-to-talk button and jerky motions of the operator's hand.

4.4.4 Test Signals

We have a number of signal samples on cassette tapes. These samples are listed below and are numbered for later reference:

(1) The amateur-radio operator's monologue in a conversation that we referred to in Section 4.4.3. The operator was driven around the Ottawa city area in a sub-compact car with windows partially opened. The time was a weekday afternoon. A recorder with manual level control was wired to the microphone output.

(2) The same operator (a male) was driven around the City of Ottawa one weekday afternoon under rainy weather conditions in a compact car. This is the AGC-impaired recording we mentioned in the last section. The operator held two microphones, one connected to the transceiver and the other to the cheap cassette recorder.

(3) The recording from DOC's monitoring station---see last section.

(4) A tape of pure mobile noise recorded in a manual-shifted Datsun one weekday morning. The recorder sensitivity was adjusted to a very high level. This tape is truly noisy.

(5) A tape of good quality speech taped mainly from CKO radio's news broadcast. This is a recording of "continuous" speech.

In a sense, none of these samples are "perfect" in that they all contain some impairments that are probably good for stress tests but undesirable for algorithm development. For instance, the operator had a habit of clicking the push-to-talk button even if he was not intending to talk. This caused a lot of false alarms registered as very short spurts in the Activity Monitor's results (see Section 4.4.7). In another instance, the noise-tape 4 is sprinkled with a lot of impulses because the microphone we used is very sensitive to any motion, much more so than the radio operator's transceiver microphone. All in all, these impairments bias and sometimes dominate the results of the Activity Monitor.

In addition to the taped samples, we also exercized the detector with Gaussian noise from a noise generator. We compared the recorded mobile noise with the generator's white noise and concluded that white noise is not characteristic of a mobile acoustic environment, at least not for our samples. This comparison was performed by observing the short-time Fourier spectrum of the mobile noise samples on a fast all-digital spectrum analyzer. We found that, in general, the mobile noise spectrum tapers off for frequencies above 200 - 300 Hz at at least 10dB/decade. Of course, one cannot categorically argue for this difference because the mobile environment is so precarious that any noise is possible; that is why we did not discard white Gaussian noise in our tests.

One last note. It is possible in our test bed (Section 4.4.1) to mix in directly spoken speech. This offers more flexibility when investigating the detector's response to a particular utterance, or class of utterances. The talker can heighten or lengthen certain syllables to magnify subtle effects.

4.4.5 Performance Evaluation

The performance of the speech detector may be characterized in a subjective and/or objective manner. In order to compare the results, the test signals must also be characterized. Both kinds of characterization must be meaningful, readily accomplished, and reproducible.

Subjective speech quality measurement can be subdivided into 2 categories: analytical and utilitarian [23]. Analytical approaches aim at finding the psychological attributes of speech, either uttered singly or in a conversational context. Utilitarian tests are more common. They ignore the psycho-acoustical factors in speech production, perception, and interaction in conversations and instead dwell on the utility aspects of the test signal. They evaluate speech quality in many different criteria; those relevant to our case are: intelligibility, acceptability, annoyance, and transparancy. Recognition tests are usually conducted to measure intelligibility. The utterances could be nonsense artificial syllables called logatoms, isolated but unrelated words, and contextual speech such as reading from a book or newscast [23-25]. In acceptance tests, subjects assign the perceived material to a category of judgment under the condition that the subjects are aware of the material's intended usage. Annoyance may not be the exact opposite of acceptance when their underlying causes are taken into consideration. One aspect in a stimulus may give rise to the subject's acceptance while another aspect of it annoys the subject, not necessarily to the extent of rejection. Transparency is concerned with whether the subjects could perceive the presence or absence of effects produced by the system under test and whether communication is impaired in any way due to this awareness.

Speech detectors degrade speech in 3 ways:

(1) Clipping of speech. It was verified that clippings less than 12 msec. are not noticeable and those less than 40 msec. do not seriously affect intelligibility [20].

(2) False activation due to noise. Perception of its effect depends on the amount of energy let through by the detector. Our detector does not have a variable hangover, so false activation invariably let through at least 200 msec. of noise, which could be fairly objectionable. Some detectors use a hangover proportional to the time the signal stays above a certain threshold to ameliorate this effect [14].

(3) "Modulation" of the speech signal by the detector switch. This "modulation" is noticed as sudden silence when the voice-switch goes off. The effect is particularly annoying when the detector cuts off inter-syllabic within what appears to the listener as continuous utterances. Annoyance is proportional to the quieting effect

or in other words, to the level of the background noise. Perceptively, the "modulation" manifests itself as a sort of clipping even though speech may not have been clipped. This form of degradation can be lessened somewhat by adding a locally generated noise source at the receive end. The noise source can be running continuously or it can be switched in when no voice transmission is arriving.

It can easily be seen that all these degradations have their implications on the subjective qualities mentioned. We have elected to measure these qualities informally, without the assembly of subjects, a list of phonetically balanced phrases or logatoms, and all the other complications involved. The task of conducting these formal tests is left to the time when specific target application and users will be better identified. However, established procedures and test material will have to be modified so as to be more representative of the requirements on mobile radio as opposed to the general context of telephoning systems.

Objective evaluation in the context of speech detection centered around the measurement of signal levels and time structures, usually as correlates to each other. While they do not answer the paramount question of acceptability, objective measures are one step closer to reveal the internal workings of the speech detector. Attack time as a function of the starting syllable determines whether a buffer to delay the signal samples is necessary to avoid excessive clipping. The dynamic range of the detector must accommodate that designed into the audio system. The minimum signal-to-noise ratio (SNR) required to detect speech signals governs the sensitivity adjustment that precedes to the detector's A/D converter. In this way, the minimum required SNR consumes part of the full dynamic range offered by the A/D converter, as will be elaborated on later.

In addition to these single-value operational parameters, the ON/OFF statistics of the voice-switch collected from a large sample of monologues in conversations held in the (possibly simulated) target system provides information on potential data throughput rate. These statistics are the gap and spurt distributions and the activity factor, defined as the ratio of the voice-switch's total active time to the total. In Section 4.4.7, we will furnish samples of these statistics measured with our Activity Monitor.

4.4.6 Subjective Results

Overall performance of the speech detector has been judged to be acceptable. The few technical persons who had listened to the detector's output concurred on this verdict. Clippings do occur occasionally, but not to the degree of impairing intelligibility, and sometimes are noticeable only to alert and trained listeners. It is not always easy for the appraisers to distinguish clipping due to the inter-syllabic quieting effect from actual clipping, but the former has become more rare since the addition of the blanking clock mechanism. However, the mechanism's hangover extension effect becomes very noticeable when the background is very noisy; the voice-switch reacts more like a squelch circuit (settling time to cutoff time 1-2 sec.) instead of a dynamic speech detector. To operate satisfactorily, the minimum signal excursion must be at least +-50mV, or in terms of level, 13 mV. Taking the peak-to-rms ratio for speech as 18 dB, the excursion range translates into a minimum signal-to-quantization-noise ratio of about 12 dB. Another measurement showed that with respect to white Gaussian noise, the minimum SNNR (signal-plus-noise to noise ratio) is about 6 dB. This value is not constant over the detector's dynamic range; it drops off to about 4 dB at high noise level, possibly an indication of the diminished significance of quantization noise at this level. In the following discussion of results from subjective evaluation, we shall primarily address the intelligibility issue, in all cases assuming that all of the above level thresholds are satisfied. The subjects of acceptability and annoyance will be ignored, apart from stating the source of annoyance. Also, there is no need to say more regarding transparency than that any unimpaired ear can detect the switching action as long as the background sound is audible.

In developing the speech detector, there was a tendency to allow it to err more on false activities than missing speech spurts. As such, intelligibility impairment due to missing a significant portion of an utterance was rarely observed. Occasional clipping of unvoiced low-level syllables at the beginning or in the middle of utterances did not seem to be as significant as the quieting effect in the degradation of intelligibility. In the listening tests that led to this conclusion, the designer tried to concentrate on the presence and absence of speech in between the two ear pieces of the headphone (Fig. 4.14) and not so much on the already painfully familiar content. This is crucial because dullness can cause the listener to believe that he has heard the complete utterances. However, the same conclusion could be drawn even when the detector was fed live newscast.

There were a noticeable number of false alarms due to clicks when testing tape #1. Spurious sounds from the amateur radio operator like coughing or clearing of the throat could be detected as speech. Thus, the detector is very sensitive to impulse noise. It might be argued that this sensitivity can be reduced by increasing the counting threshold for the number of samples above the level threshold or by smoothing the input signal with some filter comparable to LPF. Either method would not be effective because impulses of mechanical origin have a long time span----at least a few milliseconds. Another repercussion of reduced sensitivity is that the attack time suffers. When the operator turned to talk to other passengers in the car, his speech was sometimes detected, but clipped due to diminished intensity.

The detector's reaction to breathing sounds was not apparent when it was tested with mobile speech. So we fed what was believed to be a heavy-breathing person's sizzling sound (as opposed to snoring) into the test bed's microphone. Under noiseless background, the voice-switched was turned on when air was blown directly into the microphone. Therefore, breathing sounds can be consequential if they are intense enough (6 dB above noise), or if air is blown directly into a microphone. However, we were convinced by the experiment that with a "normal" amount of background noise, the detector is immune to breathing noise.

4.4.7 Objective Results

While subjective tests reveal the satisfaction of the voice users, some statistical measurement is necessary to determine the potential data transmission capacity in terms of percentage of total time being pauses (called gap activity factor) and the distribution for the durations of these pauses. The distribution of talkspurts or gaps is just the probability that a given talkspurt or gap, respectively, will be less than a certain time interval. These distributions enable the modelling of interleaved data and voice transmission and therefrom data throughput rate can be computed [3]; such throughput rate will be more realistic than the gap activity factor. The gap distribution alone indicates how much data at a given rate can be transmitted without being garbled by the next talkspurt at a given level of probability. The talkspurt distribution determines the average waiting time and hence the data buffer size. However, these distributions and activity factors are very application-dependent; distributions shown here are intended to illustrate the general trend and should not be taken as anything representative of any mobile application because of the small sample size used. The distributions presented here can be compared with Brady's work on telephony speech [26-27].

A piece of software, called the "Activity Monitor" (Fig. 4.14), that runs on a SBC8020 microcomputer board, samples the voice-switch control signal every 4 msec. over a given period of time to produce "histograms" for the talk-spurts and pauses. The Monitor keeps two tables of "bins" (counters), one for the gaps and one for the pauses. A bin is bumped up by one if the length of a pause or talkspurt is less than or equal to the upper limit represented by that bin but greater than the limit of the previous bin. The scale on which these bins are arranged is not linear: the distance between consecutive bin limits become progressively higher. Markers for the bins are stored in a table so that the time axis of the histogram is easily modified. The Monitor also keeps counters to tally spurt and gap activities.

Fig. 4.15 and 4.16 depicts the gap and spurt distributions for the three combinations of tape 5 (Radio Newscast) and white Gaussian noise: noise only; speech only; and speech plus noise. The activity factors are also shown. False-alarm activity without speech is only 3.2%, a "reasonable" figure. One expects the speech activity factor to increase with noise level. However, this isn't the case. The factor drops from 93.5% for "noiseless" speech to 86.1% for noisy speech. Thus the only effect due to the added noise is to raise the threshold. Obviously there was so much

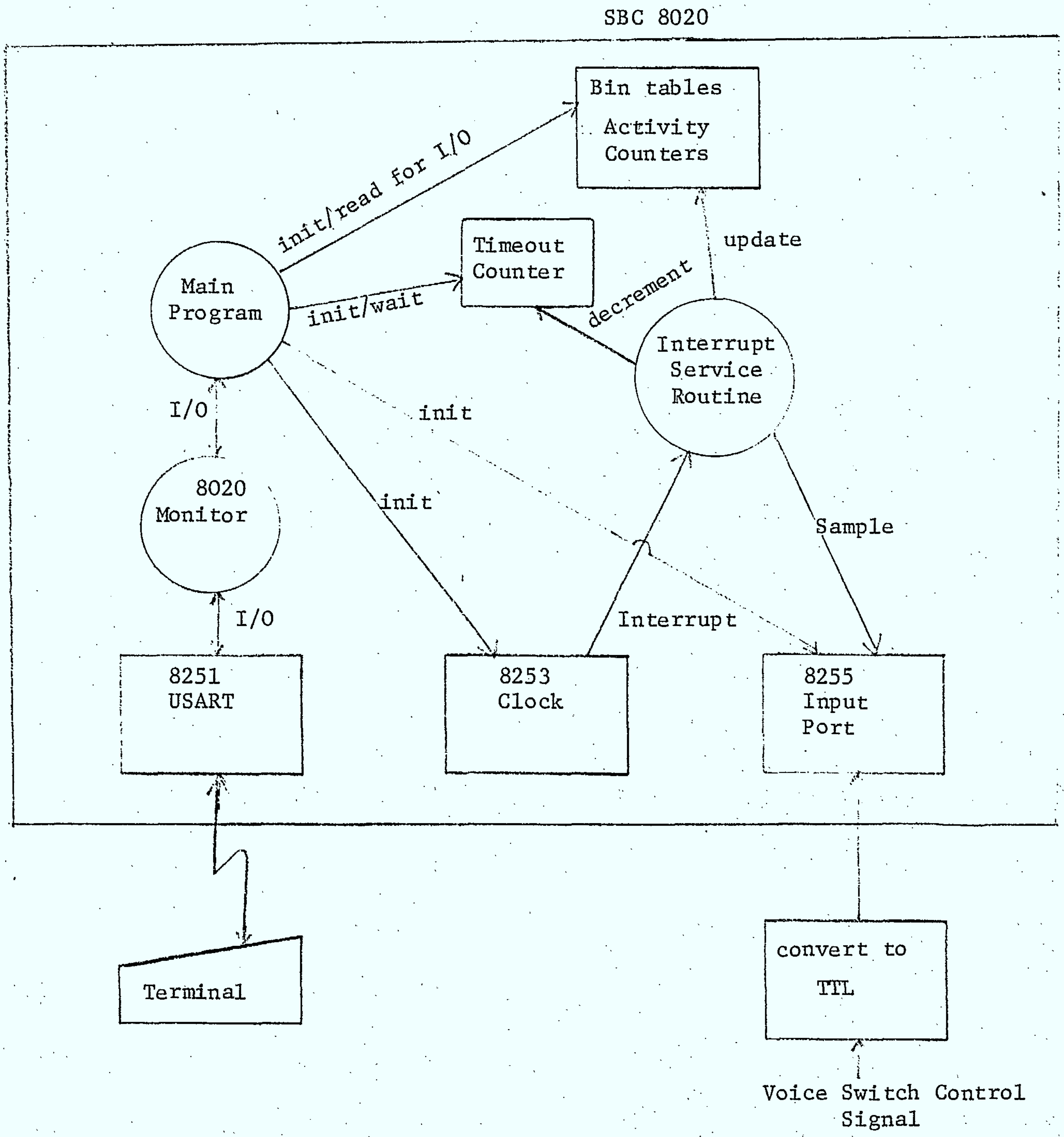


Fig. 4.14 Activity Monitor

Fig. 4.15

Gap Distributions for the 3 combinations of newscast (tape #5) and White Gaussian Noise. Noise Power = 36mV. SNR is just above audible clipping theshold for the signal + noise combination. Levels are unchanged over all three combinations. Total test time = 1700 sec. Tape counter: 0 - 540.

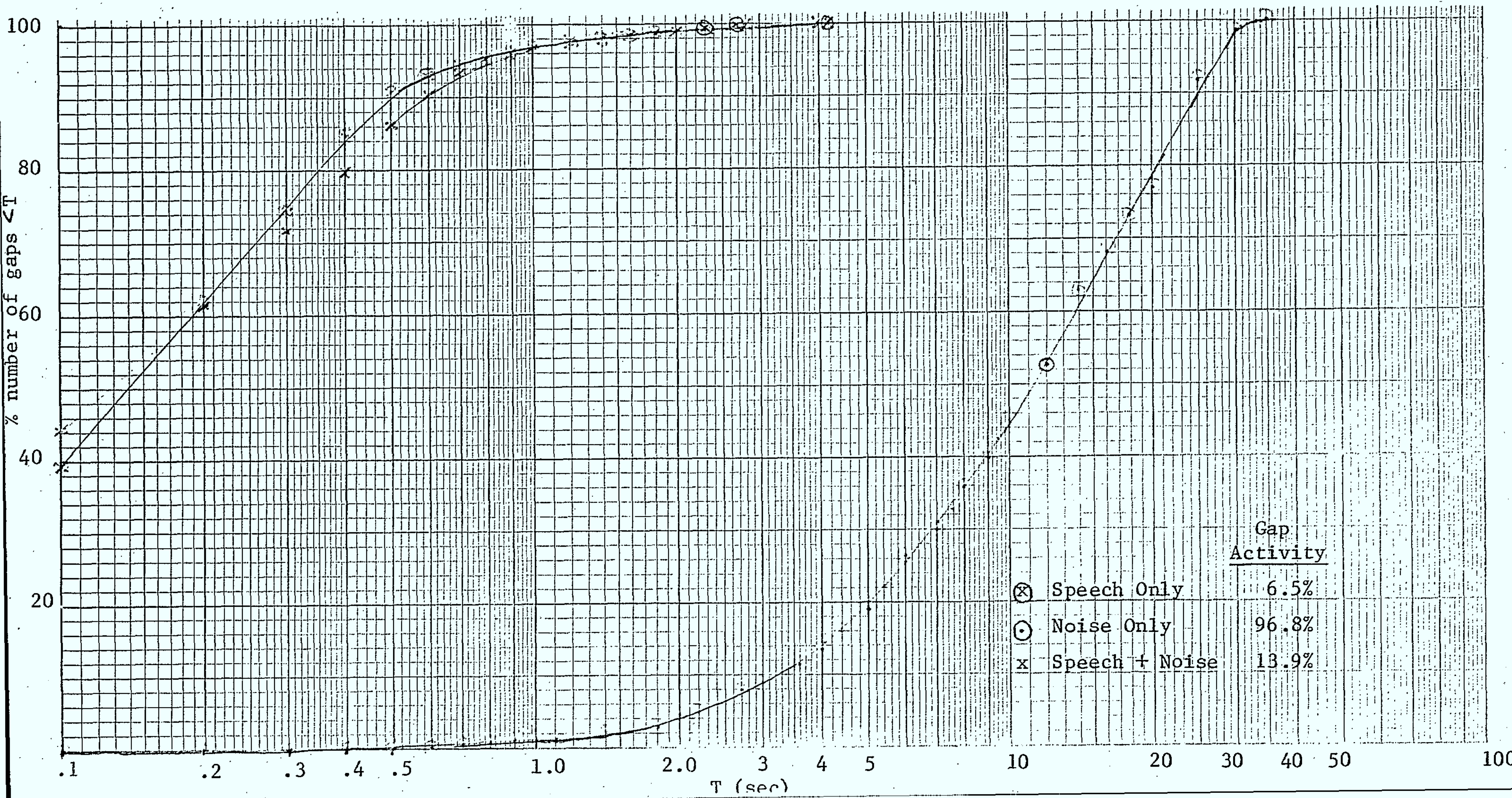


Fig. 4.16

Talkspurt distributions (see Fig. 4.15)

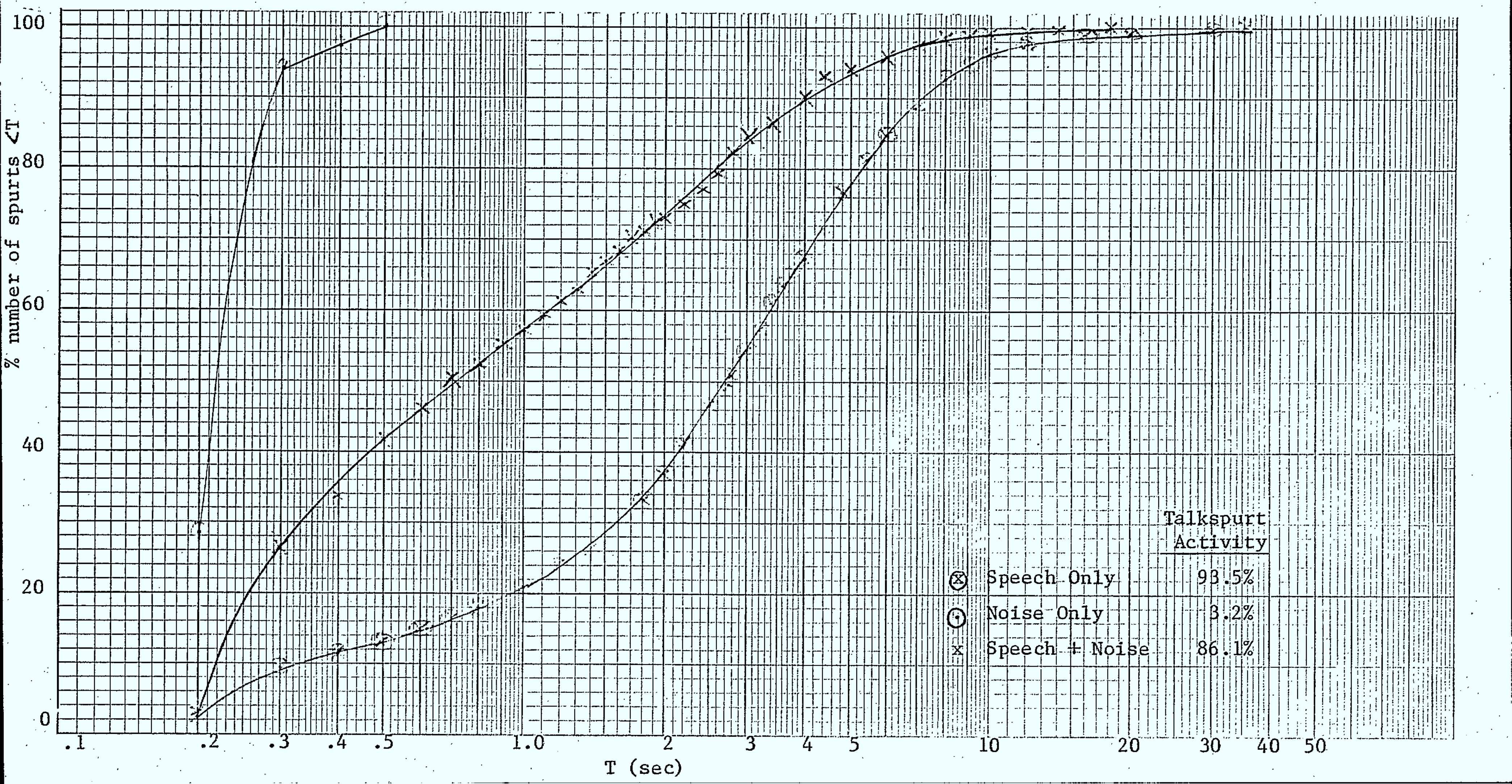


Fig. 4.17

Time Weighted Gap Distributions. Speech: Newscast (tape #5)
Noise : White Gaussian (Power: 36mV)

100% is total time of all gaps.
SNR is just above audible - clipping theshold under noisy condition.
Levels are unchanged over all three combinations.
Total test time: 1700 sec. Tape counter: 0-540

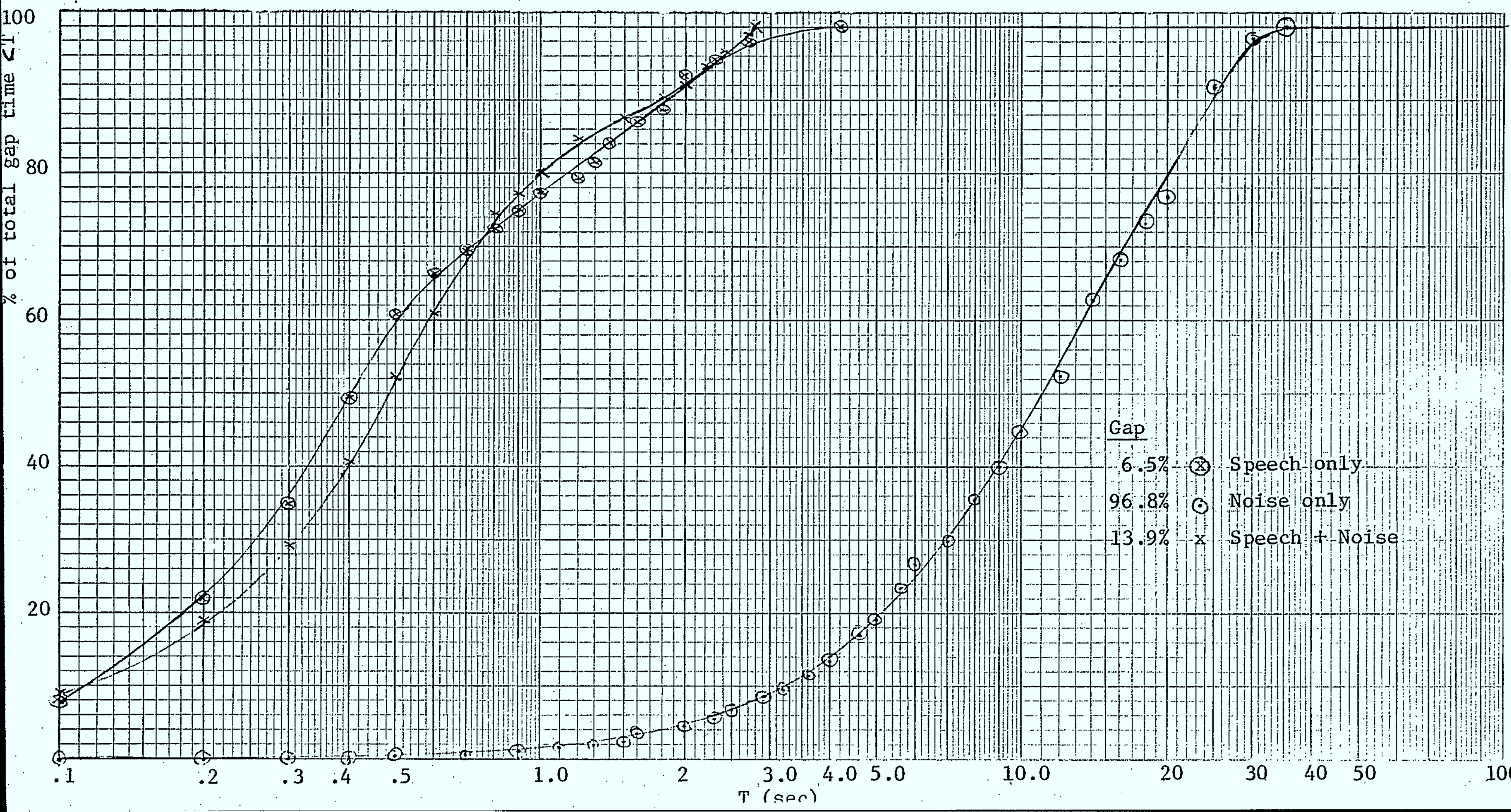


Fig. 4.18

Gap Distributions for selected active segments of live mobile radio recordings. Signal swing ± .5 volt.

	Tape	Counter	Quietness	Gap Activity	Total Test Time
⊗	1	120-196	Quiet-Moderately Noisy	57.1%	240 sec
x	2	15-125	Noisy	54.0%	360 sec

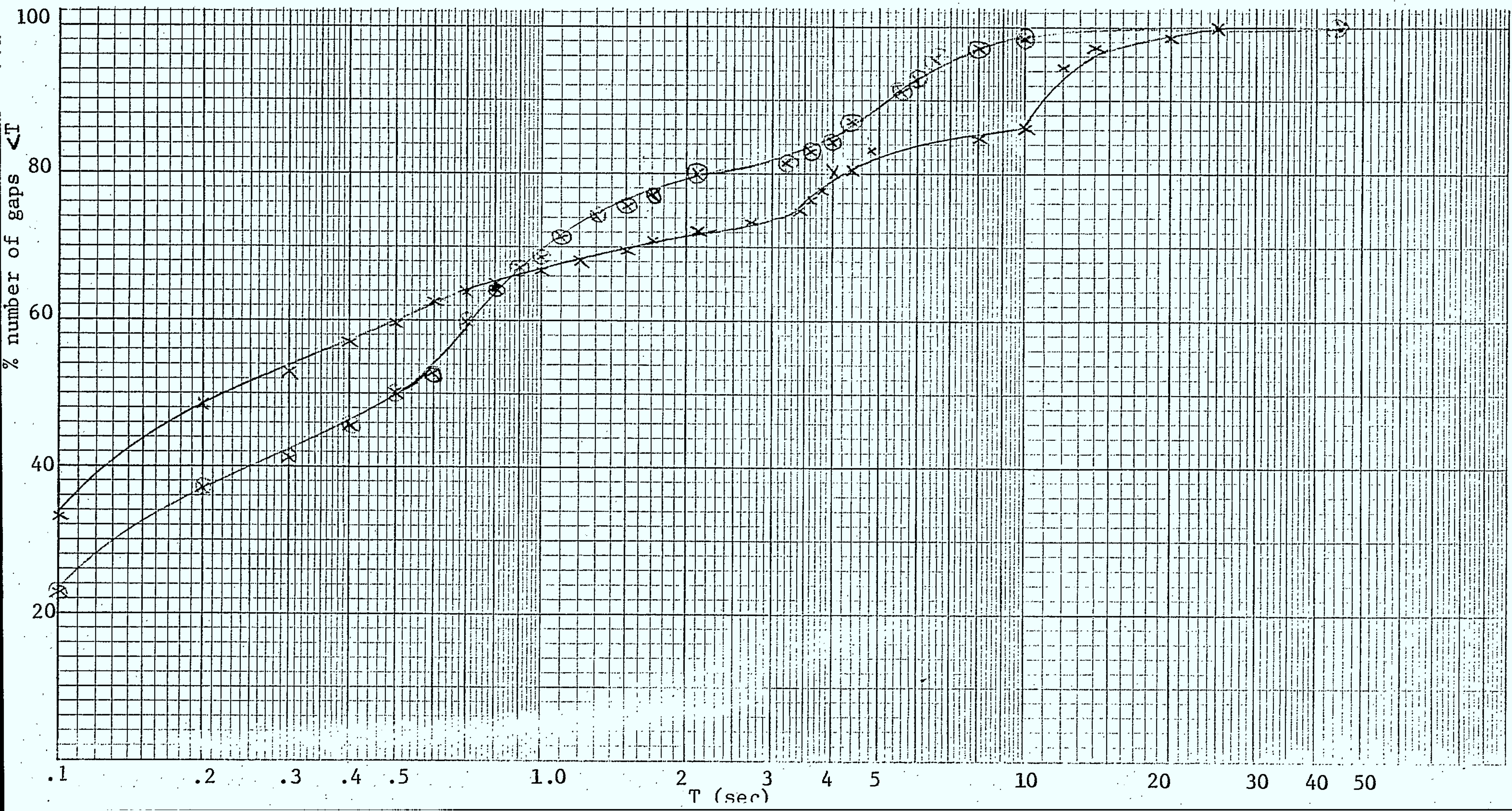
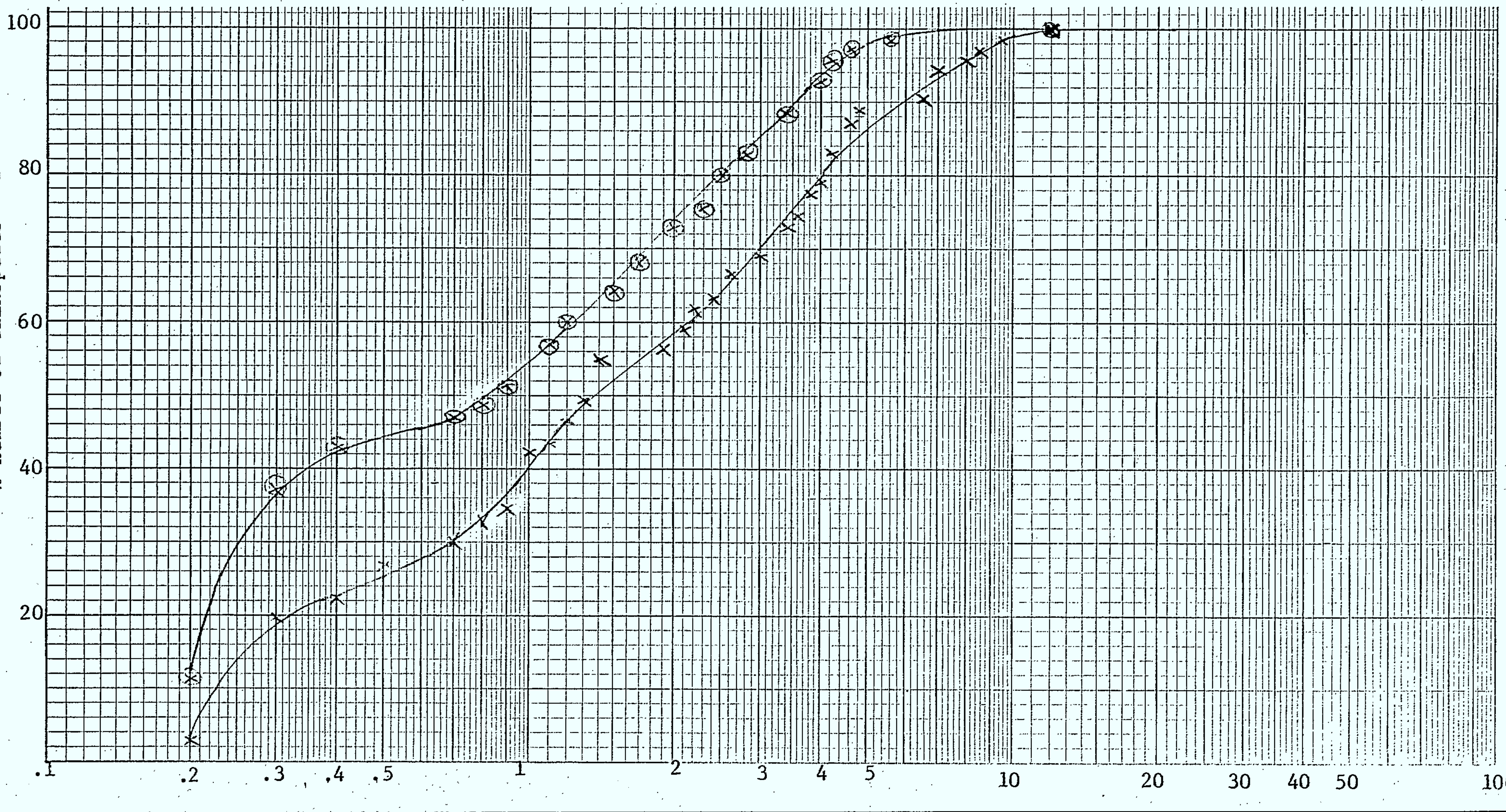


Fig. 4.19

Talkspurt Distribution (see Fig. 4.18)

Talkspurt Activity

⊗ 42.9%
x 46%



speech that false alarms were all "buried". It is interesting to note that the gap distribution for pure speech is not significantly altered by the addition of noise, but the spurt distribution completely shifts to the left, indicating that spurts are shortened. The shapes of the gap distributions for pure and noisy speech and of the talkspurt distribution for pure speech resemble Brady's distributions [27].

One must bear in mind when interpeting the distributions that the unit of the vertical axis is the fraction of the total number of gaps or spurts, not the fraction of the total activity time. For instance, even though 40% of all gaps in the noiseless newscast are less than .1 sec. long, the fraction of gaps with that length as a percentage of total gaptime hardly exceeds 1%. This is depicted in Figure 17, which shows the time-weighted "distribution" of Figure 4.15.

Fig. 4.18 and 4.19 show the distributions for two segments of tape 1 and 2. These distributions are rather "kinky" because mobile radio talk patterns are not as regular as that from newcast and also because the test interval is smaller than the last case. The segments tested were chosen as they were relatively active. The activity factors recorded and shown are typical of those quoted for telephone circuits (40% active) though our figures are a bit on the high end. Not many general and concrete conclusions can be drawn from these graphs because of the short test time. However, it can be seen that there is a greater dispersion for our distributions than that presented by Brady. This need not always be the case. For example, messages exchanged in taxi-dispatch systems are generally very short. It can also be seen that almost half of the gaps are less than .2 sec. This will definitely destroy many data packets in the process of transmission. At the time of writing, effort is under way to reduce the fraction of such short gaps either by cutting short the hangover or eliminating the hangover extension effect.

4.5 Summary

If TASI and DSI work well for telephony to economize on transmission bandwidth, there is no reason why similar benefits cannot be introduced into mobile radio. The speech detector presented in this chapter is a step forward in the achievement of that goal in that it enables a non-active voice channel to be used by others in a pool of data and/or voice users. Unavoidably, some details have been omitted. For a more in-depth study of speech detection, details on the "final version" of the speech detector, more objective test results, and the like, the readers can refer to an upcoming report in the form of a thesis for the Department of Systems Engineering at Carleton University.

The present detection algorithm performs satisfactorily with "normal" conversational speech. It is also simple enough to be realized either on a programmable signal processing chip or in a custom LSI chip. However, more work has to be done before the detector can be installed into the harsh environment of a mobile. This is particularly so with respect to the detection of non-conversational speech sounds and for the convenience of impaired talkers. Regarding the former, the speech detector has been designed to regard anything less variable than "normal" speech as noise. Whether a sound is monotonic or chromatic, as long as it is held steady for a long enough time span, it can potentially be clipped. Candidates from such a category include whistling, singing, music or simply tones from such devices as alarms. This problem can be solved by incorporating into the speech detector a manual or automatic override feature or by augmenting the detector with extra intelligence to detect unusual sounds. By manual override, we mean that a talker instructs the transceiver to bypass the detector after receiving some cue (e.g., a detector-state LED or feedback from the far-end talker) that his/her speech is being clipped. An automotive override can be implemented as a fixed ceiling (preset at manufacturing or installation time) imposed on the variable theshold so that any signal above the ceiling is passed through regardless of the signal's origin. In our context, the 2920 processor can be programmed to read the voltage of a potentiometer through one of its input channels and set a ceiling accordingly.

Detection of special sounds using analytically derived computational algorithm is a feasible approach only if implementation cost is insensitive to complexity. This is not untenable considering that a mass produced VLSI chip is not much more costly than a mass produced LSI chip. The said complexity can be illustrated by noting that detection of tones of unspecified frequency always entails some form of frequency scanning (and in our case in "real-time").

The all pervasive impairment---impulse noise---finds its way into speech detection as well. The present detector is sensitive to impulse noise. The problem can possibly be solved by median filtering [8]. However, this requires not only more computational power, but also more memories to accommodate frame-by-frame processing.

Any future work would definitely be benefited by live recordings on a larger scale. The live recordings used in our tests cannot be said to be universally representative because the amateur radio operator was conversing leisurely with friends or other operators on half-duplex channels. A large sample of recordings from a number of mobile systems in operation would enable the differentiation of talk patterns among different classes of users based on measured distributions. Potential data throughput rate could thereby be computed. More recordings would also be helpful for the characterization of the noise background [28].

5.0 Summary and Further Research

5.1 Summary

Phase A of this research*, completed in March 1981, examined the basic structure of a two hop mobile UHF transmission system in which units communicate with each other through a fixed base station. Analysis of this initial phase concentrated on the uplink model. A second part of phase A was an examination of the relationship between packet loss and speech intelligibility. Experimental transmission at 16 kbs was used to conduct subjective tests to examine this relationship.

The study reported here contains the results of phase B and represents a continuation of phase A. Analysis was concentrated on the downlink loop, and two feasible methods of shared channel access were identified. Simulation was used to verify these results. However, the main emphasis in phase B was on prototype design and development, where two major problems were addressed. The first of these is the design of a novel modulation system. Continuous phase modulation (CPM), which uses partial response signalling, has

---

* J.S. Riordan, S.A. Mahmoud, S.E. Aidarous, "Design Considerations in Packet Mobile Radio Data Networks", Report No. SC81-1, Dept. of Systems and Computer Engineering, Carleton University, Ottawa, March 1981.

the advantage of excellent spectral properties and good error performance. In an effort to achieve these properties, research in phase B focussed on the optimization of a quadrature coherent mean square error detector. The application of this detector to tamed frequency modulation was investigated in detail. The second problem investigated in phase B was the design and implementation of a speech detector capable of detecting clauses in conversation on mobile radio channels. A speech detection algorithm has been proposed and implemented in real time on a digital signal processor.

5.2 Further Research

As extensions of the work being carried out at the present time, we propose the following research items:

(i) The development of an overall model for the transmission efficiency of packetized data taken into account the synchronization time and packet retransmission probability.

(ii) Research will be taken up to evaluate a noncoherent modulation scheme for TFM modulated signals with the aim of developing efficient modems that meet the stringent requirements dictated by short data packet transmission over variable channel conditions. The advantage of noncoherent scheme is the near-instantaneous demodulation action which avoids large synchronization delays.

(iii) The bit error rate performance of the coherent and noncoherent TFM modems will be studied experimentally.

(iv) A protocol will be developed in greater detail for handling packets in the base station buffer. An accompanying analysis will be carried out to analyse resultant delays.

(v) The analysis in item (iv) will be verified through simulation, so that an overall picture of the two hop delay can be developed.

REFERENCES

[1] K. Bullington, "Engineering Aspects of TASI", B.S.T.J., March 1959, pg. 353.

[2] J. Fraser, "Overall Characteristics of a TASI System", B.S.T.J., July 1962, pg. 1439.

[3] J. daSilva, "Performance Analysis of Land Mobile Packet Radio Systems", Rep. No. SC80-3, Dept. of Systems Engineering, Carleton University.

[4] "Analysis and Design of Mobile Terminal Transmitters for Mixed Voice and Data Packets in the 800 MHz Frequency Band", March 1981, Dept. of Systems Engineering, Carleton University.

[5] R.E. Fisher, "A Subscriber Set for the Equipment Test", B.S.T.J., Jan. 1979, pg. 123.

[6] M.L. Gayford, "Electroacoustics", 1970, Butterworth & Co. Ltd., London.

[7] D.L. Richards, "Telecommunication by Speech", 1973, Butterworth & Co. Ltd., London.

[8] L.R. Rabiner, et al., "Digital Processing of Speech Signals", 1978, Prentice-Hall, Inc.

[9] L.W. Kinsler, "Fundamentals of Acoustics", 2nd ed., 1962, John Wiley and Sons, Inc.

[10] R.W. Schafer, et al., "Detecting the Presence of Speech Using ADCPCM Coding", IEEE Trans. on Comm., May 1976, pg. 563.

[11] C.K. Un, et al., "Voiced/Unvoiced/Silence Discrimination of Speech by Delta Modulation", IEEE Trans. on A.S.S.P., Aug. 1980, pg. 398.

[12] E. Fariello, "A Norel Digital Speech Detector for Improving Effective Satellite Capacity", IEEE Trans. on Comm., Feb. 1972, pg. 55.

[13] J.A. Jankowski, "A new digital voice-activated switch", Comsat Tech. Rev., Spring 1976, pg. 159.

[14] R.E. LaMarche et al., "Digital Speech Detector", U.S. Patent 4,028,496, June 7, 1977.

[15] P.G. Drago et al., "Digital Dynamic Speech Detectors", IEEE Trans. on Comm., Jan. 1978, pg. 140.

[16] L.R. Rabiner et al., "An Algorithm for Determining the Endpoints of Isolated Utterances", B.S.T.J., Feb. 1975, pg. 297.

[17] B.S. Atal et al., "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition", IEEE Trans. on A.S.S.P., June 1976, pg. 201.

[18] R.J. McAulay, "Optimum Classification of Voiced Speech, Unvoiced Speech and Silence in the Presence of Noise and Interference", Lincoln Lab. Tech. Note 1976-7, M.I.T.

[19] H. Miedema et al., "TASI Quality-Effect of Speech Detectors and Interpolation", B.S.T.J., July 1962, pg. 1455.

[20] G.G. Szarvas, "Voice-activated-switch performance criteria", Comsat Tech. Rev., Spring 1980, pg. 151.

[21] J. Jaffe et al., "Markovian Model of Time Patterns of Speech", Science, May 1964, pg. 884.

[22] P.T. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation", B.S.T.J., Sept. 1969, pg. 2445.

[23] "IEEE Recommended Practice for Speech Quality Measurements", Subcommittee on Subjective Measurements, IEEE Audio and Acoustics Group, IEEE, 1969.

[24] D.J. Goodman et al., "Intelligibility and Ratings of Digitally Coded Speech", IEEE Trans. on A.S.S.P., Oct. 1978, pg. 403.

[25] L. Pols, "Intelligibility of Intervocalic Consonants in Noise", Proc. of the Int. Conf. on A.S.S.P., 1979, pg. 460.

[26] P.T. Brady, "A Technique for Investigating On-Off Patterns of Speech", B.S.T.J., Jan. 1965, pg. 1.

[27] P.T. Brady, "A Statistical Analysis of On-Off Patterns in 16 Conversations", B.S.T.J., Jan. 1968, pg. 73.

[28] J.T. Walker, "Mobile Telephone Control Unit Design Guidelines for the Advanced Mobile Phone Service", IEEE Trans. on Veh. Tech., Nov. 1978, pg. 276.

[29] D.L. Richards, "Statistical Properties of Speech Signals", Proc. IEE, May 1964, pg. 941.