TRANSMISSION OF SPEECH IN PACKET-SWITCHED NETWORKS

BY

ROGER CECIL FERRY TUCKER

A thesis submitted for the degree of

Doctor of Philosophy

of the

University of Aston in Birmingham

Department of Electrical and Electronic Engineering

August 1984

TRANSMISSION OF SPEECH IN PACKET-SWITCHED NETWORKS

ROGER CECIL FERRY TUCKER

A thesis submitted to
The University of Aston in Birmingham
for the degree of
Doctor of Philosophy 1984

Summary

Packet switching speech not only allows straightforward multiplexing of speech and data, but also enables efficient use to be made of transmission capacity. This thesis shows how the transmission efficiency of a packet-switched speech network can be optimized. To do this, an analytical model of the packet queueing process is developed, and subjective tests on the effects of packet losses are described. Together these provide a powerful tool for the performance assessment of a packet-switched speech link.

The analytical model is based on a two-dimensional Markov chain, and models the queue as a continuous variable. This enables the delay distribution and amount of packet loss to be found by solving a first-order matrix differential equation. The accuracy of the analysis is investigated through the use of simulations, and is found to be good for the range of parameters of interest.

The subjective tests clarify inconsistencies in the literature, and give reliable indications of the effects of packet losses. A preliminary test shows that 8ms packets with gaps filled by repeating the previous packet give the best speech quality. A listening test shows that speech with packet losses of up to 20% can be understood with only moderate effort. Finally, a conversation test shows that 2.5% loss is acceptable to almost everyone, 10% loss is acceptable to 70% of people and 25% loss is acceptable to 33% of people.

Some factors that effect the performance of a packet-switched speech link are investigated, namely the way packet loss and delay are used, packet length and receiver buffering. Ways to ensure adequate data performance, and the trade-off between TASI advantage, delay and packet loss are also investigated. From these investigations, principles for optimizing the transmission efficiency of a packet-switched speech network are established.

Key words

PACKET SWITCHING, SPEECH, QUEUEING ANALYSIS

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor Professor J. E. Flood both for guiding the direction of this research, and for the criticism and encouragement he has given. I would also like to thank Dr. M. H. Ackroyd for his supervision of the early stages of this work, and Mr. R. W. Whorwood and Professor D. L. Richards for many helpful discussions on subjective testing.

I am very grateful to GTE Laboratories Incorporated for their financial support, and for allowing such freedom in choosing the direction of the research. I am also grateful to the Department of Electrical and Electronic Engineering of the University of Aston for the use of its facilities, and to Mrs. Helen Turner and all the other secretaries who have helped type this thesis.

Thanks are also due to Grace Towers and to my parents for the support and encouragement they have given me, and to the many other friends who have given freely of their time to be subjects in the tests. Most of all, I am thankful to God for his help and guidance, particularly through the critical stages of this research.

CONTENTS

APPENDICES

LIST OF SYMBOLS

a_K	scaling constant for Kth eigenvector
A	average traffic in Erlangs
b_j	probability of state j
$b(j,v,\alpha)$	binomial probability given by $\binom{v}{j} \alpha^j (1-\alpha)^{v-j}$
$\underline{c}_K$	Kth eigenvector
c_{jK}	jth element of Kth eigenvector
$d(t)$	overall p.d.f. of delay
d_π	overall probability of no delay
d_Θ	overall probability of maximum delay
D	one-way delay
h	capacity of link
j	number of active talkers
j_{ave}	average number of active talkers
j_ℓ	lowest number of talkers with probability $\geq 10^{-5}$ of occurrence
j_u	greatest number of talkers with probability $\geq 10^{-5}$ of occurrence
$\ell(t)$	overall p.d.f. of loss with a delay of t
L	fractional packet loss
m	maximum queue length allowed
N	number of calls at which blocking occurs
$o(\Delta t)$	any function for which $\lim\limits_{\Delta t \to o} \dfrac{o(\Delta t)}{\Delta t} = 0$
$p(i,j)$	transition rate from state i to state j
$p^*(i)$	total flow-rate out of state i
$p_c(v)$	probability of v calls

p_D	probability of difficulty with a roundtrip delay of 2D
p_L	probability of difficulty with a fractional packet loss of L
p_{LD}	probability of difficulty with a roundtrip delay of 2D and a packet loss of L
$p_u(v)$	probability that a caller is one of at least v callers
$q(x)$	overall equilibrium p.d.f. of queue length x
$q_j(x)$	equilibrium p.d.f. of queue length x with j speakers
$\underline{q}(x)$	vector of $q_j(x)$
$q(t,x,j)$	p.d.f. of queue length x and j speakers at time t
q_π	overall probability of no queue
q_Θ	overall probability of maximum queue
$Q_j(n)$	probability of a queue of n packets at the end of a frame
R	state matrix
s	service time of one packet
t	time or queueing delay
v	number of off-hook callers
v_L	smallest number of talkers that gives a packet loss greater than L
x	queue length
X_i	values of packet loss obtained by simulation
$\bar{X}$	mean value of X_i
α	average fraction of people talking
β	constant
λ_K	Kth eigenvalue

λ	reciprocal of mean silence length
μ	reciprocal of mean talkspurt length
π_j	probability of no queue with j speakers
Θ_j	probability that queue is on its limit with j speakers
ρ	efficiency, defined by $(v\alpha/h\ )$

LIST OF FIGURES

Page

LIST OF TABLES Page

CHAPTER ONE

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The current interest in packet switching for data has grown from the success of ARPANET and other data networks. With packet switching, any information to be transmitted is formed into a packet of length up to 1000 bits and a header is attached. The header contains the packet's destination and source (the latter enables retransmission of corrupted packets), and various other information, e.g. control bits and a sequence number. In addition to the header, framing bits and error check bits are also attached to the packet. The packet is then transmitted through the network with each node reading the header and forwarding the packet to another node. Packets may have to wait for transmission and so a very significant characteristic of packet switching is a variable network delay.

Packet switching has proved to be a versatile switching system, allowing easy maintenance, modification and extension. It allows a wide range of data rates to be used on the same network. It is reliable, since a connection can be guaranteed at all times, even if the network is heavily loaded or if several links are out of service. It guarantees error-free transmission, since error correction can be carried out on a link-by-link basis using retransmission requests. Finally, it makes very efficient use of transmission capacity. This is particularly useful when the transmitted data arrive in short, well-spaced bursts, as with interactive data.

The growth of data communications has led to an increasing interest in the integration of data and digital speech into the same network. Due to its many advantages, packet switching is the obvious switching method to use for data. The choice for speech is far less obvious. Where rented lines are used, transmission costs are a substantial part of the cost of operating a network covering distances of hundreds of miles (i.e. a wide-area network). If current trends continue (one paper reports that in nine years the cost of computer memory and switching has fallen 30 times faster than transmission costs [1]), the cost of communication networks will soon be dominated by transmission costs. It is therefore becoming increasingly important that switching methods should make efficient use of transmission capacity [2, 3].

Consequently, it is questionable if circuit switching, which has been used almost universally for speech transmission since telephony began, is the most suitable switching method for an integrated digital network. For an average conversation, each speaker is silent 60% of the time, and so with circuit switching 60% of the transmission capacity is used to transmit silence. Because of this, a number of alternative switching methods for digital speech in an integrated network have been proposed that form a continuous spectrum from circuit switching to packet switching. A good review of these is contained in reference (2).

The general aim of these switching systems is to remove the silences from each speech channel, and multiplex either other speech channels or data packets into the gaps. (For this reason these are referred to as speech interpolation (SI) systems.) This enables more speech calls to be carried than a link would have capacity for if circuit switching were used. The ratio of number of calls to circuit-switched capacity is usually referred to as TASI advantage, since TASI (time assignment speech interpolation) was the first system that used these techniques (4). Two common characteristics of all SI systems are an overhead for signalling, and the loss of some speech when the number of active channels exceeds the channel capacity. In some SI systems, the effect of the speech loss is reduced by losing bits from every speech sample, instead of losing short sections of speech completely.

In this thesis, the use of packet switching for speech is considered. Packet switching has a number of advantages over other switching methods. As packet switching is likely to be used for data, using it for speech unifies the switching. (This is not to say that data packets and speech packets should be treated in the same way, as their requirements are very different.) Also, packet switching has the potential to make better use of transmission capacity than other switching methods, for three reasons.

Firstly, buffering is possible with packet switching. This gives an alternative to speech loss when the demand for transmission momentarily exceeds the capacity of a link. Consequently, a link can carry more traffic with packet switching than any other SI system. However, the buffering causes a delay to be inserted in the conversation, and this is in itself a degradation that must be taken into account.

Secondly, packet switching allows a simple method of speech loss that has a smaller effect than that associated with the simple methods of other SI systems. In particular, it has been shown that packet loss has much less effect than the "front-end clipping" that occurs in TASI systems [5], since the losses can be well distributed through the talkspurts. Since higher levels of speech loss can be allowed, more traffic can be carried.

Packet switching also allows more sophisticated forms of speech loss, since a packet can be arranged such that when the tail end is discarded, the quality of the speech is reduced but no speech is lost (this arrangement is called "embedded encoding" [5-7]). In principle, the quality of the very best DSI (digital speech interpolation) systems [8] can be matched in this way. In addition, whereas it would be quite difficult to build a network of DSI systems, packet switching is particularly suited to network operation [9].

Thirdly, packet switching allows speech coders with different output bit-rates to use the same network. More significantly, it is unnecessary for any speech coder to have a fixed bit-rate output, since packets can be of varying lengths. This enables speech coders to be used that vary the output bit rate to match the requirements of the speech signal, which results in a lower average bit rate with no drop in quality [10-14]. The potential of packet switching to exploit variable-rate coders has been seen only by a small number of researchers [1, 15-20].

This thesis describes research into how packet switching can be used for speech to make the best use of transmission capacity. Consequently, the details of how data and speech may best be integrated will not be directly considered. However, when considering packet switching as a technique for integrating speech and data, it is generally thought that speech packets should have priority over data packets in transmission queues [15, 17, 21, 22]. This is simply because the delay requirements of speech are more severe than those of data. Consequently, the speech performance of a network will be unaffected by the presence of data packets.

One implication of this is that if data traffic is dominant (in terms of overall bits/sec transmitted), the speech will effectively operate at a TASI advantage less than 1. In this case there will be no speech loss and minimal delay, and the speech performance will be almost as good as circuit-switched speech. Thus, packetising the speech creates no difficulty. However, it will be assumed here that either speech is the dominant traffic type or that, for some other reason, a high TASI advantage is required.

The next two sections outline the present knowledge of issues concerning delay in a packet-switched speech network (section 1.2), and the factors that effect speech quality (section 1.3). It should be noted that the concern here is with a network designed to carry speech. The issues concerning transmission of speech packets over a network designed for data, such as ARPANET, are somewhat different. As this is a very specialised application of packet-switched speech, it is not considered at all in this work.

1.2 DELAY IN A PACKET-SWITCHED SPEECH NETWORK

1.2.1. Effect of Delay on Speech Conversations

Before discussing the causes and the control of delay in a packet-switched speech network, it is important to clarify the effect that delay has on a conversation. Various figures from 200ms to 500 ms are quoted as the maximum delay that can be allowed in a conversation without adverse effects [(23-28)]. Many of these estimations are not based on experimental evidence, but general impressions. However, there is no shortage of experimental work on the effect of delay on a conversation. Long propogation times in speech conversations were investigated in the 1920-30 era when loaded cables were used, and again in the 1960s when communication satellites were first used. Some of the studies considered delay with echo-suppression; others considered echo-free delay. Since echo cancellors [(29,30)] have now been developed that produce very much better quality speech than echo suppressors [(30-32)], the results for echo-free delay have more significance for future systems. Unfortunately the results of these studies into echo-free delay do not altogether present a clear picture of the effects of delay.

D L Richards [(33, 34)] and Brady [(35)] both classified delays by the number of confused situations occuring in the conversations. D L Richards found that the number of confusions was roughly proportional to delay, whereas Brady found that 600 ms roundtrip delay produced a similar number of confusions to 1200 ms roundtrip delay. Krauss and Bricker [36] studied the number of words and length of utterances used to complete a descriptive task, and found that 600 ms roundtrip delay had no effect but 1800 ms roundtrip delay had a marked effect.

Klemmer [37] gave people the opportunity to reject calls they had difficulty with, and found that very few (<4%) calls with 600 ms and 1200 ms roundtrip delay were rejected in the first weeks of his test. However, when a delay of 2400 ms was used for a week, 27% of calls were rejected. Furthermore, for the remaining weeks of the test 13% of 1200 ms delay calls were rejected, but still no 600 ms delay calls were rejected. Klemmer called this effect "sensitization to the delay"; it can be explained by people becoming aware that conversational difficulties were caused by the connection. Without this awareness, people tend to think that the other person is simply being uncooperative and impolite [35].

A reasonable conclusion to reach from these results is that conversations are adversely affected by roundtrip delays of 600 ms, though users unaware of delay and its effects will not notice the delay until it is 1200 ms or more. There is no doubt that 2400 ms roundtrip delay is a serious degradation.

Using these tests as their basis, the CCITT has recommended that roundtrip delay should not exceed 800 ms [38]. However, this is a slightly arbitrary threshold. It should also be noted that all these results apply only to constant delay. No results are available on the conversational effects of a varying delay.

1.2.2. Sources of Delay

Whereas some delays in a packet-switched speech network are fixed, others are variable. The largest fixed delay would probably be the delay incurred whilst waiting for a whole packet to be formed at the transmitting end and played out at the receiving end. This is known as the packetisation delay, and is equal to the period of speech contained in a packet. Transmission delay, the length of time taken to transmit a packet once it has been formed, is insignificant compared to packetisation delay unless the link capacity is very low. Propagation delay, which is most significant over satellite links, is common to all switching techniques and need not be considered here.

Variable delays occur at the switching nodes. Processing delay, the time taken to read the packet header and place the packet on the appropriate output queue, is dependent on the speed and complexity of the packet switches. As faster and more complex microprocessors are produced, processing delay will become of little significance. It will be assumed here that it can be reduced to a few milliseconds at no great cost and is therefore small compared to other delays [19].

Queueing delay is the delay incurred whilst a packet waits in the output queue of a node (i.e. whilst waiting for transmission). This delay is very significant, not only because it is potentially the source of largest delay, but also because the ability to queue for transmission gives packet switching a major advantage over other

SI systems.

Finally, the speech-packet receiver attempts to remove the variation in delay incurred at the switching nodes. To do this it must further delay some packets and perhaps discard others which have been delayed too much. Thus although the average delay is increased by the receiver, the peak delay will, if anything, be decreased.

1.2.3 Packetisation Delay and the Problem of the Header Overhead

The constraining factor on packet length, and therefore packetisation delay, is the packet header. If a large TASI advantage is to be achieved, a large header overhead is not acceptable.

Strictly speaking, the packet header does not contain framing bits and error check bits as these belong to a different level of the ISO OSI (international standards organisation open systems interconnection) reference model (39). However, as the issue here, and throughout the rest of the thesis, is the header overhead rather than protocols, the term "header" will be used to describe <u>all</u> the overhead bits.

It is generally thought that the headers used with data packets are larger than is necessary for use with speech packets. Error detection for the coded speech is not necessary unless the expected error rate is exceptionally high. Even then, the increase in delay due to the retransmission of corrupted packets would be undesirable. Thus neither the source address nor the error check bits are needed.

Even the destination address of the packet can be shortened, since it is accepted that virtual-call routing is the best packet-routing strategy for speech [2, 15]. In a virtual call, the route a packet takes through a network is decided on the basis of traffic levels at the time of setting up the call, and the route remains fixed for the duration of the call. If the packet header has to contain the routing details, it is of course considerably lengthened. However, if each node on the packet's route contains the routing details, a shortened address can be used in the header that is changed at each node according to a simple look-up routing table [2, 15, 19]. Such a scheme does not perform as well as a variable-routing strategy when a network is lightly loaded, but performs better when a network is heavily loaded [2]. This is especially true for speech, where instantaneous traffic levels do not vary as much as for data.

It is possible to shorten further the packet header by grouping all the speech packets transmitted on a link together into one large packet [21]. A single bit in the header of the large packet indicates whether a call is transmitting a packet or not in that particular group of packets. Even though extra bits are needed to protect the header against errors, the header per packet can be as small as 4 bits. If the packets are permitted to be of variable length, the header will have to be somewhat larger.

Typical values for headers for single speech packets are 16 to 32 bits (3, 21, 40), rising to 64 bits if X25 protocols must be observed (15). The question of the trade-off between delay and header overhead has been studied by a few researchers (21, 24, 26). Their work will be discussed in section 1.2.6, and the problem is explained in some detail in section 4.5.

Ignoring the packet header, transmission delay is equal to the packetisation delay divided by the link capacity (measured in no. of calls). As it is proportional to packetisation delay, there is no need to consider it seperately.

1.2.4 Queueing Delay

Packets queue for transmission for one of two reasons. Firstly, the capacity of the outgoing link may be plenty to carry the traffic, but the near-simultaneous arrival of two packets means that one of them must queue, and this in turn means that others may have to queue. In this case the queueing time is never more than the time taken to transmit one or two packets, which will be small compared to the packetisation delay. Secondly, the capacity of the outgoing link may not be enough to carry the traffic, and packets have to queue until the traffic level goes down (i.e. the number of active talkers drops below the channel capacity). In this case the queueing time may be quite considerable. Consequently, the queueing delay considered will be this second case, the first case not introducing enough delay to be significant.

There are different opinions as to how the queueing delay should be controlled (besides the preventative measure of call blocking). Many papers (15, 22, 26, 27, 41-43) have suggested that the queueing delay of a packet should not be controlled in the network, but that the receiver should discard packets delayed more than a certain maximum delay. These proposals invariably make the assumption, based on experience with ARPANET (44), that the delay of a packet is independent of the delay of the packets preceding it. However, this assumption is not likely to be valid for a network designed for speech, for three reasons.

Firstly, assuming a virtual-call policy is used, a packet cannot arrive out of order (15). Secondly, if speech packets have priority over data packets, the speech-packet queues will tend to be either steadily increasing (if no. of talkers > link capacity) or steadily decreasing (if no. of talkers < link capacity) over short periods of time (9). Thirdly, packet lengths would rarely be as long as those used for ARPANET (134 ms (44)), and so there is less time for network queues to change between packets. It is therefore necessary to reconsider the strategy of losing packets at the receiver in the light of this.

The other approach to controlling delay is to limit the size of each queue in the network. This can be done simply by discarding packets evenly over all channels whenever the queue is about to exceed its limit (3, 9). Thus, packets are discarded before they are transmitted, rather than afterwards. It would appear that the only disadvantage of this approach is that a little more processing is needed at the switching node.

Whichever approach to controlling queueing delay is used, it is clear that there is a trade-off between delay and packet loss (9, 45) since, as the delay limit is increased, the packet loss is reduced. Thus, for a given packet loss, the TASI advantage increases as the delay limit is increased.

It appears that the increase in TASI advantage achieved by allowing large queueing delays is not always very great. Weinstein and Hofstetter (9) investigated the trade-offs between delay, packet loss and TASI advantage for a single link with a limited queue by using simulations. (Their work has subsequently been largely verified analytically by Janakiraman et al. (46, 47).) They found that increasing the maximum delay from zero up to 500 ms made little difference to the TASI advantage when the link capacity was high (>40 calls), although it made considerable difference when the link capacity was low (4 calls). However, the 0.5% packet loss allowed was rather small and it is not obvious how different the trade-offs are at larger levels of packet loss.

1.2.5 Receiver Buffering

Whichever method is used to control queueing delay, there is always the problem of how much to buffer the received packets in order to reach the best compromise between delay and gaps in the output speech. Fundamentally, there are two approaches (43). The simplest is to buffer the first packet of a talkspurt by an amount independent of the delay the packet has already incurred. Since no information on network delays is needed this approach is called NTI - null timing information (43). Naylor and Kleinrock (45) have found an empirical rule for determining the best buffering delay, based on the delay variations of the packets in the previous talkspurt. Although this rule works well for delays on ARPANET, it is uncertain if it would work well for the correlated delays of a packet-switched speech network.

The other approach is to include a few bits of information in the packet header indicating at what time it was sent, called a "timestamp". Each packet can then be buffered until its total delay is equal to a "control time" and then played out. This approach is called CTI - complete timing information (43) (if transmitter and receiver clocks are not synchronised it is called ITI - incomplete timing information (43)). Of course, any packet arriving already delayed by more than the control time would cause a gap in the output speech. The control time would have to change as the loading of the network changed. Again, the analysis presented in references (42) and (43) cannot be applied directly to a packet-switched speech network, as it assumes independent delays. However, it is apparent that CTI is the better of the two schemes (48).

Suda et al. (24) used simulation to study both approaches to receiver buffering. Unfortunately, a close look at their simulated network reveals that all the links are operated with a TASI advantage less than 1. Thus the queueing was due to the first cause described in section 1.2.4, and it was never more than a few ms.

Consequently, the question of how the buffering schemes would operate in a well-utilised packet-switched speech network remains unanswered.

In a network where each queue has a maximum delay, it would make sense for the control time to be equal to the maximum delay, as then there would be no variation in delay at all. Whether or not this is the optimum solution cannot be discovered from the literature. However, it would be a simple and reliable approach.

One interesting way to overcome the variable network delay is to buffer whole talkspurts instead of individual packets. This technique, called "buffered TASI" [49, 50] or "transparent message switching" (TMS) [2], is not usually suggested in the context of packet switching, but it can easily be adapted to operate with packet switching. The effect of buffering whole talkspurts is to cause the variable delays to affect only silent periods, which is acceptable so long as the change in silence period is less than 50% [8]. The drawback of TMS is that delays will tend to be concentrated onto a few channels at a time, rather than spread evenly over all channels as in conventional packet-switching. Thus, the variation in delay will be greater and more packet loss would have to take place for a given delay limit.

1.2.6 The Analysis of Queueing Delay

The accurate prediction of delay (and therefore packet loss) is crucial in order to optimise the performance of a packet-switched speech network, and to know when packet switching has significant advantages over other SI techniques. Whilst it is a simple matter to calculate packetisation, transmission and propogation delays, it is difficult to predict queueing delays with any accuracy. This section reviews and criticizes the methods used to do this.

The literature falls into four categories. Two papers analyse the buffered-talkspurt queue using a finite-population M/M/m queueing model (49, 50). The others are concerned with the conventional speech-packet queue. Four use one-dimensional queueing models (21, 40, 52, 53), two use a two-dimensional model (46, 54) and three use simulation techniques (9, 24, 40).

Only one paper does not assume a constant number of calls (50). However, this paper shows that accurate results can be obtained by assuming a constant number of calls and multiplying by the probability of that number of calls occuring. This justifies the assumption made in the other papers that variations in the number of calls are slow compared to the variations in queue length, and therefore the number of calls can be considered constant.

To analyse the buffered talkspurt queue using the finite-population M/M/m model (49, 50), it must be assumed that talkspurt and silence distributions are both exponential. In the light of Brady's work (55) this seems a reasonable approximation for talkspurts, but rather a poor one for silences. In fact Gruber showed (56) that silence distributions are best described by the weighted sum of two geometric distributions.

It must also be assumed that a speaker does not start to speak until his previous talkspurt has been transmitted. The greater the efficiency of the link, the more unrealistic is this assumption. For instance, with an efficiency of 1 the queueing delay should be infinite, but the finite-population M/M/m model predicts that approximately 40% of talkspurts suffer no delay at all! However, it was shown in reference (49) using simulation that at an efficiency of 0.72 this assumption is valid. Finally in reference (49), when calculating the amount of speech loss caused by fixing a maximum queueing delay, it is assumed that speech loss does not affect the queueing delay (although implied by the mathematics this assumption is never stated). Again, a simulation in reference (49) showed that at very low levels of speech loss (less than 0.5%) this assumption is valid, though at higher levels of loss it is unlikely still to be valid.

The conclusion from this is that for queues operating at moderate efficiencies and with very little speech loss the finite-population M/M/m model fits adequately. However, for high efficiencies and moderate levels of speech loss, the model is too inaccurate to be of use.

To use a one-dimensional model for a speech-packet queue it must be assumed that packet arrivals are independent, and this is the fundamental problem. As a talkspurt will typically consist of many packets, packet arrivals will be highly correlated. It is interesting to note that no-one who used a one-dimensional model admitted that it is unrealistic to assume independent arrivals.

Coviello [20, 21] assumed an M/M/1 model which he claimed would provide an upper bound to performance. Although this would be true if arrivals were independent, it is nonsense since arrivals are correlated. Kim [53] and Seguel et al. [40] assumed an M/D/1 model. Seguel et al. tested the assumption with simulation and found it only gave accurate results when there was no congestion and therefore hardly any delay. Where there were significant amounts of delay, the M/D/1 model hopelessly underestimated queueing delay. Kim also tested the assumption with simulation, but only considered the cases where there was no congestion! A fair comment on the M/D/1 model is that it can predict the first kind of queueing delay mentioned in Section 1.2.4 but not the second kind, which is the more important one.

Minoli has published three papers which assume a geometric arrival process [26, 27, 52]. Unlike Kim and Seguel et al., he does not check his assumptions using simulation. Minoli's original paper on the geometric arrival, constant server queue [52] presents an analytical solution to the queueing model which he admits, at the end of the paper, he never used since it is quicker to use an iterative numerical method.

In Minoli's later paper "Issues in packet voice communication" [27] he claims it to be a fact that "each speaker's behaviour is an independent Bernoulli trial with a probability p/p+r of supplying a nonempty packet" (p.732; p/p+r is the speech activity). He then proceeds to present the analytical solution to his model rather than the much more useful numerical solution. Besides this, he gives a formula for mean delay (p.735) from his original paper without stating that it is only true for one extremely specialised condition ($\Delta = 1$ in his terminology). Finally, he compares his solution with other models, including the M/D/1, and then concludes that these models "predict poor (sic) performance than can be actually attained" (p.739). However, it is clear from the work of Seguel et al. [40] that quite the reverse is true, i.e. the M/D/1 model predicts far better performance than can actually be attained.

Where Minoli considers the queue length to be restricted to just 2,3, or 4 packets (p.737), it is possible that his analysis is quite accurate. This is because then several packets (one from each active channel) separate the correlated packet arrivals of any given channel. Thus, with such a restricted queue length, the correlation could have very little effect.

Minoli also uses his unrestricted queue results to determine the "optimal packet length" in reference [26]. Again, he quotes a formula for mean delay without stating that it is only true for $\Delta = 1$. However, not only is his model very inaccurate, but his justification of using average delay instead of a 95th or 99th percentile is based on an example where the TASI advantage is only 0.9! It is no surprise in this case that the optimum packet length is roughly the same for average delay and the two percentiles. For such a low TASI advantage, the packet length must be very short before any significant amount of queueing takes place. Then, when the packet is so short that the header is over half the packet length, a reduction in packet length of a few bits inevitably sends the link into congestion. Thus, any queueing model and any percentile would give an optimum packet length roughly twice the header size! Besides this, the knee of the delay versus packet length graph would be very much less sharp for the TASI advantages near 2.0 and headers smaller than 100 bits that were used in his results. Thus, his "representative" curve vastly exagerates the importance of packet length. It must be concluded that Minoli's results are invalid.

An interesting consequence of using a one-dimensional queueing model is that queueing delay becomes a function of service time, and therefore packet length. This is one of the fundemental bases of both Coviello's (21) and Minoli's (26, 27) results. Since the correlation in talkspurt arrivals is also a function of packet length, it remains to be seen whether or not packet length affects the queueing process. As Weinstein and Hofstetter point out (9) it seems very unlikely that it should , especially when short (i.e. < 50ms) packets are used.

A two-dimensional queueing model (46, 54) is able to look beyond the packet-arrival process to the talkspurt-generation process. Weinstein (57) found that the number of active talkers (i.e. those generating talkspurts) can be modelled very well by a Markov chain. This in turn means that the number of packets arriving per frame can also be modelled by a Markov chain (where, as in time division multiplexing, a frame is the time interval between consecutive packet arrivals on channel no.1). The length of the queue at the end of a frame is governed only by the length at the start of the frame and the number of packet arrivals during the frame. Therefore, the queue behaviour can be modelled as a two-dimensional Markov process involving both the queue length and the number of packet arrivals per frame.

The main drawback of this model is that it cannot take into account <u>when</u> each packet arrives in a frame. Consequently it is not able to model short queues well. Janakiraman et al. (46) got around this problem by assuming that all the packets of a frame are placed in the buffer at the same time and that they are all serviced at the same time. On the other hand, Tanaka et al. (54) simply admit that the model is only an approximation to a real arrival process, and justify their results with simulation.

The state equations are easily formed, and in theory can be solved so long as the queue length is bounded. However, this does require a lot of computation. Tanaka et al. (54) only considered small queues and small channel capacities, and so the computation would not have been too much problem. However, Janakiraman et al. (47) used their analysis to try to verify the simulations of Weinstein and Hofstetter (9). It is very interesting to note that the largest channel capacity they considered was 20, although the simulations went as high as 40. With 20 channels, a packet size of 20ms and a maximum delay of 500ms, the state equations involved 20,000 unknowns. To solve these equations, 5.4×10^6 multiplications and the same number of additions would be needed per iteration. (This figure can easily be calculated from Eq 24 in reference (46). It is much less than 4×10^8 because the state matrix is sparse.) With 40 channels, the state equations would involve 80,000 unknowns and 43×10^6 multiplications per iteration to solve them. Whether or not it was the large amount of computation or the large arrays required that discouraged Janakiraman et al. from computing the 40 channel case is not stated. However, it is clear that for a comprehensive study of queue behaviour that involves large channel capacities and queues, this approach is not very satisfactory.

One additional drawback of the two-dimensional model used by Tanaka et al. and Janakiraman et al. is that it is not possible to compute the queue distribution of an unbounded queue. This is because the number of queue states would be infinite. However, as in any practical system the queue must be restricted at least by the amount of memory (i.e. buffers) available, this is not a very serious limitation.

The simulation studies ought to be the most reliable studies, although even these are based on speech models rather than real conversations. Also, none of the authors give confidence limits to their results.

As mentioned in section 1.2.5, the work of Suda et al. (24) is not very helpful, as they do not consider a single instance where the TASI advantage is greater than 1. It is also important to note that their optimal packet length is similar to Minoli's (26) for one reason only. In their figures 5 and 6 the queueing delay is significant compared to the packetisation delay only for the last 20 bits of the packet-length axis. Thus, the packet length would be similar whatever the queueing model used! It does not validate either theirs or Minoli's work.

The simulations of Seguel et al. (40) are useful, but only relate to two channel sizes, both fairly large (24 and 30 channels). The simulations of Weinstein and Hofstetter (9) cover a good range of channel sizes and show very clearly how delay and channel size affects the TASI advantage, although (as stated in section 1.2.4) the 0.5% loss used is rather small. However, the analytical results of Janakiraman et al. (47) do not agree with Weinstein and Hofstetter's. Janakiraman et al. put the discrepancies down to a difference in model, although the only significant difference is the packet arrival and service strategy mentioned earlier. Whilst this might be expected to make a small difference when the queueing delay is limited to 20ms or so, it can hardly make a significant difference at a queueing limit of 500ms. Therefore one or both sets of results must be in error. As no confidence limits are given for the simulation, and as some interpolation must have been used to obtain exactly 0.5% loss, it is the simulation results that are the more suspect.

All this emphasizes that the usefulness of published simulation results is limited by the particular parameters chosen and the difficulty of verifying the results. On the other hand, to repeat them with different parameters and a number of independent runs to gain confidence limits requires very large amounts of computation. The two-dimensional model and its resulting analysis requires much less computation and is preferable to using simulation. It would appear that the authors of these three simulation studies were not aware of its existence at the time of their work.

1.3 FACTORS THAT AFFECT SPEECH QUALITY IN A PACKET-SWITCHED SPEECH NETWORK

1.3.1 Packet Loss and Other Packet-Switched Speech Impairments

As mentioned in section 1.2.4, packet loss can occur either in the network to control the length of queues or at the receiver to avoid waiting for late packets. It is interesting to note that in neither case are the losses likely to occur randomly. In the queue, the losses will occur continuously whenever the queue is on its limit and the number of talkers exceeds the channel capacity (e.g. one in four packets would be lost if five people talked and the channel had capacity for four). At the receiver the losses are likely to often follow one another due to correlations in network delay. In spite of this, all the tests on packet losses (5, 58-61) have assumed random losses, which means the results may not be totally applicable to a packet-switched speech network.

Other impairments besides packet loss may be introduced at the receiver. One is buffer underflow, where gaps are introduced into the speech but no speech is lost. This effect has not been widely investigated, although it seems reasonable that it would affect the speech in a very similar way to packet loss. This is consistent with the results of some informal tests (58). Another impairment is "silence interval modification" (51) or "variable talkspurt delay" (59) where silence lengths are changed. So long as no silence is changed by more than 50%, it would appear that this impairment is acceptable (51). Both buffer underflow and variable talkspurt delay can be reduced by receiver buffering.

Packet loss is therefore the most important impairment of packet switching. There are three main issues concerning the acceptability of packet losses. These are the effect of packet length, the way in which the gaps caused by lost packets should be filled, and the amount of loss that is acceptable.

The published work on the acceptability of packet losses and "interrupted" (62) or "reiterated" (63) speech (which are in essence the same as packet losses) does not present a very clear picture on any of these issues. This is partly because some give intelligibility scores and others give opinion scores and partly because the actual test used differs widely.

On the question of packet length, all the published work agrees that packets less than 50ms in length give the most acceptable losses. Both intelligibility scores, (62) and opinion scores (59) have shown this. However, from informal opinion tests (5) 16 to 32 ms is the best length, from intelligibility tests (62) 5 to 50ms packets degrade speech equally, and from formal opinion tests (59) 4ms is the best length. Thus, within the range 4 to 50ms, there is little agreement on the best packet length.

The main issue of gap filling is whether silence should be inserted in the gaps, or something else. The possible alternatives to silence are a repetition of the previous packet (58, 60, 61), a repetitiion of a small part of the previous packet (60, 61), or interpolation between every other speech sample if odd and even samples are transmitted in different packets [5]. A comparison between reference (61) and reference (5) suggests that the first and third alternatives are about equal, although the first is much more simple. The second alternative was not successful when as little as 1.5ms was repeated [61]. With vocoders, the effect of repeating the last packet can be improved on by using an unvoiced excitation signal [58].

The main alternative to inserting silence into gaps is therefore repeating the previous packet. Intelligibility tests using both hi-fi headphones [63] and telephone receivers [60, 61] have failed to show which is better (mainly because the intelligibility scores for both are quite high). However, opinion tests [60, 61] have shown that for packets shorter than 10ms it is more acceptable to repeat the previous packet than insert silence.

The amount of loss that is acceptable appears to depend completely on what is meant by "acceptable". From an intelligibility point of view, packet losses of up to 50% are acceptable as they still allow 95% of words to be understood correctly [63]. Informal listening tests have shown [5, 61] that levels of 5% or 10% loss are not objectionable (so long as gaps are filled with repeated packets) and should therefore be acceptable to most people. However, it has been found from formal listening tests [60] that even 2% loss with repeated packets is unacceptable to 20% of subjects. It must be pointed out though that in the same test, no loss at all was unacceptable to 10% of subjects! With silence in the gaps, 2% loss is unacceptable to almost 50% of people [60], and 10% loss is not acceptable to anybody [59, 60].

The poor scores for packet losses in the formal tests may well be caused by the tests themselves. In neither case was the standard "listening effort" opinion scale used as is recommended in reference (34). Indeed, in reference (60) a binary acceptable/unacceptable decision was used, as discussed in a paper by Reiss and Klemmer [64]. However, the tests of Reiss and Klemmer were conversational, where an acceptable/unacceptable decision has considerably more meaning than in the listening test used in reference (60).

A conservative conclusion from all these results would be that losses up to 2% are probably acceptable to most people. Even this figure is four times higher than the loss allowed with TASI [4].

1.3.2 Embedded Encoding

As has already been mentioned in section 1.1, a more sophisticated form of speech loss can be used with packet switching if embedded encoding is used. Then, instead of discarding whole packets to limit queue length, the last part of every packet in a queue can be discarded. Although the speech quality is impaired, no speech is lost. For a waveform coder, this would probably mean grouping the most significant bits (MSB) of each sample at the front of the packet, and the least significant bits (LSB) at the end. If prediction were used, the coder would have to be modified [5, 7]. With a vocoder, it could be the residual or baseband signal that goes at the end of a packet [6].

Since a long period of slightly-degraded speech is preferable to short periods of poor-quality speech, it has been suggested that the packet-shortening process should begin before the queue reaches its limit [40]. Once the limit has been reached, a few packets might have to be lost altogether, or at least quite badly degraded.

It has also been suggested [6, 15, 50, 65] that the bit rate of the coder can be changed according to the network load. This avoids transmitting long packets part of the way through the network only for them to be shortened at a later link. However, for transmission capacity to be used efficiently, speech interpolation will always be needed. Thus packet loss or embedded encoding will always be necessary, even if the coder bit-rate can vary with network load.

Some DSI systems [8] consider the requirements of the actual speech signals when squeezing all the active speech channels onto a link. The most bits are lost on the channels with the highest signal-to-noise ratios (SNR), in order to maintain as high a SNR as possible overall. This same idea could be used with embedded encoding; this would be similar to the variable-rate scheme proposed in reference (10).

1.3.3. Silence Detection and Echo

It has been assumed throughout the previous discussion that it is always possible to separate speech from silence. The presence of background noise and echo can make this very difficult. If a level detector is used, the level is set above the noise, and the speech-to-silence transition is delayed at the end of each talkspurt to capture any low-level sounds that may have been there. This delay is called hangover. Modern detectors can distinguish speech from background noise to some extent (65, 66), and little hangover is necessary. The advantage of this is that the speech activity is reduced, and so the TASI advantage obtainable is increased.

Some DSI systems do not make a speech/silence decision, but simply reduce the bit rate for small signals (8). Since packets can be of variable length, a similar system could be used with packet-switched speech in circumstances when the background noise is of too high a level to be distinguished from speech. Ideally, the bit rate of the speech coder should match the certainty with which the detector assesses the signal as speech.

Echo is caused mainly by two-wire to four-wire transformations, and will be a problem until digital transmission to the telephone set becomes widespread. In private networks this may be quite soon, but in the public networks of the world it may never happen because of the tremendous reinvestment needed.

Without delay, echo is only a problem for speech detectors, and these can to some extent distinguish echo by its signal level. However, with delay, echo becomes very distracting and must be highly attenuated or preferably eliminated completely for comfortable conversation to take place (64, 68). Echo suppressors, which attenuate whichever signal path is weaker, have been in use on long-distance links for many years. Nowadays echo cancellors [29, 30], which hardly affect a conversation at all, are being used on some satellite circuits. As faster and cheaper signal-processing integrated circuits are produced, it will be possible to cancel echos without great expense. Certainly, it seems reasonable to assume that echo need not be a major contribution towards the degradation incurred over a packet-switched speech network.

1.4 AIMS OF PRESENT RESEARCH

The aim of this thesis is to show how packet switching can be used for speech to make the most efficient use of transmission capacity. To do this, an accurate method for the analysis of a speech-packet queue that does not require excessive amounts of computation is described in Chapter 2. Chapter 3 describes tests on packet losses that not only clarify the rather confused picture of the effects of losses portrayed by the literature, but use loss distributions that would be found on a real packet-switched speech link. Chapter 4 then uses the analysis and the results of the tests to show both how packet switching can make best use of transmission capacity, and the circumstances under which significantly less channel capacity is needed compared to other SI systems.

Packet-switched speech has other applications besides the integrated speech/data network considered in this thesis. These include local-area networks (LANs) [61, 69-73], satellite systems [65], military systems [3], mobile radio [74, 75], voice messaging [3] and the use of data networks for speech transmission [16, 23, 76-78]. Some of the results of Chapter 4 will be applicable to these uses, but none of the particular problems encountered in them will be considered.

CHAPTER TWO

AN ACCURATE METHOD FOR THE ANALYSIS OF A SPEECH-PACKET QUEUE

CHAPTER 2

AN ACCURATE METHOD FOR THE ANALYSIS OF A SPEECH-PACKET QUEUE

2.1 INTRODUCTION

In Section 1.2.6 various methods were discussed for analysing the behaviour of a conventional packet speech queue. It was seen that a two-dimensional model gives more accurate results than a one-dimensional model, and requires less computation than a simulation. However, it still requires a considerable amount of computation when large channel capacities and queues are analysed. For example, a link with a capacity of 40 channels and a maximum delay of 500 ms with 20 ms packets requires a set of 80,000 state equations to be solved, involving 43×10^6 multiplications and additions per iteration.

It will be shown in this chapter that by modelling the queue length as a continuous quantity, the state equations can be solved with considerably less computation. For the example quoted above a set of 80 state equations are formed (after some initial matrix calculations which do not need to be repeated for different queue lengths) involving 6.4×10^3 multiplications and additions per iteration. On the computing facilities available this involved a few seconds of computation instead of possibly several hours. (The discrete method was never used as its computational requirement for large channels and queues appeared prohibitive.)

As mentioned in Section 1.2.6 the basis of the two-dimensional model is that the number of active talkers i out of v speakers (i.e. off-hook calls) can be modelled by a Markov process. For a continuous model this becomes a birth-death process. If

$$\frac{1}{\mu} = \text{mean talkspurt length}$$

$$\frac{1}{\lambda} = \text{mean silence length}$$

$$p(i,j) = \text{transition rate from state } i \text{ to state } j$$

the birth and death rates are given by[57]

$$p(i,i+1) = (v-i)\lambda \qquad i \neq v$$

$$p(i,i-1) = i\mu \qquad i \neq 0$$

as shown in Fig. 2.1.

Since a speaker activity model is being used, it must be assumed that either all the packets originate at the node considered, or that they have travelled over links without any significant amount of queueing or packet loss. If they have already undergone queueing delays and packet loss, the effect would most likely be to reduce rather than increase the queueing delay. This is because the previous queueing and packet loss would have tended to smooth out the variations in traffic levels (unless the packets had all been queued in different queues, but this is unlikely).

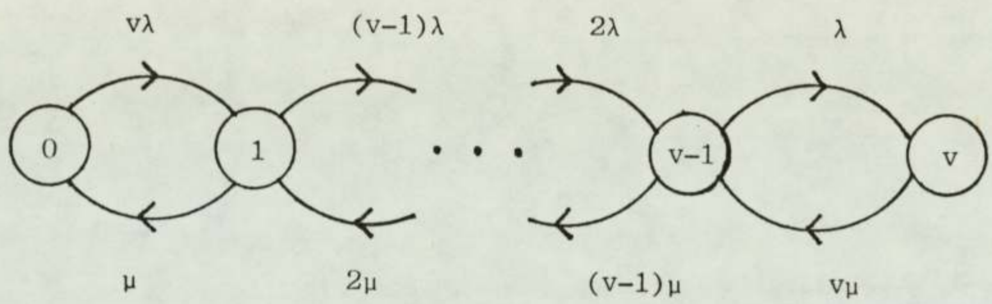


Fig. 2.1 Talker activity model

It is also assumed that the speech coding rate is fixed. As variations in traffic level due to variable-rate coding would be faster than variations due to talkspurts and silences, this is a reasonable approximation to make. Besides this, without a knowledge of how the bit-rate varies, it would be impossible to take the variations into account. The header overhead is not taken into account, as it can be added later by suitably scaling the channel capacity.

Finally, it is assumed that the number of calls is fixed. As pointed out in Section 1.2.6, this is not only a very common assumption, but one justified by results shown in reference (50) that show that variations in the number of calls do not significantly interact with the operation of the queue.

2.2 DERIVATION OF STATE EQUATIONS

Before deriving the state equations all the transition rates of the talker activity model need to be defined:

$$p(i,i+1) = (v-i)\lambda \qquad i \neq v$$

$$p(i,i-1) = i\mu \qquad i \neq 0$$

$$p(i,i) = 1-((v-i)\lambda + i\mu)$$

$$p(i,j) = 0 \qquad j<i-1 \text{ and } j>i+1$$

We also define $p^*(i)$ as the total flow-rate out of state i

$$p^*(i) = p(i,i+1)+p(i,i-1) = (v-i)\lambda+i\mu$$

also

$$p^*(i) = 1-p(i,i)$$

For large values of v the talker activity model is valid independent of talkspurt and silence distributions, but for small v ($v<25$) it is more accurate the closer these distributions are to exponential[57]. The validity of the talker activity model is checked in Section 2.8 for values of v down to 5.

The units of λ and μ could be $\sec^{-1}$, but it is useful to define λ in terms of μ , i.e. replace λ by $\frac{\lambda}{\mu}$ and μ by 1. The queueing delay may also be defined in terms of μ^{-1} , making the effect of μ very clear. This approach is taken throughout the thesis.

The queue length will be modelled as a continuous quantity by letting the packet length (and therefore the frame length) tend to zero. Let $q(t,x,j)$ be the p.d.f. of queue length x and number of speakers j at time t, and let Δt be the packet length (length here refers to the total amount of speech in the queue or packet, and is measured in time units). Also, let h be the capacity of the link (i.e. the number of packets that can be transmitted in one frame). Then for small Δt,

$$\begin{aligned} q(t+\Delta t,x,j) = \; & q(t,x-(j-1-h)\Delta t,j-1)p(j-1,j)\Delta t+ \\ & q(t,x-(j+1-h)\Delta t,j+1)p(j+1,j)\Delta t+ \\ & q(t,x-(j-h)\Delta t,j)(1-p^*(j)\Delta t)+o(\Delta t) \\ & \qquad x>(j+1-h)\Delta t \end{aligned} \tag{2.1}$$

Rearranging (2.1) gives

$$\begin{aligned} q(t+\Delta t,x,j)-q(t,x-(j-h)\Delta t,j) = \; & q(t,x-(j-1-h)\Delta t,j-1)p(j-1,j)\Delta t \\ & + q(t,x-(j+1-h)\Delta t,j+1)p(j+1,j)\Delta t \\ & - q(t,x-(j-h)\Delta t,j)p^*(j)\Delta t+o(\Delta t) \\ & \qquad x>(j+1-h)\Delta t \end{aligned}$$

Dividing both sides by Δt and letting $\Delta t \to 0$ gives

$$\frac{\partial q}{\partial t} +(j-h)\frac{\partial q}{\partial x} = q(t,x,j-1)p(j-1,j)+p(t,x,j+1)p(j+1,j)-q(t,x,j)p^*(j) \qquad x>0$$

Let $q_j(x) = \lim_{t\to\infty} q(t,x,j)$. By setting $\frac{\partial q}{\partial t} = 0$ and letting $t\to\infty$ we get the equilibrium solution given by:-

$$(j-h)\frac{\partial q}{\partial x} = q_{j-1}(x)p(j-1,j)+q_{j+1}(x)p(j+1,j)-q_j(x)p^*(j) \tag{2.2}$$

Equation (2.2) can also be written

$$\underline{q}'(x) = R\underline{q}(x) \tag{2.3}$$

where

$$\underline{q}(x) = \begin{bmatrix} q_0(x) \\ q_1(x) \\ \cdot \\ \cdot \\ \cdot \\ q_v(x) \end{bmatrix}$$

and

$$R = \begin{bmatrix} \frac{-p^*(0)}{-h} & \frac{p(1,0)}{-h} & 0 & 0 & 0 \\ \frac{p(0,1)}{1-h} & \frac{-p^*(1)}{1-h} & \frac{p(2,1)}{1-h} & 0 & 0 \\ 0 & \frac{p(1,2)}{2-h} & \frac{-p^*(2)}{2-h} & \frac{p(v-1,v-2)}{(v-2)-h} & 0 \\ 0 & 0 & \frac{p(2,3)}{3-h} & \frac{-p^*(v-1)}{(v-1)-h} & \frac{p(v,v-1)}{(v-1)-h} \\ 0 & 0 & 0 & \frac{p(v-1,v)}{v-h} & \frac{-p^*(v)}{v-h} \end{bmatrix}$$

If h is an integer (as is often the case), the (h+1)th row will be infinite. The reason for this is that the LHS of equaton (2.2) is zero, indicating that $q_h(x)$ is simply a linear combination of $q_{h-1}(x)$ and $q_{h+1}(x)$ given by

$$q_h(x) = \frac{(q_{h-1}(x)p(h-1,h) + q_{h+1}(x)p(h+1,h))}{p^*(h)} \tag{2.4}$$

Because of this relationship, it is unnecessary to include $q_h(x)$ in equation (2.3), so it can be removed. This entails removing the (h+1)th row and column of R and altering the $q_{h+1}(x)$ and $q_{h-1}(x)$ terms according to equation (2.4) to compensate for the missing $q_h(x)$ terms. Once the solution to equation (2.3) has been found, $q_h(x)$ can easily be recovered by using equation (2.4).

It will be assumed that h is an integer. The analysis for the case when h is non-integer is, if anything, more straightforward.

2.3 SOLUTION OF STATE EQUATIONS

Let λ_K be an eigenvalue of R and $\underline{c}_K$ be its corresponding eigenvector. Since the dimension of $\underline{q}(x)$ is v, equation (2.3) has v solutions of the form

$$\underline{q}(x) = \underline{c}\ e^{\lambda x}$$

The general solution is a linear combination of these v solutions. If a_K is the scaling constant for the eigenvector $\underline{c}_K$, the general solution is

$$\underline{q}(x) = \sum_{K=1}^{v} a_K\ \underline{c}_K e^{\lambda_K x} \tag{2.5}$$

Whilst the λ_K and $\underline{c}_K$ can be found using standard eigenvalue and eigenvector algorithms, the a_K are chosen by defining boundary conditions, as described in Sections 2.4 to 2.6.

The overall queue-length distribution is the sum of the individual elements of $\underline{q}(x)$:

$$q(x) = \sum_{j=0}^{v} q_j(x) \tag{2.6}$$

Equations (2.5) and (2.6) do not quite tell the whole story. We also need to define the point probabilities π_j that there is no queue at all and, if the queue has a limit, point probabilities Θ_j that the queue is actually on the limit. Thus, the overall probabilities of no queue and maximum queue, q_π and q_Θ, which are given by

$$q_\pi = \sum_{j=0}^{v} \pi_j \quad \text{and} \quad q_\Theta = \sum_{j=0}^{v} \Theta_j$$

are needed to supplement the p.d.f. of equation (2.6). The methods for calculating the π_j and Θ_j will be explained in the sections on boundary equations ((2.4) to (2.6)).

Queue length can easily be converted to queueing delay by substituting $h \times t$ for x. However, the p.d.f. of delay is not simply a scaled version of the p.d.f. of queue length. This is because the distribution of queue length seen by a packet joining the queue is different from the distribution seen by an outsider. For instance, the state $q_0(x)$ can never be found by an arriving packet, as $j=0$ excludes any packet arrivals. Thus, $q_0(x)$ cannot make any contribution towards the delay distribution.

The overall delay distribution $d(t)$ is found by scaling each $q_j(ht)$ by $\frac{j}{j_{ave}}$ before summing (where j_{ave} is the average number of speakers). This gives

$$d(t) = \sum_{j=0}^{v} \frac{j}{j_{ave}} q_j(ht)$$

Similarly, the probabilities of no delay and maximum delay, d_π and d_Θ, are given by

$$d_\pi = \sum_{j=0}^{v} \frac{j}{j_{ave}} \pi_j$$

and

$$d_\Theta = \sum_{j=0}^{v} \frac{j}{j_{ave}} \Theta_j \ .$$

2.4 BOUNDARY EQUATIONS FOR AN UNLIMITED QUEUE

Whilst is any practical queue the queue length cannot be unlimited (due to finite buffers), it is useful to be able to analyse such a system (as in Section 4.1).

So long as the link is not permenantly overloaded, each $q_j(x)$ must tend to zero as x tends to infinity. It can be seen immediately from equation (2.5) that the scaling constants a_K corresponding to all the positive eigenvalues must be zero for this to be true. Since there are (v-h) negative eigenvalues, there are (v-h) non-zero scaling constants which have to be found by solving boundary equations.

A rough sketch of the $q_j(x)$ is given in Fig. 2.2. The point probabilities π_j at x=0 represent the probability of no queue. These are zero for j>h, since the queue must always be increasing for these states.

The (v-h) boundary equations can be found by integrating each $q_j(x)$ for j>h and setting the result equal to the overall probability of being in state j, b_j, where b_j is given by the binomial probability $b(j,v,\alpha)$ where $\alpha = \frac{j_{ave}}{v}$, the average fraction of people talking. We have therefore, for h<j<v,

$$\int_0^\infty q_j(x)dx = \int_0^\infty \sum_{K=1}^{v-h} a_K c_{jK} e^{\lambda_K x} dx = \sum_{K=1}^{v-h} -a_K c_{jK}/\lambda_K = b_j$$

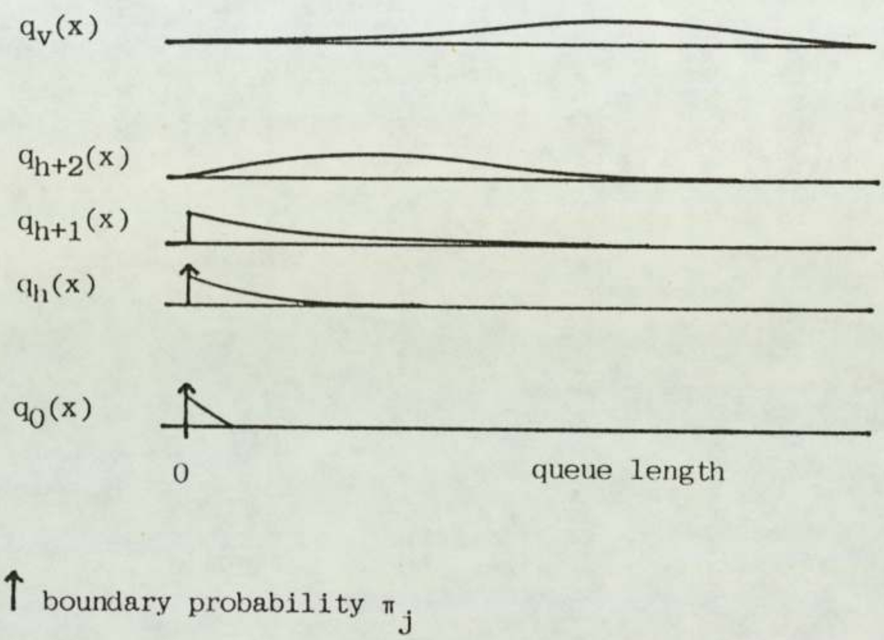


Fig 2.2 Rough sketches of queue distributions for an unlimited queue.

where c_{jK} is the jth element of the Kth eigenvector. These equations can be solved using standard numerical methods to yield the (v-h) constants a_K.

Once the a_K have been found, the unknown point probabilities $\pi_j (j<h)$ can be easily found by subtracting $\int_0^\infty q_j(x)$ from b_j for each $j<h$. This gives

$$\pi_j = b_j - \sum_{K=1}^{v-h} - a_K c_{jK}/\lambda_K = b_j + \sum_{K=1}^{v-h} a_K c_{jK}/\lambda_K \qquad 0<j<h$$

2.5 BOUNDARY EQUATIONS FOR LIMITED QUEUES WITH PACKET LOSS

Here we derive the boundary equations for a queue that is prevented from exceeding a specified length m by discarding packets. The approach taken for the unlimited queue is of no use. The queue is finite and all v eigenvalues must be used; also there are queue-limit point probabilities as well as zero-queue point probabilities.

Sketches of the queue distributions are shown in Fig. 2.3. All but two of the boundary equations can be formed immediately by observing that

$$\lim_{x \to 0} q_j(x) = 0 \qquad j>h+2$$

$$\lim_{x \to m} q_j(x) = 0 \qquad j<h-2$$

Since $q_j(x)$ is continuous up to the boundaries, we can write

$$q_j(0) = \sum_{K=1}^{v} a_K c_{jK} = 0 \qquad 0<j<h-2 \tag{2.7}$$

and

$$q_j(m) = \sum_{K=1}^{v} a_K c_{jK} e^{\lambda_K m} = 0 \qquad h+2<j<v \tag{2.8}$$

The remaining two boundary equations can only be formed through a rather involved procedure. The general approach is to integrate two of the $q_j(x)$, add in the appropriate point probability, and equate the result to the probability of state j, b_j (similar to the unlimited-queue equations). The values of j used are h+1 and h-1,

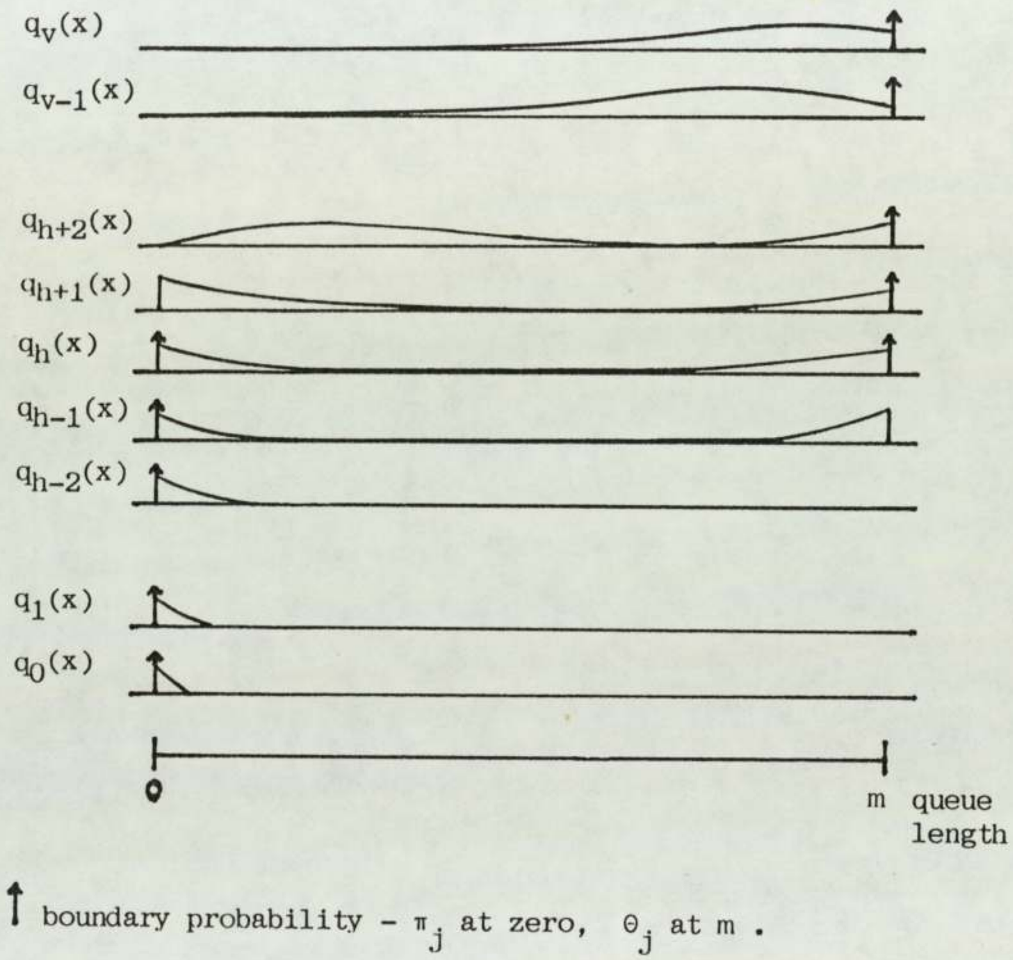


Fig. 2.3 Rough sketches of queue distributions for a limited queue with packet loss

whose integrals are given by

$$\int_0^m q_{(h+1)}(x)dx = \int_0^m \sum_{K=1}^{v} a_K c_{(h+1)K} e^{\lambda_K x} dx = \sum_{K=1}^{v} a_K c_{(h+1)K} \frac{(e^{\lambda_K m} - 1)}{\lambda_K}$$

and

$$\int_0^m q_{(h-1)}(x)dx = \int_0^m \sum_{K=1}^{v} a_K c_{(h-1)K} e^{\lambda_K x} dx = \sum_{K=1}^{v} a_K c_{(h-1)K} \frac{(e^{\lambda_K m} - 1)}{\lambda_K}$$

Thus, the remaining boundary equations are

$$\sum_{K=1}^{v} a_K c_{(h+1)K} \frac{(e^{\lambda_K m} - 1)}{\lambda_K} + \Theta_{h+1} = b_{h+1} \tag{2.9}$$

$$\sum_{K=1}^{v} a_K c_{(h-1)K} \frac{(e^{\lambda_K m} - 1)}{\lambda_K} + \pi_{h-1} = b_{h-1} \tag{2.10}$$

However, Θ_{h+1} and π_{h-1} are unknown. Fortunately, relationships exist between Θ_{h+1} and $q_{h-1}(m)$, and π_{h-1} and $q_{h+1}(0)$. Thus, Θ_{h+1} and π_{h-1} can be replaced by functions of $q_{h-1}(m)$ and $q_{h+1}(0)$ for which there are expressions.

To find these relationships, we must consider once again a queue with very small packet length (and therefore frame length) Δt. If $Q_{h+1}(1)$ is the probability of h+1 talkers and one packet in the queue at the end of a frame, it is easy to show that the queue probabilities at the zero boundary are

$$\pi_{h+1} = 0$$

$$Q_{h+1}(1) = \pi_h p(h,h+1)\Delta t$$

$$\pi_h = \pi_{h-1}p(h-1,h)\Delta t + \pi_h(1-p^*(h)\Delta t)$$

The last equation can also be written

$$\pi_h = \pi_{h-1}p(h-1,h)/p^*(h) \qquad (2.11)$$

and so

$$Q_{h+1}(1) = \pi_{h-1}p(h,h+1)p(h-1,h)\Delta t/p^*(h) \qquad (2.12)$$

Whilst $Q_{h+1}(1)$ is the probability that there is one packet in the queue with h+1 talkers, it is also the probability of a queue length between 0 and Δt with h+1 talkers. We can therefore write

$$Q_{h+1}(1) = \int_0^{\Delta t} q_{h+1}(x)dx$$

For very small Δt the integral can be replaced by a first-order approximation

$$\int_0^{\Delta t} q_{h+1}(x)dx = (q_{h+1}(0) + q_{h+1}(\Delta t))\Delta t/2 + o(\Delta t^2)$$

Substituting for $Q_{h+1}(\Delta t)$ in equation (2.12) and dividing by Δt gives

$$(q_{h+1}(0)+q_{h+1}(\Delta t))/2 = \pi_{h-1}p(h,h+1)p(h-1,h)/p^*(h)+o(\Delta t)$$

Finally, letting $\Delta t \to 0$ gives the required relationship

$$q_{h+1}(0) = \pi_{h-1} p(h,h-1) p(h-1,h)/p^*(h)$$

or

$$\pi_{h-1} = \frac{q_{h+1}(0) p^*(h)}{p(h-1,h) p(h,h+1)}$$

A very similar process yields the other relationship

$$\Theta_{h+1} = \frac{q_{h-1}(m) p^*(h)}{p(h+1,h) p(h,h-1)}$$

Using these expressions to substitute for π_{h-1} and Θ_{h+1} in equations (2.9) and (2.10) the final two boundary equations are complete. These can be solved using standard numerical methods to yield the a_K.

The non-zero π_j and Θ_j can be found by subtracting $\int_0^m q_j(x)$ from b_j as in the unlimited queue case, except for π_h and Θ_h which must be found directly from π_{h-1} and Θ_{h+1} (using equation (2.11) and its equivalent for Θ_h). The fraction of packets lost is then given by

$$\text{Loss} = \frac{1}{j_{ave}} \sum_{j=h+1}^{v} (j-h)\Theta_j$$

The average packet delay, which is used in Section 4.5, is given by

$$\text{Ave Delay} = \int_0^m \sum_{j=0}^{v} \frac{j}{j_{ave}} t q_j(x) dx + \frac{m}{h} \sum_{j=h}^{v} \frac{j}{j_{ave}} \Theta_j$$

Substituting ht = x gives

$$\text{Ave Delay} = \int_0^{m/h} \sum_{j=0}^{v} \frac{j}{j_{ave}} htq_j(ht)dt + \frac{m}{h} \sum_{j=h}^{v} \frac{j}{j_{ave}} \Theta_j$$

$$= \int_0^{m/h} \sum_{j=0}^{v} \frac{j}{j_{ave}} \sum_{K=1}^{v} ht\, a_K c_{jK} e^{\lambda_K ht} dt + \frac{m}{h} \sum_{j=h}^{v} \frac{j}{j_{ave}} \Theta_j$$

Integrating by parts and rearranging gives

$$\text{Ave Delay} =$$

$$\frac{1}{h} \left\{ \sum_{j=0}^{v} \frac{j}{j_{ave}} \left[m\Theta_j + \sum_{j=0}^{v} \left(\frac{a_K c_{jK}}{\lambda_K} e^{\lambda_K m} (m-1/\lambda_K) + a_K c_{jK}/\lambda_K^2 \right) \right] \right\}$$

(2.13)

2.6 BOUNDARY EQUATIONS FOR LIMITED QUEUES WITH EMBEDDED ENCODING

If embedded encoding were used it would be quite reasonable to shorten packets only when the queue was on its limit. In this case, the results from the previous section can be applied directly and the probability that a packet will be shortened to h/j of its original length is

$$\frac{j}{j_{ave}} \Theta_j$$

As stated in Section 1.3.2, it may be desirable to reduce the probability of having to remove a large proportion of a packet by starting to shorten packets before the queue reaches its limit, although more bits will be lost this way. The derivation of the boundary equations for such a system is rather complicated and as it is of rather limited use, only an outline will be given here.

Reducing each packet to 1/x of its original length is equivalent to increasing the channel capacity by x. Assuming the decision to reduce packet size is based on queue length, the queue distributions are roughly as shown in Fig. 2.4. When the queue reaches length m', the packets are shortened and the capacity effectively changes from h' to h. The point probabilities at the changeover point represent the cases where reducing all the packets would cause the queue to decrease again and so only enough are reduced to ensure the queue does not increase.

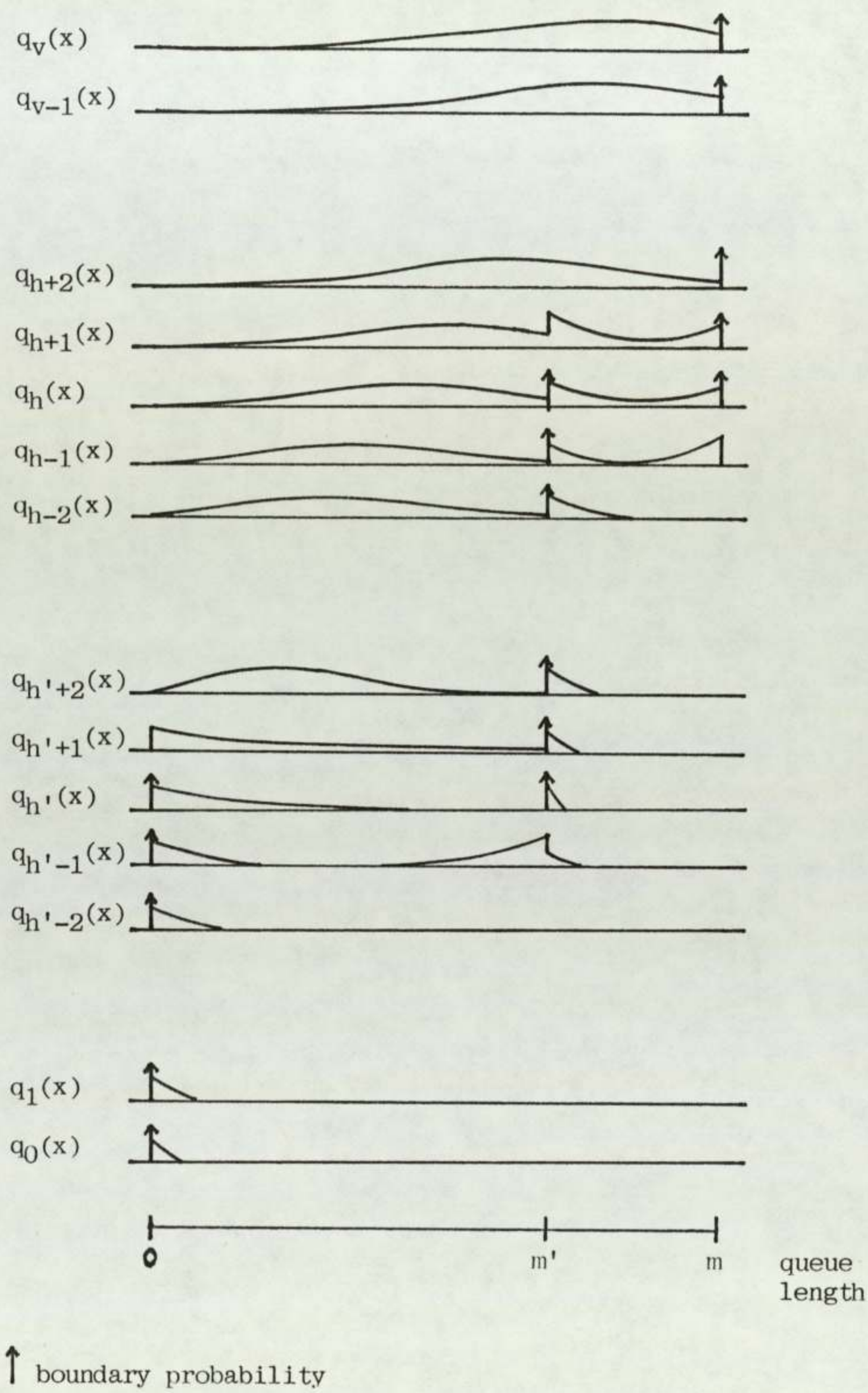


Fig 2.4 Rough sketches of queue distributions for a limited queue with embedded encoding

Before the boundary equations can be formed, the eigenvalues and eigenvectors for channel capacities of both h' and h must be found. The 2v boundary equations are then formed as follows. For $j \geq h'+2$ the zero boundaries must be zero and for $j \leq h-2$ the limit boundaries must be zero. For $j \geq h+2$ and $j \leq h'-2$ the distributions must be equal at the changeover queue length m'. This gives 2v-4 equations. Two more can be found by integrating $q_{h'-1}(x)$ and $q_{h+1}(x)$ and using the relationships derived in the previous section. The final two can be found by deriving the relationships between the steps at $q_{h+1}(m')$ and $q_{h'-1}(m')$ and the point probabilities at $q_{h-1}(m')$ and $q_{h'+1}(m')$, and then integrating $q_{h-1}(x)$ and $q_{h'+1}(x)$.

If there are not one but n changeover points, the procedure is the same; however, there are n+1 sets of eigenvalues and eigenvectors and (n+1)v boundary equations to find and solve. With v=20, queues with up to 3 changeover points have been successfully analysed.

2.7 COMPUTATIONAL ASPECTS

2.7.1 The Singularity of the State Matrix

The fact that R in equation (2.3) is a singular matrix can easily be seen by multiplying each row by its denominator, leaving the transition rates unweighted. The sum of each column is then

$$p(i,i-1) + p(i,i+1) - p^*(i)$$

which is zero by definition (see definition of $p^*(i)$ in Section 2.2). It follows that each row is a linear combination of all the other rows, and so the matrix is singular. It can easily be shown that the matrix adjusted for integer values of h is also singular.

The significance of this singularity is that one eigenvalue will be zero. The NAG subroutine (F02AQF[79]) used to calculate the eigenvalues does not return the value of zero, but typically a value in the range 10^{-5} to 10^{-12}. If this value were allowed to remain, it would cause gross inaccuracies in the calculation of $(e^{\lambda_K m}-1)/\lambda_K$ in equations (2.9) and (2.10). Two measures are therefore needed. Firstly, this very small eigenvalue must be detected and set to zero. Secondly, the formula $(e^{\lambda_K m}-1)/\lambda_K$ must be replaced by m, which is its true value when $\lambda_K=0$. (This can easily be shown by expanding $e^{\lambda_K m}$ with a power series.)

2.7.2 Improbable States

Including very improbable values of j (i.e. close to 0 and v for v>20) in equation (2.2) does not improve accuracy very much, and is inadvisable for two important reasons. The first is that for v>30 the elements of each eigenvector close to j=0 and j=v are smaller than 10^{-36}. The effect of this is that the eigenvector subroutine cannot reach convergence and may overflow as well. The second is that the computation required for the eigenvector calculation goes up with the cube of the number of states, and so including any more states than is necessary is a waste of computer time.

A short test showed that packet losses can still be computed to four decimal places of accuracy if all states of probability less than 10^{-5} are excluded from R. Using this limit, packet losses for values of v as large as 800 have been obtained.

For very large v, the calculation of the b_j to find at what point the 10^{-5} probability is reached is both time consuming and prone to overflow. Consequently, it is better to adjust the speech activity model itself to have a lower and upper limit and to find the limits by an iterative procedure. It was found that the lower limit j_ℓ and the upper limit j_u always lay closely within the bounds

$$j_u \leq \mathrm{MIN}(v,(.268+.327\lambda/\mu)v+15)$$

$$j_\ell \geq \mathrm{MAX}(0,.45(v-40)\lambda/\mu)$$

and so these were used as starting points for the iterations. The b_j are no longer given by a binomial distribution but by the formula

$$b_j = \frac{\binom{v}{j}\left(\frac{\lambda}{\mu}\right)^j}{\sum_{K=j_\ell}^{j_u} \binom{v}{K}\left(\frac{\lambda}{\mu}\right)^K}$$

2.7.3 Computation of Large Exponentials

As v increases, the eigenvalues of R increase and the value of m for a given delay increases. Thus, even for values of v as small as 10, the exponential function $e^{\lambda_K m}$ (Equations (2.8) to (2.10)) can overflow for large delay limits. For large values of v, delay limits of a few ms will cause overflow. This problem can be solved only because $e^{\lambda_K m}$ always occurs as part of the product $\underline{c}_K e^{\lambda_K m}$, and the magnitude of $\underline{c}_k$ is unimportant.

This leads to the following solution. For negative eigenvalues, $\underline{c}_K e^{\lambda_K m}$ will naturally underflow and be set to zero without causing any further difficulties. For positive eigenvalues, the product $\underline{c}_K e^{\lambda_K m}$ is given the values of $\underline{c}_K$ and $\underline{c}_K$, when it occurs alone, is scaled by $e^{-\lambda_K m}$. This turns the potential overflow into an underflow by effectively scaling $\underline{c}_K$ by $e^{-\lambda_K m}$.

One further complication is that one of the negative eigenvalues goes positive when the link utilisation goes above 100% (ignoring the reduction in utilisation caused by packet losses). For the special case when the link utilisation is exactly 100% the eigenvalue is zero, and in this case the two smallest eigenvalues must be detected and set to zero.

2.7.4 Program Structure

Figure 2.5 gives a flowchart of the basic program structure for the limited queue case. A listing of this program is given in Appendix 2. The subroutines are all from the NAG library[79]. It can be seen that once the eigenvalues and eigenvectors have been computed, a whole range of delay limits are used. Normally these go from 0 to $1\mu^{-1}$ either in 20 constant steps of $0.05\mu^{-1}$ or in a similar number of increasing steps starting at $0.018\mu^{-1}$. The computation of the eigenvalues and eigenvectors requires about twice as much CPU time as the solving of each set of linear equations. Thus, the computation time required for the eigenvalues and eigenvectors is only about one tenth of the overall computation time and is therefore not very significant.

The basic structure was used either just once through, or repeated a number of times with constant v and decreasing h, or constant h and increasing v. With constant h, the value of v was simply incremented for each time through the loop. With constant v, the value of h was selected to reduce the packet loss by about 20%. This ensured that accurate constant packet-loss curves could be produced by interpolating between successive values of h. Note also that the first box in Fig. 2.5 does not need to be repeated for each value of h.

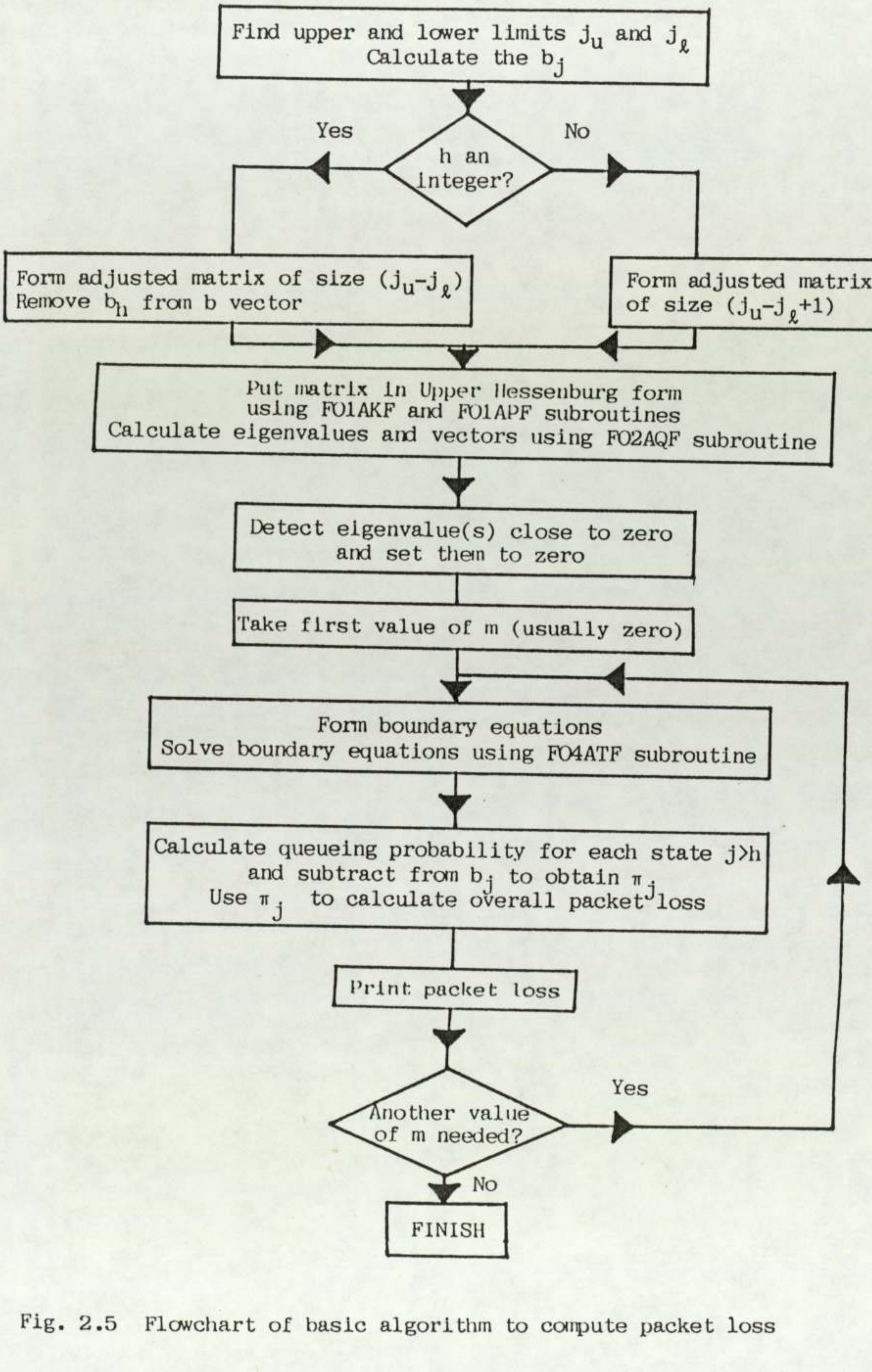


Fig. 2.5 Flowchart of basic algorithm to compute packet loss

2.8 SIMULATION TO CHECK ACCURACY OF THE ANALYSIS

2.8.1 Introduction

The models on which the analysis is based are not necessarily accurate for all parameters. For instance, it is possible that the analysis is inaccurate for small v, as Weinstein (57) found the talker model to be good only for $v>25$. The continuous queue model is likely to be more accurate for short packets than long packets. This will be particularly true for small delay limits as only the number of packets arriving in each frame is modelled and not the arrival process itself.

The accuracy has been checked using 8 independent simulations for each of a varied set of parameters. To make the simulations as realistic as possible, talkspurt and silence lengths were recorded from 10 half-hour conversations. The speech detector used was almost identical to that used by Brady (55). The speech on each channel was full-wave rectified and checked every 8 ms to see if it had crossed a threshold. If it had, it was assumed to be speech unless both the packet before and after recorded silence, in which case it was assumed to be noise and was classed as silence. Finally, any silence lasting less than 200 msec was assumed to be a stop consonant or very-low-level speech and so classed as speech. The measured values of μ^{-1} and λ^{-1} were 0.96s and 1.69s. Comments on the significance of the speech detector are made in Section 2.8.5.

The simulation programs used a packet-arrival strategy similar to the operation of a time-division multiplexer (this was also used by Tanaka et al. in their simulations [54] and by Minoli in his model [52]). That is, each frame is divided into v evenly-spaced time slots, and packets from the nth channel arrive only in the nth time slot. The 8 independent simulations were obtained by shifting the data of each speaker by amounts between 0 and 30 minutes. This ensured that the sequence of talkspurt and silences was not lost, but that the input to the multiplexer was completely different for each run.

The 90% confidence limits were formed using the student's t-distribution with 7 degrees of freedom. If the 8 packet losses are given by $X_1,\ldots,X_8$ and the mean is $\bar{X}$, the confidence limits are given by

$$\bar{X} \pm 0.253 \sqrt{\left(\sum_{i=1}^{8} (X_i-\bar{X})^2\right)}$$

2.8.2 Accuracy for small values of v

Because of Weinstein's simulations that suggested the talker model was poor for v<10 , the simulations were first of all run for v=5, 10 and 20. The values of maximum delay used were 10, 50, 100, 200, 400 and 800 ms.

Figs. 2.6 to 2.8 show the results of these simulations. Packet loss is used rather than the delay p.d.f. not only because numbers

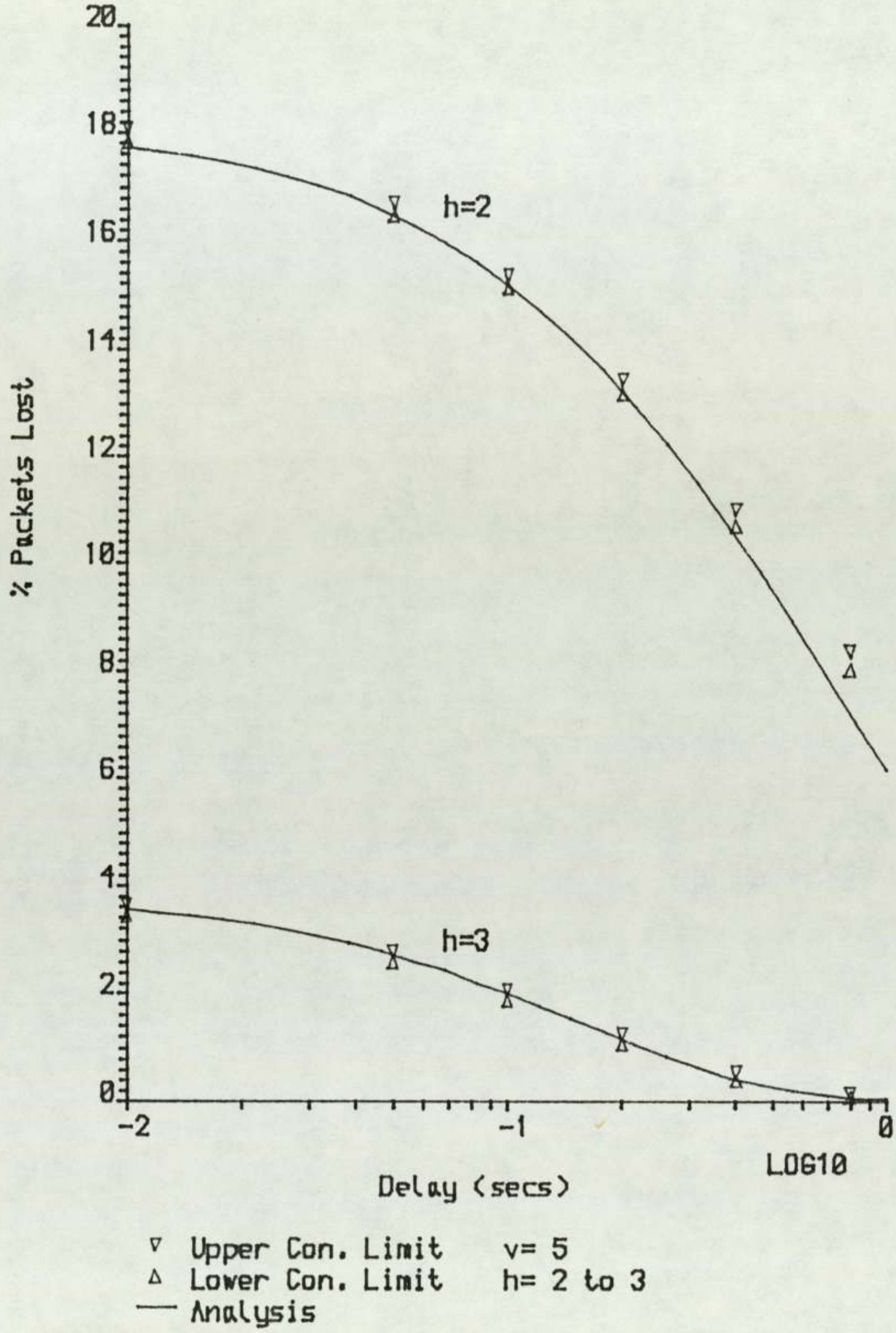


Fig. 2.6 Comparison of analysis and simulation results for 5 callers

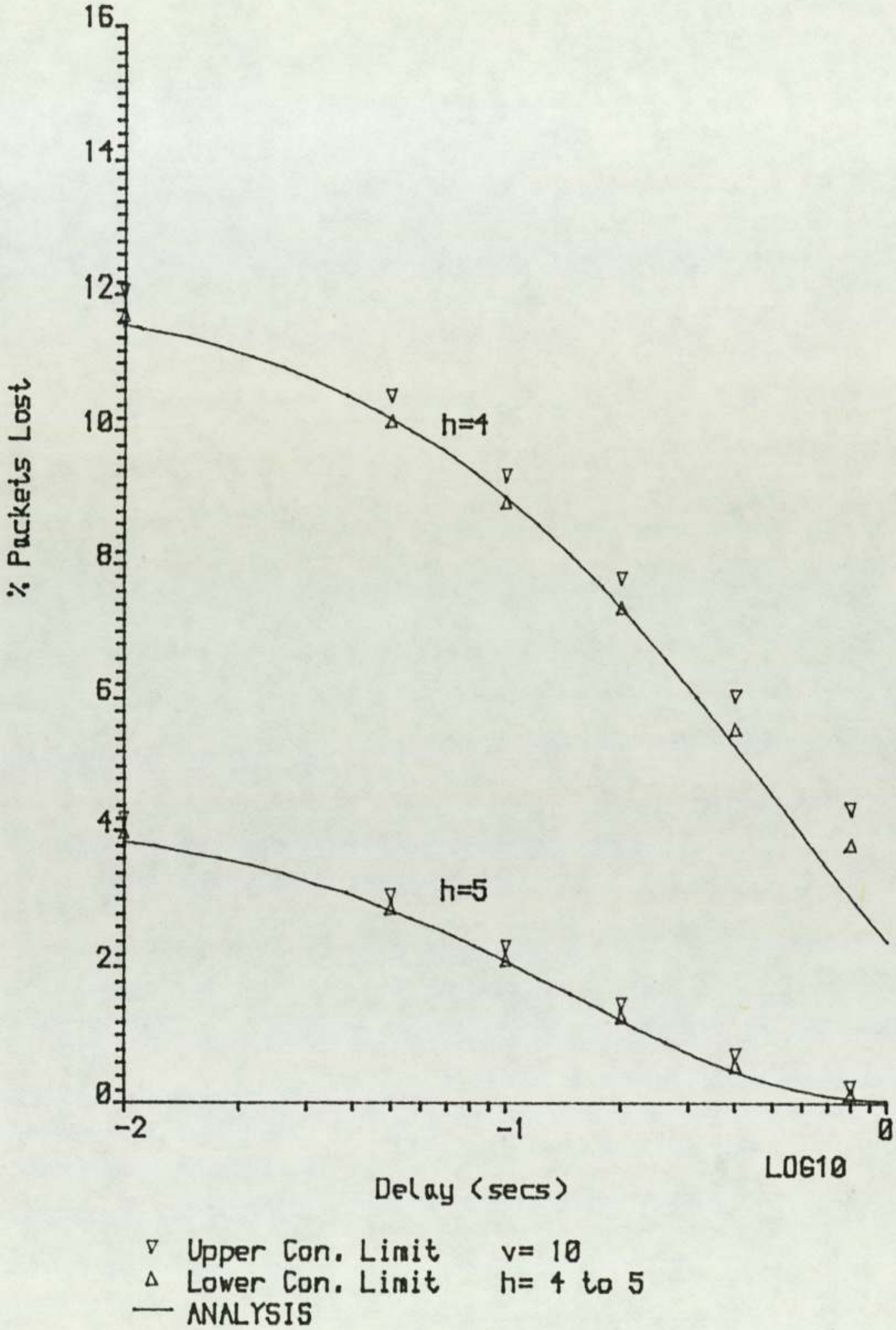


Fig. 2.7 Comparison of analysis and simulation results for 10 callers

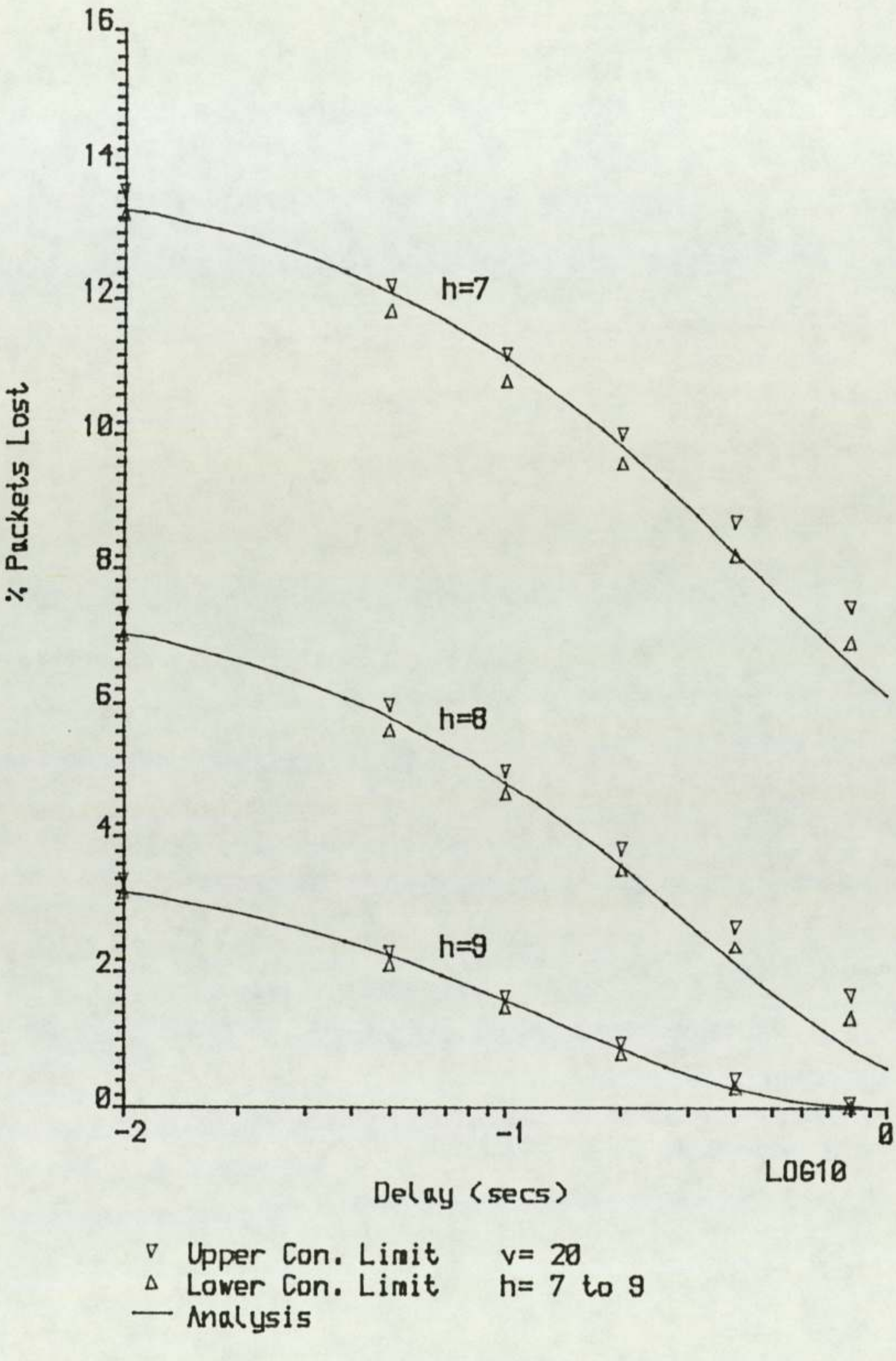


Fig. 2.8 Comparison of analysis and simulation results for 20 callers

are easier to compare than distributions, but also because packet loss has more significance than delay p.d.f.

The analysis falls within the confidence limits for delays up to 200 ms, but then begins to underpredict the loss. At 800 ms it is too low by as much as 0.8% loss in the worst case (v=10, h=4). However, there is no noticable loss of accuracy as v decreases, which suggests that the analysis is reasonably insensitive to the talker model.

2.8.3 Accuracy for large values of v

To test the analysis for large v, 30 minutes of operation of a link with 100 calls were simulated. As it was impractical to collect data from another 40 conversations, the data for these simulations were synthesized from the talkspurt and silence distributions of the ten half-hour conversations.

Fig. 2.9 shows the simulation results for h=36. Although the results do seem to be closer to the analysis than for smaller v, this may be due to the use of synthesized rather than real data. In any case, the results are no worse than for smaller v, demonstrating that the accuracy of the analysis is good for both large and small numbers of talkers.

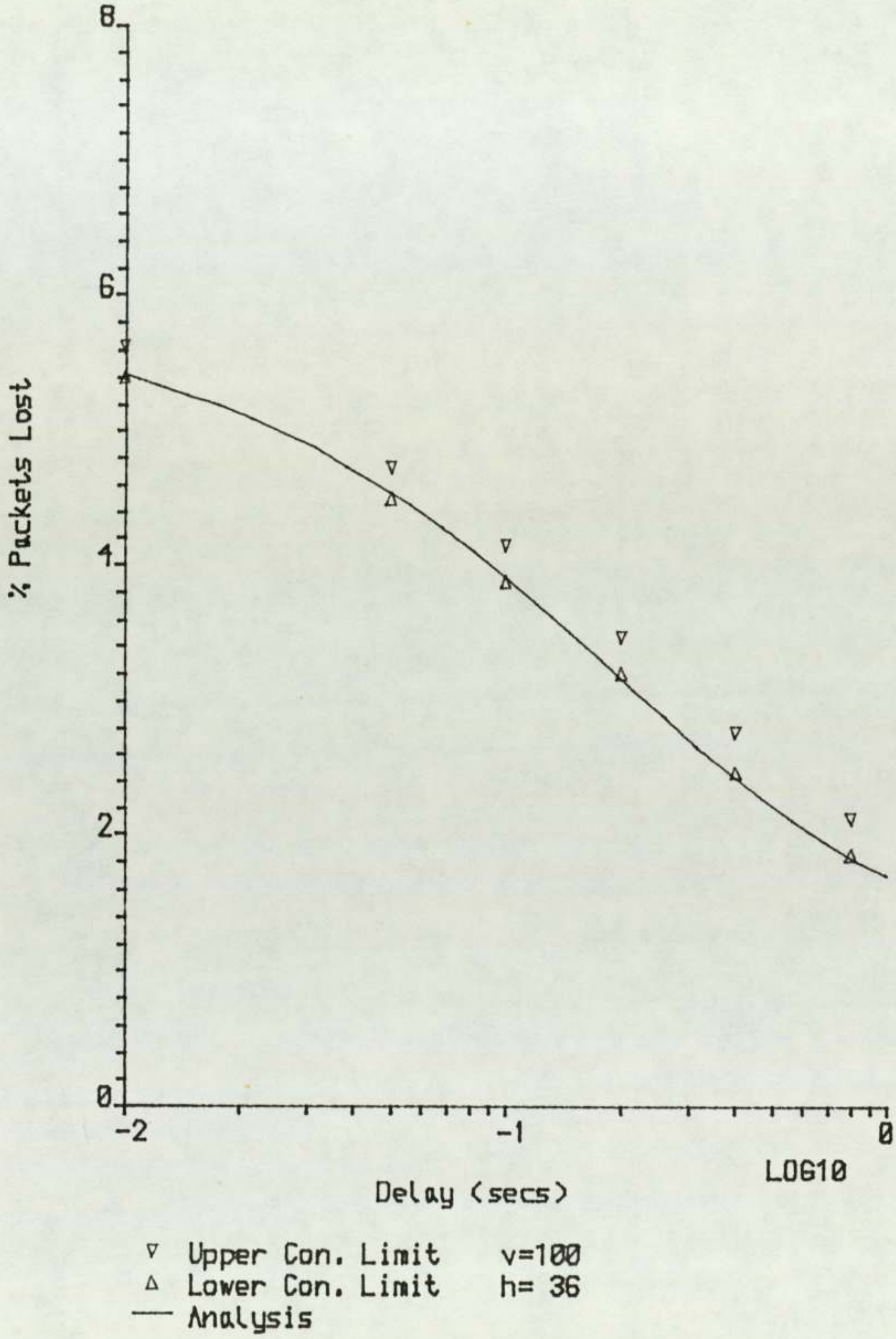


Fig. 2.9 Comparison of analysis and simulation results for 100 callers

2.8.4 Different Packet Lengths

The simulations for v=100, h=36 were carried out with packet lengths ranging from 2 ms to 128 ms. Fig. 2.10 shows the results of these simulations. It appears that the analysis is reasonably accurate for packet lengths where the maximum delay is more than about half the packet length.

The loss of accuracy for large packets and small delay limits is, as mentioned earlier, due to the lack of a model for when packets arrive in each frame. However, even if this could be modelled, the accuracy of the analysis would depend on the model being the correct one. As the packet arrival process is dependent on the packet switch itself, it is probably better that it is not included in the analysis.

The result that packet length does not normally affect packet loss is not surprising . As mentioned in 1.2.6, the result from one-dimensional models that delay, and therefore packet loss, is a function of packet length is highly suspect. It is evident from Fig. 2.10 that this is indeed not the case (except when the maximum delay is small relative to the packet length), and this emphasizes the inadequacies of the one-dimensional model.

Different packet lengths were not tried for small values of v, and there seems no reason to expect the effect to be different. However, the accuracy of the small-v simulations (which used 16 ms packets) at 10 ms suggests that similar results would be obtained.

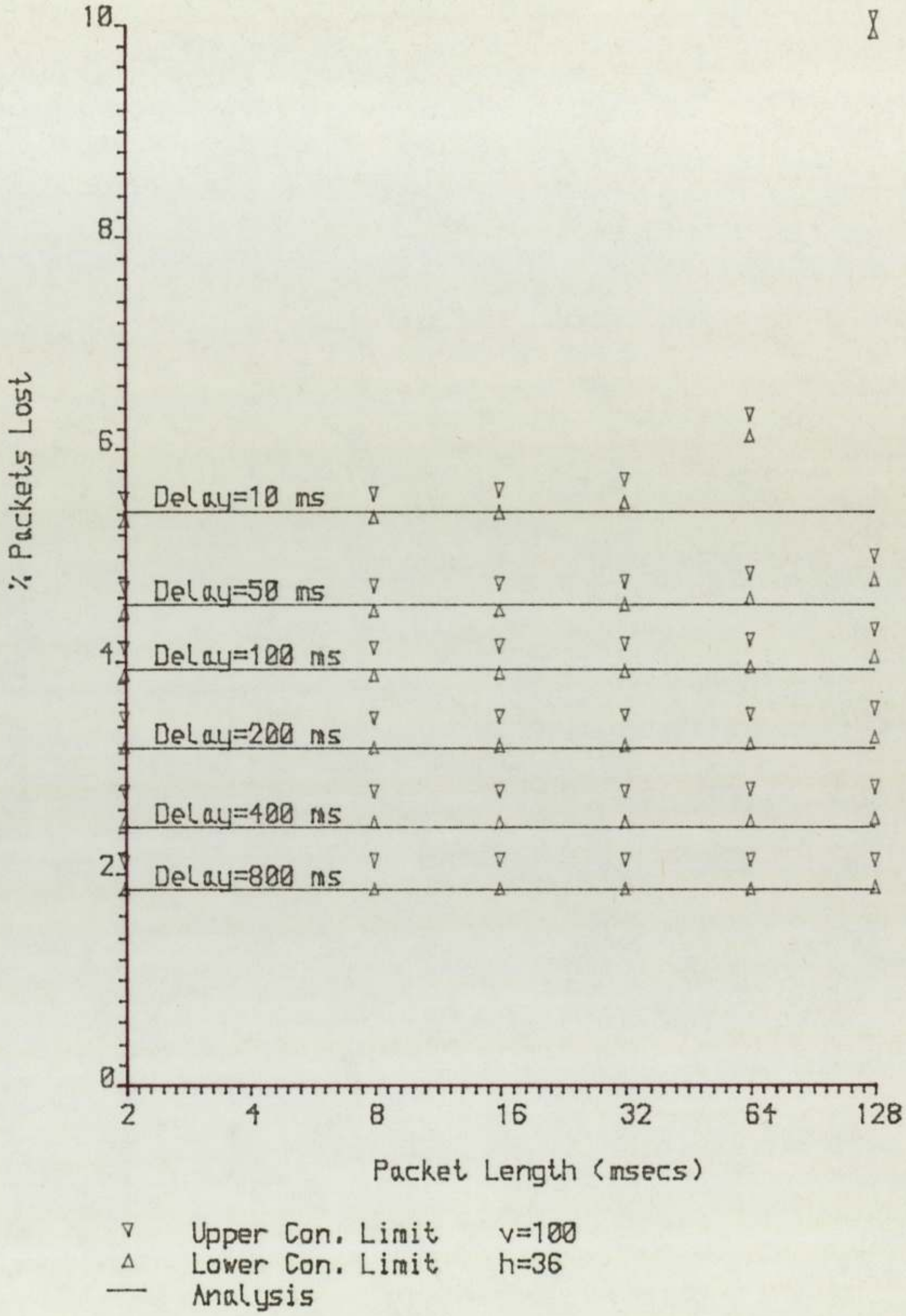


Fig. 2.10 Comparison of analysis and simulation results for different packet lengths

2.8.5 Effect of speech detector

The success of the analysis for all values of v must indicate that the talkspurt and silence distributions of the ten conversations were close enough to the exponential distribution to make the accuracy of the analysis independent of v. However, a recent paper (66) has shown that as the amount of hangover is reduced from 200 ms to 32 ms, the talkspurt and silence distributions become very unlike the exponential distribution. The reason for this is quite obvious; the short silences representing stop consonants that were previously classed as speech are now classed as silence. Thus the low end of the silence distribution takes a greater probability, and the tail end of the talkspurt distribution is almost lost altogether.

With a hangover of 32 ms the average talkspurt length was found to be 0.46 sec, compared to 1.01 sec for a hangover of 200 ms (66), the value used in the simulations. If this new value was used in the analysis, it would show that the multiplexing was at least twice as good than with 200 ms hangover since, as stated in section 2.2, delay is specified as a multiple of average talkspurt length. In reality the improvement can only be slight, as opening small gaps in the speech cannot affect the multiplexing very much.

In view of this, it is safest to use a value of μ^{-1} of at least 0.9 sec., regardless of the actual value of μ^{-1}. However, the load on the link should be given as accurately as possible. As this is dependent on the ratio λ/μ, the average silence length λ^{-1} must also be adjusted to give the correct value of λ/μ.

2.8.6 Summary of simulation results

To summarize these results, the analysis is accurate except for the following cases:

(a) when the delay limit is greater than 400 ms, the analysis begins to underestimate packet loss. It is accurate enough to be useful until at least 800 ms.

(b) when the delay limit is less than half the packet length, the analysis underestimates the packet loss. It is accurate enough to be useful (for packet lengths up to 128 ms at least) for delay limits down to one third of the packet length.

(c) If very little hangover (i.e. $\ll$ 200 ms) is used when measuring the average talkspurt length (μ^{-1}), the analysis can considerably underestimate the packet loss. More accurate results can be obtained if a value of about 0.9 ms is used for μ^{-1}.

2.9 EMPIRICAL RELATIONSHIPS FOR DELAY AND PACKET LOSS

One disadvantage of the two-dimensional model is that no closed-form equations exist for delay and packet loss, and consequently it is difficult to understand how these change as various parameters are changed. Because of this, empirical formulas have been established that, although not always accurate, at least give a clear indication of the effect of the different parameters.

For the unlimited queue case, the p.d.f. of delay d(t) can be approximated by an exponential distribution of the form

$$d(t) \approx \frac{h\ (1 - \rho^2)}{(\rho - \alpha)} \exp\left(- \frac{h\ (1 - \rho^2)\ t}{(\rho - \alpha)} \right) \qquad t>0 \qquad (2.14)$$

where the activity, $\alpha = j_{ave}/v$ and efficiency $\rho = v\alpha/h$. This formula is usually accurate for values of $t \approx 1\mu^{-1}$, and most inaccurate for small values of t (at t = 0 there is a point probability which this formula ignores). It is clear that the channel capacity h directly affects the amount of delay for a given ρ , and that as $\rho \to 1$ the slope gets ever smaller. For values of ρ close to α , the slope is extremely large. Of course, for $\rho < \alpha$ there is no queuing at all. Note that because of the results of Section 2.8.4, packet length is not included as a parameter.

It is interesting to compare Equation (2.14) with Coviello's (21) distribution of

$$d(t) = s(1 - \rho) \exp(-s(1 - \rho)t) \qquad t > 0$$

where s is the service time of one packet. Although his expression is not very good for ρ close to 1 or ρ close to α, it is of the right form. Its major error is that it expresses delay as a function of service time, and since this is a function of both channel capacity and packet length, delay becomes a function of packet length.

For the limited queue case, the reduction in packet loss with delay is exponential for $t>0.05\ \mu^{-1}$ and $\rho<0.95$ or $\rho>1.05$. For $0.95<\rho<1.05$ (this interval is smaller for large v) the loss versus delay curves are not exponential. Consequently, to keep the empirical formula simple, this range of values of ρ will be excluded. If $\ell(t)$ is the fractional packet loss with a delay of $t\mu^{-1}$, we have empirically

$$\ell(t) = \beta \exp\left(\frac{-7h\alpha|1 - \rho|t}{\sqrt{(\rho - \alpha)}}\right) + \ell(\infty)$$

$$t>0.05\mu^{-1},\ \rho<0.95,\ \rho>1.05 \qquad (2.15)$$

where β is approximately $0.9(\ell(0)-\ell(\infty))$ since as $t\to 0$ the exponential fit tends to underestimate packet loss. $\ell(\infty)$ is zero if $\rho<1$, and is $(\rho-1)/\rho$ if $\rho>1$. $\ell(0)$ can easily be worked out from the formula[57]

$$\ell(0) = \frac{1}{v\alpha} \sum_{j=h+1}^{v} (j-h)b_j$$

It can be seen that the exponent of Equation (2.15) is almost symmetric about $\rho = 1$. It can easily be shown that $\ell(0)-\ell(\infty)$ is almost symmetric about $\rho=1$ as well, at least for $(\ell(0)-\ell(\infty))>0.01$. Thus the loss v delay curve for $\rho=1+x$ is similar to the loss v delay curve for $\rho=1-x$ offset by $x/(1+x)$.

It might appear from equation (2.15) that the usefulness of delay increases as h increases since the exponent increases with h. This is not necessarily the case since although the percentage reduction in packet loss is greater for large h, the actual reduction may be much less due to a decrease in $\ell(0)$.

CHAPTER THREE

TESTS ON THE EFFECTS OF PACKET LOSSES

CHAPTER 3

TESTS ON THE EFFECTS OF PACKET LOSSES

3.1 INTRODUCTION

3.1.1 Overall Requirements

It was shown in Section 1.3.1 that the literature on the acceptability of packet losses is not at all in agreement over the various issues concerning packet loss. The best packet length ranges from 4 ms(59) to 16-32 ms(5). The best way of filling in gaps appears to be to repeat the previous packet(60,61), but there is no evidence to show this increases the intelligibility of the speech. It is possible that it causes a listener to misunderstand some words without realising it, which is worse than not being able to understand at all. The tests concerning how much loss is acceptable suggest amounts from 2%(60) to 10%(5,61), depending on the test.

There was a need for carefully conducted tests to resolve these issues, particularly the question of how much loss is acceptable. There was also a need to take into account how the losses take place, since the random losses used in all the tests are not realistic, as will be seen in the next section. The aim of the tests described in this chapter is to give meaning to the analysis by producing reliable assessments of the effects of packet losses as they would occur in a real queue.

The details of the tests given in this chapter are only those that are needed to gain an understanding of each test and the significance of its results. Details of the preparation of the tests, the subjects, the test structures, the results and the analysis of variance of the results are given in Appendix 1.

3.1.2 The Variability of Loss Levels

The values of packet loss produced by the analysis are only averages and, as such, need to be treated carefully. Even for an average packet loss of 2%, the actual packet loss may rise as high as 10% or fall to zero for short periods. This variability in packet loss is greatest when the capacity of a link is low, since it is then that the loading on the link varies the most (see Fig. 3.1). A long queue also tends to increase variability, since it removes short bursts of losses completely but lets through long bursts.

As an example of how losses are distributed, Figs 3.2 and 3.3 show the percentage of packets lost in each 320 ms interval of one speech channel in just under half an hour of queue operation. Although the average packet loss is virtually the same in each case, it can be seen that the longer queue used in Fig. 3.3 gives a different loss distribution. The bursts of losses are more spread out, but the degradation introduced by each burst is worse, demonstrated by the large number of bursts exceeding 25% loss.

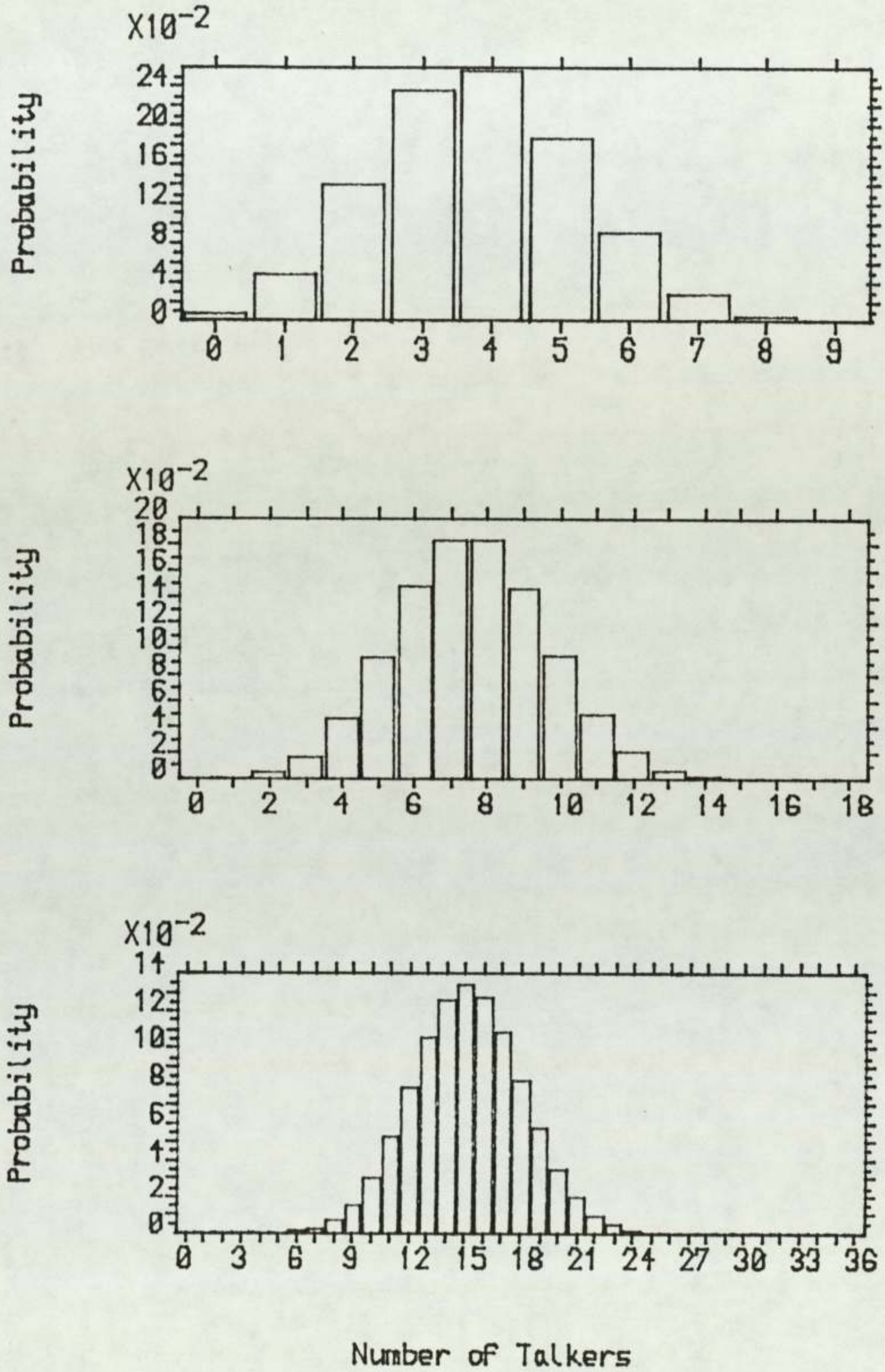
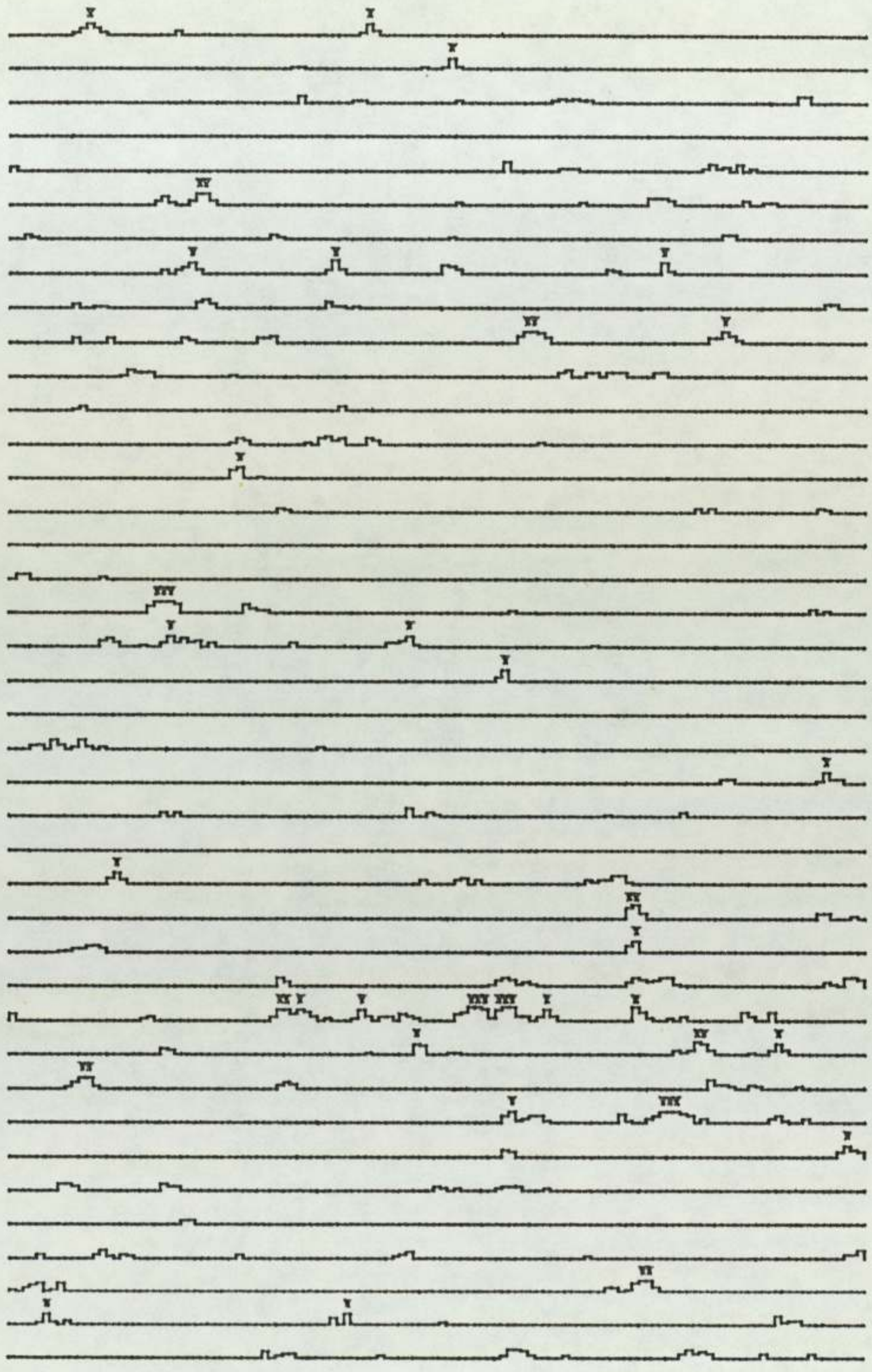
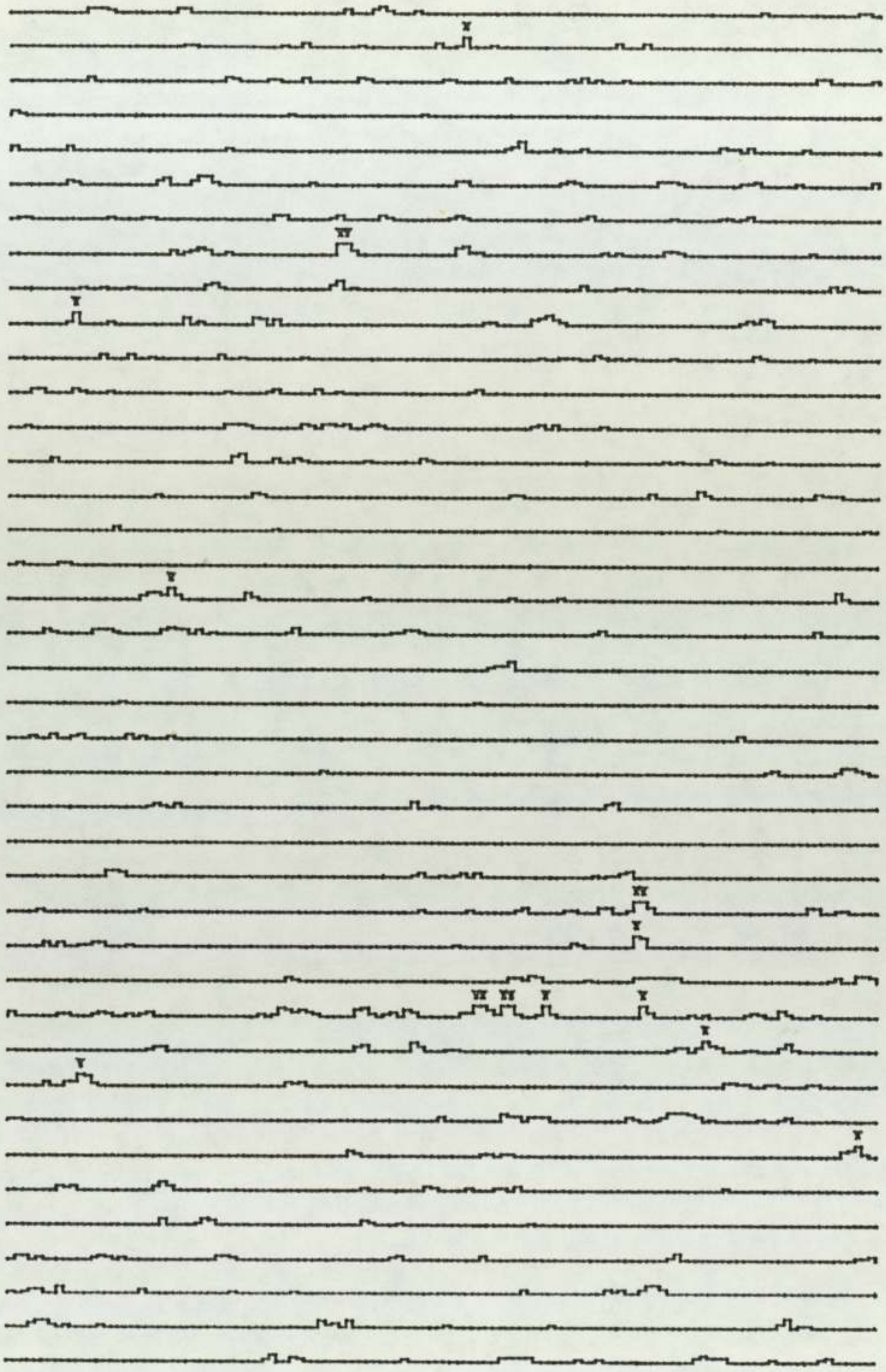


Fig. 3.1 Probability distribution of no. of talkers producing packets for 9, 18 and 36 callers, showing how the loading of a link varies



▾ more than 25% of packets lost. Average packet loss=3.29%
Pkt. length = 16ms, v = 20
Each line represents 40 seconds of one speech channel
Spacing between lines is equivalent to 100% of packets lost

Fig. 3.2 Percentage of packets lost in each 320ms interval of one speech channel in just under half an hour of queue operation
Channel capacity = 9, Max. delay = 10ms



▿ more than 25% of packets lost. Average packet loss=3.48%
Pkt. length = 16ms, v = 20
Each line represents 40 seconds of one speech channel
Spacing between lines is equivalent to 100% of packets lost

Fig. 3.3 Percentage of packets lost in each 320ms interval of one speech channel in just under half an hour of queue operation
Channel capacity = 8, Max. delay = 200ms

Although it is probably possible to extend the analysis to quantify these variations, it is doubtful that this would achieve anything. It would be a long and difficult task to test subjectively all the different amounts of variation as well as all the different amounts of packet loss. Then, having tested them, all the results would have to be combined somehow to give an overall quality score. Experience with early subjective tests on packet losses (and the published work) has shown that consistent results are so difficult to obtain that such a sophisticated approach is most unlikely to yield useful results.

If the variations themselves are not to be tested, a test must be used that is long enough to include, for each average packet loss, the variations that make up that average. The drawback with this approach is that a large test is needed to test just a few parameter values because of the amount of speech required. This restricts the number of different parameters values that can be tested.

3.1.3 Additional Factors to be Tested

From the literature it is clear that different packet lengths, fill-in techniques and levels of packet loss need to be tested. However, there are two underlying assumptions made in all work on packet speech (including Chapter 4 of this thesis) that also need to be tested.

The first is that different loss distributions, caused by either different queue limits or different channel capacities, do not affect opinion scores. In other words, it must be verified that packet loss alone really is a suitable parameter to use in assessing speech quality.

The second is that, if a packet-switched speech network is connected to the public network, the speech degradations of the public network do not interact with the packet loss of the packet-switched speech network. Although it would be useful to test packet loss with impulsive noise, crosstalk and other degradations, only loudness loss will be tested as it is probably the most common degradation.

3.1.4 Speech Coder Considerations

One other factor that affects the acceptability of packet losses is the type of speech coder. With predictive coders (e.g. delta modulation, ADPCM) a packet loss means a loss of step size and prediction information; so the speech following a loss will also be degraded(74,80). This effect can be avoided by transmitting the necessary information in each header (this also allows forward rather than backward prediction to be used(5)).

On the other hand, vocoders ought to be able to fill in for lost packets very well, as the previous packet's parameters can be continued without any discontinuity in the speech (at least for packets containing only one or two vocoder frames(58)). Similarly, since embedded encoders produce a gentle loss of quality rather than abrupt losses, the effect of speech loss should also be reduced when these coders are used.

It has not been possible to test packet losses with different coders because of a shortage of both time and equipment. However, by using the 12-bit, 8 k samples/sec linear A/D speech interface to the computer with no further speech processing, the results provide a lower bound for vocoders and embedded encoders, and an upper bound for predictive coders. This upper bound can be reached if packet headers contain step-size and prediction coefficients.

3.2 PRELIMINARY TEST

3.2.1 Aim

The aim of this test was to find the best packet length in the interval 4 to 32 ms, and the best fill-in technique (i.e. whether it is best to repeat the previous packet or to leave a gap when a packet is lost). Since it is the relative intelligibilities of fill-in techniques that is unknown, a test was used that could give intelligibility scores as well as opinion scores.

3.2.2 Test Material

The test was a listening test constructed from sentences of the form 'the black ship could hit the table', where the adjective, main verb and two nouns were randomly put together from lists of common words. The subjects had to repeat each sentence after they heard it, and the degradation was scored on how many of the four main words were repeated correctly (as well as the subject's opinion of the sentences).

Early versions of the test were found to give inconsistent results, and it was evident that most of the incorrect words were not caused by lost packets. This is probably why previous intelligibility studies had not managed to choose between the two fill-in techniques(60,63). Consequently, several measures were taken to ensure the sentences were of as near equal difficuty as was possible.

First of all, words of three or more syllables were excluded from the word list, as these tended to be distinctive and very easy to hear correctly. 120 sentences were made up from the resulting word lists; these were spoken both by a male and a female speaker, and tape-recorded.

Two people then listened to each sentence with no packet loss at all, and any sentence misheard by both people was rejected. In addition, another two people listened to each sentence with 50% packet loss, 32 ms packets and gaps left in place of lost packets. The sentences that were repeated with fewest mistakes were rejected as they were too distinctive. As only 54 out of the 120 sentences spoken by each voice were required for the test, this enabled two sets of sentences of fairly equal difficulty to be used.

In a further effort to gain consistent results, the subjects chosen were all females aged between 16 and 50 from a similar social background, who did not use the telephone for business and who knew the voices well. Since the test consists only of comparisons, it was assumed that these limitations would not affect the outcome of the test.

3.2.3 Structure of Test

The sentences were grouped into 6 groups of 18 sentences, the first 9 of each group being a male voice and the second 9 a female voice. Packet losses of levels 20, 30 and 40 per cent were then inserted into each of the four main words of each sentence. These losses either lasted for the whole word or for an interval of 160 ms at a random position in the word. This procedure ensured that each group of 18 sentences could have the same loss patterns, and so only the packet length and fill-in technique varied between groups. Three different packet lengths were used, 8 ms, 16 ms and 32 ms. 8 ms was used rather than 4 ms, as 4 ms seemed too short to be of practical use.

A Latin square design was not used for the test. Instead, sentence groups were arranged in pairs so that subjects could make an accurate judgement as to which of the pair was preferable. Whilst this increased the reliability of the comparisons, it decreased the sample size of the intelligibility scores, as not all of them could be used for each comparison.

Nine different pairs were tested; 8 v 16 ms, 16 v 32 ms and 8 v 32 ms lengths for both silent gaps and repeated-packet gaps, and silent v repeated packet gaps for 8, 16 and 32 ms packets. As each pair was tested both ways round to balance the test, there were 18 pairs in all. As any one person could only listen to 6 groups of sentences (i.e. 3 pairs of groups) to avoid hearing the same sentence twice, 6 people were required just to test each pair in each order once. In fact 12 people were used, so each pair in each order was tested twice (see Appendix 1 for further details of the test structure).

3.2.4 Method of Testing

The subjects were visited in their own homes and listened to the sentences through a telephone headset connected to a tape recorder. The speech level was not measured but simply adjusted to a comfortable level before the test began. No sidetone was used as it seemed unlikely that it would have any effect on a listening test (although in view of comments in reference (61) this may not be altogether true). As the subjects repeated the sentences, the number of wrong words was noted.

It was originally intended that the male voice and female voice sentences should be taken as a whole, but it very quickly became apparent that the subjects wished to compare them separately. Because of this, the tape recorder was stopped after each set of 9 sentences to allow the subjects to give a score on an Excellent-Good-Fair-Poor-Bad scale to that half group of sentences. Although this was merely to assist them in stating a preference between two slightly separated half-groups (since they were in the order male-female-male-female), the scores have proved to be useful for confirming the preferences.

3.2.5 Results

Table 3.1 lists the results of the tests. The first results column shows the number of times each group was preferred to the other one (counting male and female voices separately). The scores

TABLE 3.1 Results of Preliminary Test

Pkt Length	Fill-in	No. of Preferences	Opinion Score Confidence Level	Intelligibility Confidence Level
8 ms 8 ms	gap repeat	1 5	– 1%	– 2%
16 ms 16 ms	gap repeat	– 8	– 0.1%	– 1%
32 ms 32 ms	gap repeat	– 7	– 0.1%	– 0.1%
8 ms 16 ms	gap gap	5 2	20% –	25% –
16 ms 32 ms	gap gap	5 –	5% –	20% –
8 ms 32 ms	gap gap	5 –	5% –	20% –
8 ms 16 ms	repeat repeat	1 2	NS –	25% –
16 ms 32 ms	repeat repeat	3 3	NS –	2% –
8 ms 32 ms	repeat repeat	4 1	20% –	2% –

do not always add up to the maximum of 8, since some people could not tell any difference between the groups. The second results column gives the confidence level, based on opinion scores, that one group is better. The third results column gives the confidence levels based on the intelligibility scores. As usual, the smaller the confidence level, the more certain the result.

There can be little doubt that the repeated-packet strategy is the better of the two fill-in methods. Not only is it subjectively preferable, as reported in reference (61), but it is quite definitely more intelligible. As for packet length, it appears that 8 ms packets are better than 32 ms packets, and that 16 ms packets probably lie somewhere between the two.

In view of these results, the next two tests both use 8 ms packets with gaps filled by repeated packets.

3.3 LISTENING TEST

3.3.1 Aim

The aim of this test was two-fold. Firstly, to find the acceptability of different levels of packet loss using the standard listening effort scale[34]. Secondly, to verify that opinion scores are not significantly affected by different loss distributions.

3.3.2 Test Material

A tape was made of 120 sentences taken from novels and magazines, half spoken by a male and half by a female (as before). Gaps were left between sentences to allow time for them to be repeated. (The repetition this time was to aid concentration and not for intelligibility testing.) Thus, the whole tape lasted about 25 minutes.

The tape was played over six different simulated packet-switched speech links and the output recorded. The data for the simulations were the same as used for the simulations described in Section 2.8 . The receiver buffer in the simulations buffered all the packets out to the maximum delay. Thus the only effect of delay was to change the packet loss distributions. The links with their average packet losses are shown in Table 3.2. It can be seen that the first two pairs have the same packet loss for different queue lengths and channel capacities. The third pair do not have equal packet loss even though the maximum delay of the smaller channel is set much higher than could ever be tolerated. Even so, they were included in the test to give an idea of whether or not packet losses this high are acceptable.

3.3.3 Structure of Test

The tape recordings from each link were divided into 12 groups of ten sentences. These groups were then mixed together to form six tests using the male voice and six tests using the female voice, based on a standard Latin square design (see Appendix 1 for more details).

TABLE 3.2 Measured Packet Loss for Simulated Links with 21 Calls

Channel capacity	Maximum delay (ms)	Average Packet loss
9	8	3.5
8	256	3.5
8	8	7.6
7	768	7.4
7	8	14
6	2000	19.1

As has already been demonstrated, packet loss varies a great deal over 25 minutes, and each group of ten sentences had a quite different loss. Table 3.3 shows the range of losses for each link.

Besides this variability between groups, there was a variability within each group as packet losses tend to occur in bursts. This often caused the subjects great difficulty in assessing the ten sentences as a whole, and led to considerable inconsistencies in the results.

TABLE 3.3 Variation of Packet loss in Listening Test

Channel Capacity	Maximum Delay (ms)	Minimum Packet Loss (%)	Maximum Packet Loss (%)
9	8	1.2	5.0
8	256	0.8	6.2
8	8	3.3	10.1
7	768	3.2	12.3
7	8	7.9	17.8
6	2000	10.7	26.4

3.3.4 Method of Testing

Subjects were visited in their homes and listened to the test tapes through a telephone headset connected to a tape recorder. The subjects repeated each sentence, but mistakes were not recorded. Again, the volume level was adjusted to a comfortable level before the test began, and no sidetone was used. After each group of 10 sentences they were asked to score the sentences according to the listening effort scale given in Table 3.4. Since it was hard to get consistent results with this test, 36 subjects were used and so each of the 6 tests were repeated 6 times. This time, the subjects were of both sexes, and from a wide variety of professions (see Appendix 1 for more details).

TABLE 3.4 Listening Effort Scale

4	Complete relaxation possible, no effort required.
3	Attention necessary; no appreciable effort required.
2	Moderate effort required.
1	Considerable effort required.
0	No meaning understood with any feasible effort.

3.3.5 Results

Figure 3.4 summarises the results of the test. The opinion scores lie within a narrower range than expected. Some of the subjects commented that they never gave a 'no effort' score because

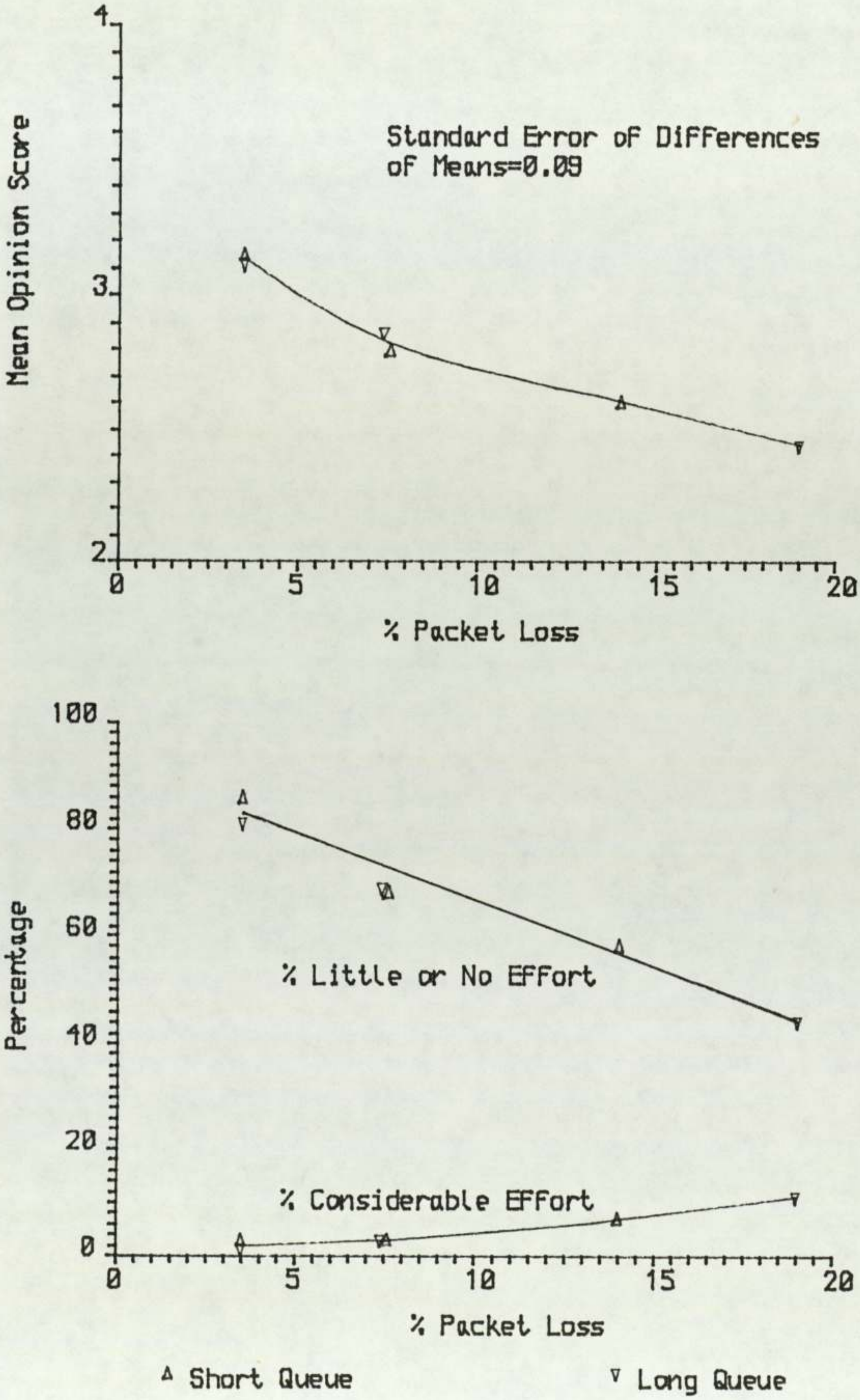


Fig. 3.4 Results of Listening Test

some attention was always necessary to be able to repeat the sentences. It also appears that the reason for very few 'considerable effort' scores is simply that, however annoying they are, packet losses do not seriously impair intelligibility. In retrospect, the listening effort scale was obviously not a very suitable one, even though it is a standard one to use for listening tests[34].

The conclusion that can be reached concerning packet losses is that speech with levels of loss up to 20% can generally be understood with moderate effort. However, it is still not clear whether loss levels this high are really acceptable or not.

The question of different loss distributions is not altogether settled. The mean scores for the links with similar packet losses are very close. Figure 3.5 shows histograms of the actual scores for the links with similar packet losses. (Where people have given a score halfway between two points on the opinion scale, half the scores have been rounded up and half rounded down.) It can be seen that these are also very close. However, the standard error indicates that this closeness may be a result of chance. The 90% confidence limits on the difference between the means are $\pm$ 0.149 (see Appendix 1). Thus there is a 10% probability that the means in actual fact differ by more than half the difference between the means for 8 channels, short queue, and 9 channels, short queue. However, given the difficulty of testing the different loss distributions, it does seem reasonable to assume that they do not significantly affect opinion scores.

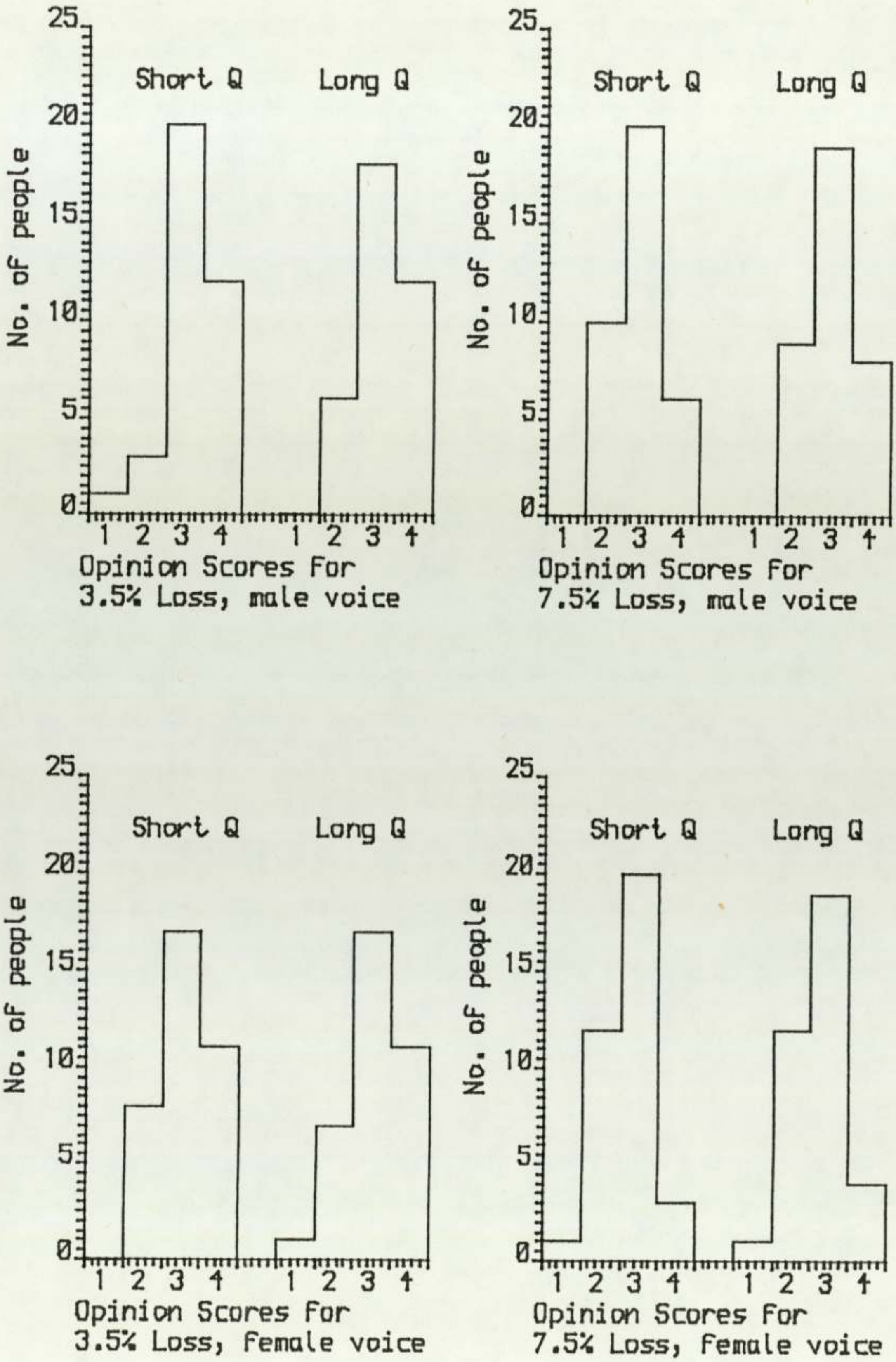


Fig. 3.5 Histograms of Listening Test results

The bursty nature of packet losses creates difficulties whatever test is used. It was clear in this one that subjects were very unsure how to score a group of sentences where some were clear, others were slightly degraded and perhaps one was badly degraded. Thus, the only way of obtaining very reliable results is to test the losses under the conditions they actually occur, i.e. in conversation.

3.4 CONVERSATION TEST

3.4.1 Aim

The aim of this test was to obtain reliable opinion scores for various levels of packet loss through the use of a conversation test. In addition to different levels of packet loss, two different settings of loudness loss were used. The first was approximately the ideal loss and the second the worst loss that should be encountered in a public network. The purpose of using the worst loss was to find out how packet loss and loudness loss interact. However, it also enabled the effect of packet loss to be compared with the effect of a better understood degradation.

3.4.2 Test Conditions

Telephone sets were placed in two rooms at opposite ends of the building. These were connected to each other through the speech interface of the computer which was programmed to simulate a packet-switched speech link, as before. Whilst the telephones were given the facility to ring each other, dialling and ringing tone were not provided.

The carbon microphones in the handsets were replaced by microphones which had the frequency characteristics but not the unreliability of a carbon microphone. The hybrid transformers in the telephone sets were by-passed to provide a four-wire link from one headset to the other. The sidetone path was created artificially and had 1.5 dB more loss than the main speech paths.

The two overall reference equivalent (ORE) values used were 14 dB ORE (8 dB overall loudness rating (OLR)) and 34 dB ORE (28 dB OLR). Injected room noise was not used, as each room already had a room noise of 50-60 dBA. However, a constant noise level at each earpiece of -77 dBmp was maintained by injecting noise just before each earpiece. This was just sufficient to mask quantisation noise and room noise coming through both sidetone and speech paths. It was not intended to simulate circuit noise, but to give a known constant noise level throughout the test. This gave average SNRs at each earpiece of approximately 54 dB and 34 dB for the two values of loudness loss.

3.4.3 Structure of Test

Four levels of average packet loss were tested: 0, 2.5%, 10% and 25%, each at the two values of loudness loss already mentioned. Thus each pair of subjects carried out eight conversations. The range of losses over all the conversations is shown in Table 3.5. Again, there is considerable variation, even though the period over which the losses are measured is quite a few minutes.

TABLE 3.5 Variation of Packet Loss in Conversation Test

Channel Capacity	Maximum Delay (ms)	Minimum Packet Loss (%)	Maximum Packet Loss (%)
21	10	0	0
11	10	1.3	4.1
9	10	6.9	14.6
7	10	19.3	31.8

As sixteen pairs of subjects were used, the settings were arranged in two different Latin squares (see Appendix 1 for details). Each test was divided into two halves of four conversations each to allow the subjects a break in the middle.

The packet-switched speech link was simulated using the data collected for the simulations described in Section 2.8. As the speech detector used for the tests was a simple level detector with 200 ms hangover, 200 ms hangover was also added to the data. The simulation was restarted only at the beginning of each group of four settings. In-between, it was halted while the subjects filled in their opinion forms and the link parameters were changed, and then started from where it was stopped. This enabled about 20 to 30 minutes of queue operation to be simulated.

3.4.4 Method of Testing

The test was a standard conversation test described in reference (34). The subjects were given a conversational task to perform, described in detail in Appendix 1. The subjects took it in turn to ring one another when given the signal to by a central control switch. If they talked for more than about eight minutes, the computer put pips on the line to warn them they were about to be cut off. This was necessary not only because the test might take too long otherwise, but also because only about 8 minutes worth of data could be stored in the computer memory at once. Each call was assessed by the subjects on the standard Excellent-Good-Fair-Poor-Bad scale. In addition, subjects were requested to state whether or not difficulty was experienced during the conversation, and if so, why.

3.4.5 Results

Figure 3.6 shows the Mean Opinion Score (MOS) and the % good or better, % poor or worse and % difficulty for the 14 dB ORE (ideal) connection. Presumably, those who said they had no difficulty either did not notice the distortion or at least thought it did not impede the conversation in any way.

Figure 3.7 shows the MOS and the % good or better, % poor or worse and % difficulty for the 34 dB ORE (worst case) connection. The % difficulty scores can be separated into "% who commented on distortion" and "% who commented on the loss or background hiss" as

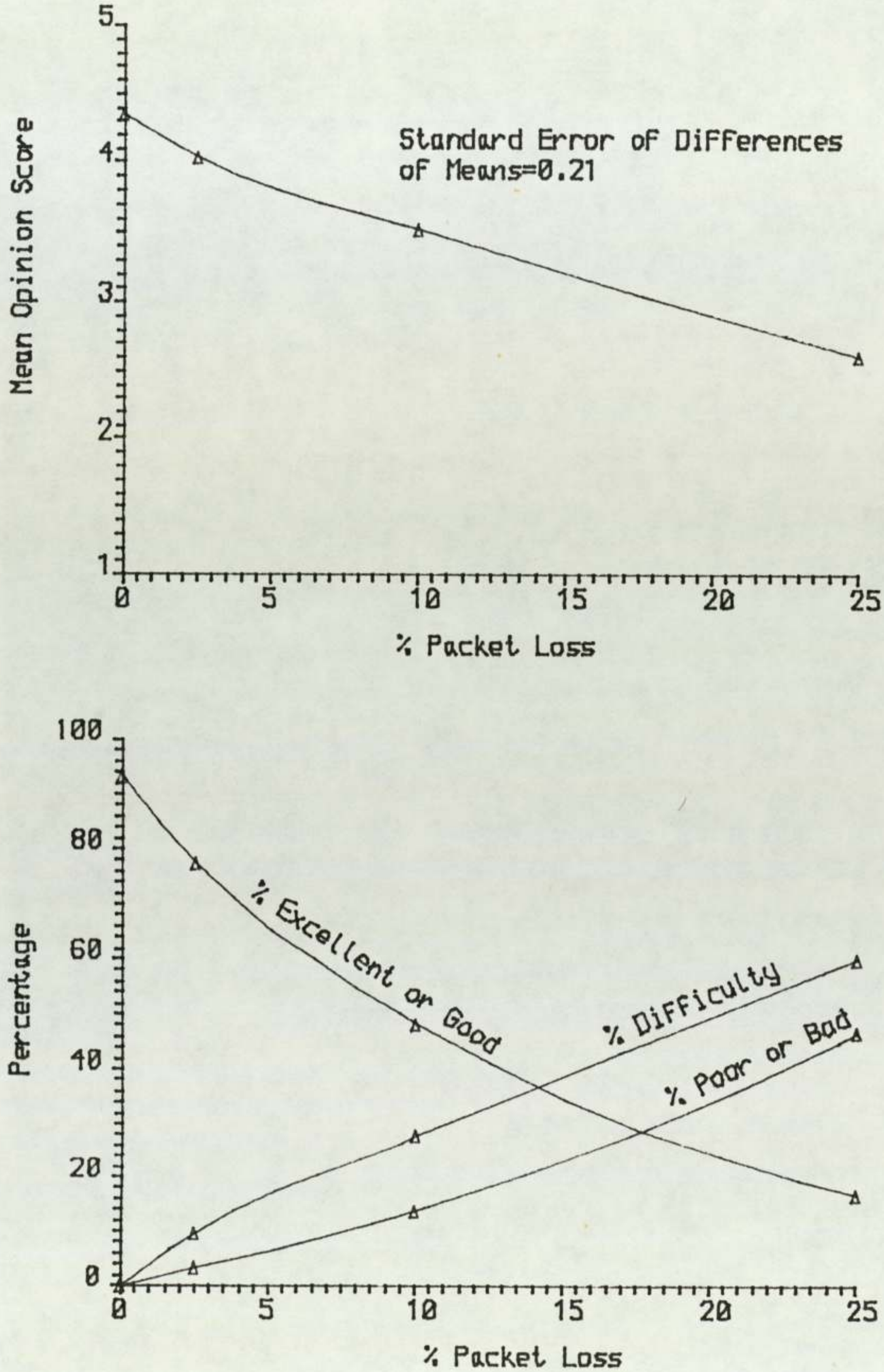
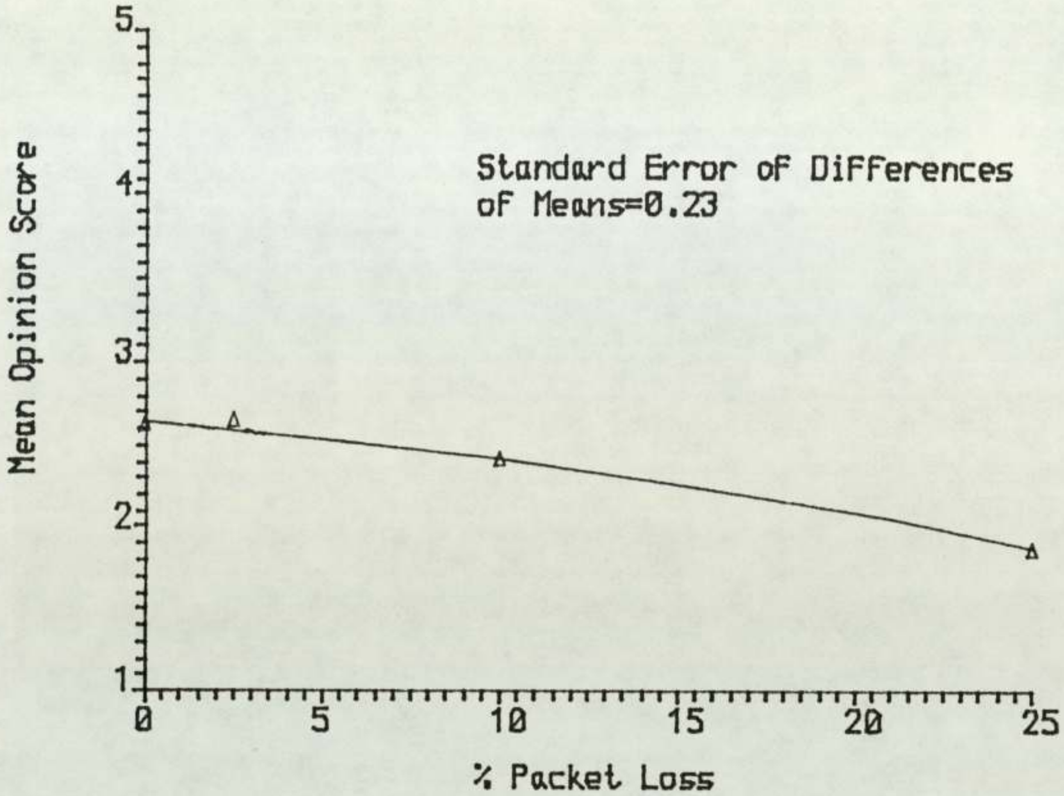


Fig. 3.6 Results of Conversation Test at 14dB ORE



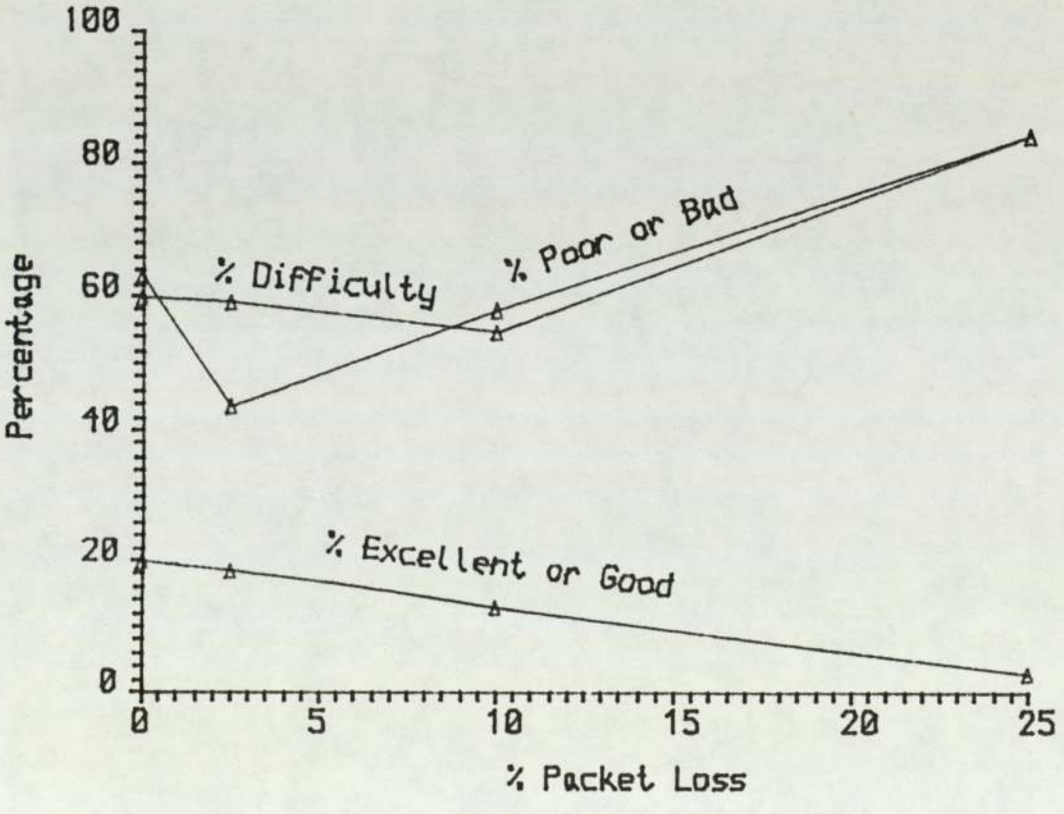


Fig. 3.7 Results of Conversation Test at 34dB ORE

shown in Fig. 3.8. (It has been assumed for these graphs that those who were vague in their comments were actually commenting on both. Of course many people did comment on both.) The percentage who commented on the loss or background hiss remains fairly constant, whereas the percentage who commented on the distortion is about the same as the "% difficulty " for the 14 dB ORE connection. Only for 25% loss did anyone comment on the distortion without mentioning the low sound level as well.

If packet loss and loudness loss are totally independent degradations, the probability of difficulty with a combination of the two is simply the sum of the individual probabilities minus the product(34). The expected difficulty curve produced on this basis for the 34 dB ORE connection is shown in Fig. 3.9, along with the actual difficulty curve. From the difference between curves, it can be concluded that only a high level of packet loss has an effect independent of loundess loss.

Further insight into how people reacted to the packet losses can be gained from the comments explaining why they had difficulty. Tables 3.6 and 3.7 list a synopsis of these for 25% loss and 10% loss. (The few comments on 2.5% loss simply referred to occasional crackle or croakiness.)

The zero-packet-loss mean opinion score (MOS) for 14 dB ORE corresponds well to the MOS at 14 dB given in reference (34).

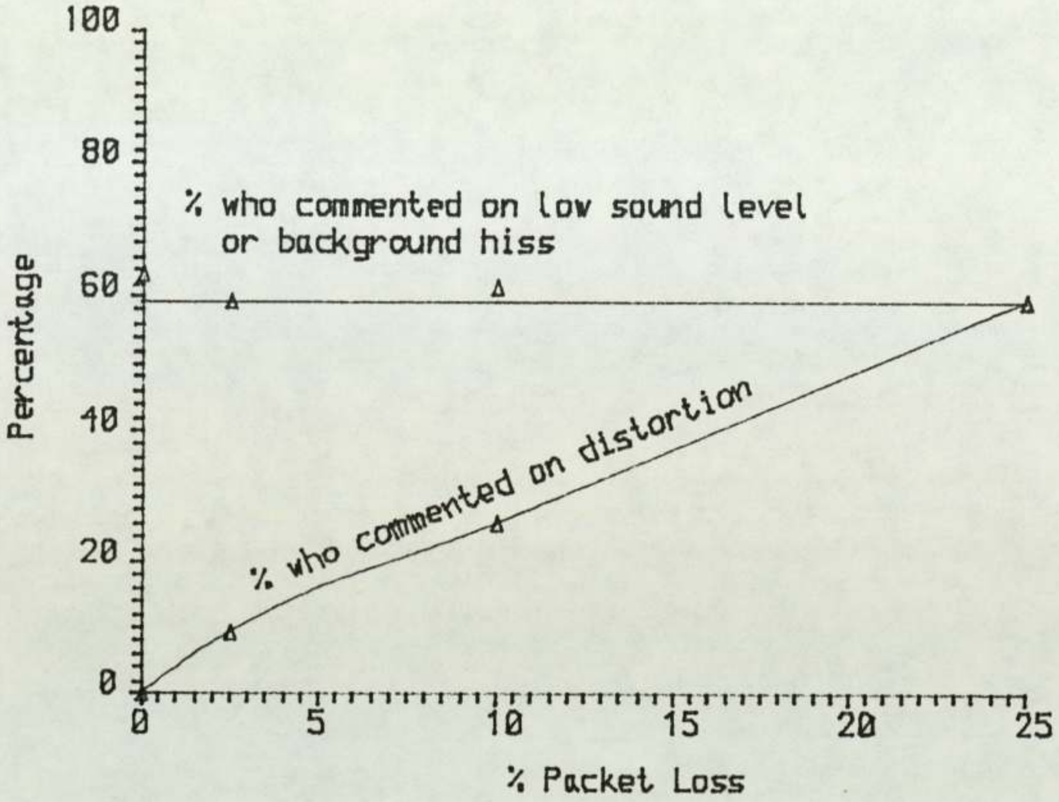


Fig. 3.8 Analysis of comments on 34dB ORE connection

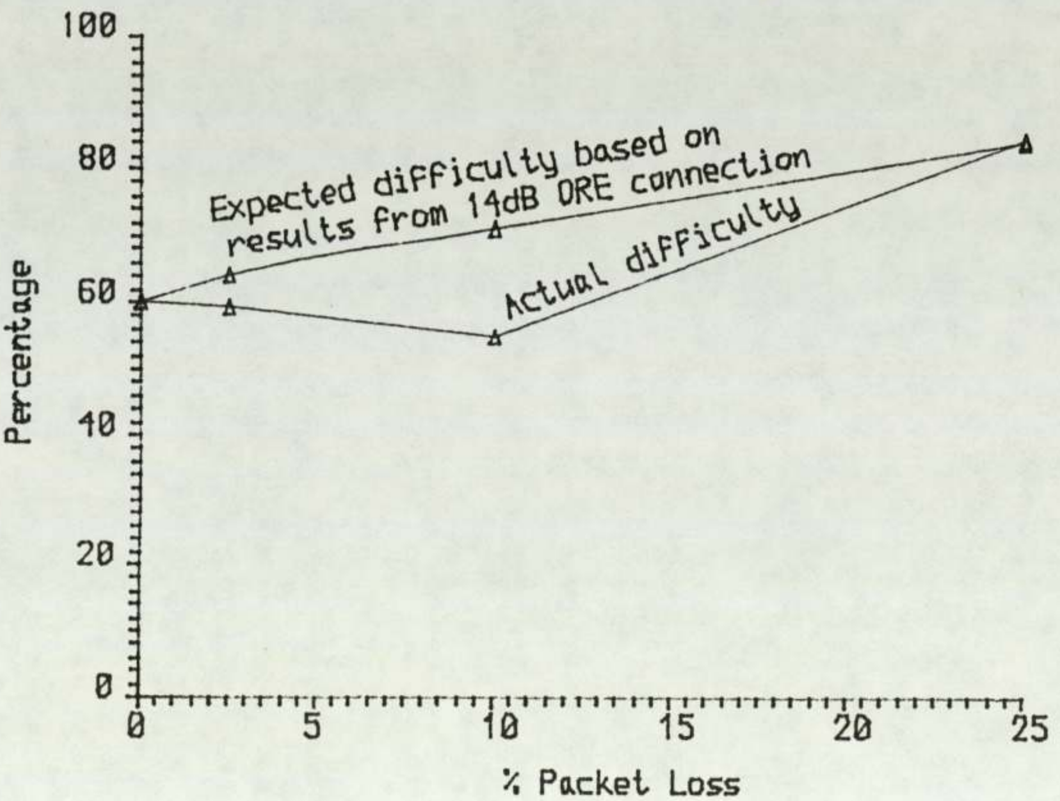


Fig. 3.9 Expected and actual % difficulties at 34dB ORE

TABLE 3.6 Comments on 25% Loss

Fuzzy speech. Very croaky (but could hear as I got used to it). Good for high pitch, croaky and off-putting for low pitch. Very husky. Coarse line. Voice sounded gruff. Rattly and croaky. Rasping. Bad crackle. Interference. Voice distortion. Voice doubled. Voice fuzzy - hearing two voices. Resonance (on high frequencies). Sounded like a Dalek. Tinny 'Dalek' quality, sort of buzz. Voice seemed to quaver/vibration of voice. "Wobbly" and not very pleasant. Was glad when conversation ended as the voice was getting on my nerves. Echo to each word making voice indistinct, particularly 's' sounds. Sounded as though down a tunnel with a touch of distortion. Ringing - could understand but not recognize voice.

TABLE 3.7 Comments on 10% Loss

Slight buzzing
(Slightly) husky
A little croaky
Fuzzy quality to some words
Some words distorted
Crackle made it hard to catch odd word now and again
Lacked clarity at times - a bit noisy.
Voice repeated slightly later
Sounded like a radio off-tune
Didn't sound like the voice I know
Distortion irritating.

This confirms that the results for the 14 dB ORE connection are reasonably reliable. However, the zero-packet-loss MOS for 34 dB ORE is much worse than given in reference (34). This may be due to uncontrolled test conditions (probably room noise), but it is more likely to be due to the particular set of subjects used. Very few used the phone for business and they would therefore not be very experienced at conversing over bad lines. Certainly, it was often noticed that one person tended to shout because they heard a quiet voice and the other person went still quieter because they heard a loud voice. Fortunately, several useful conclusions can be drawn despite this discrepancy.

3.5 CONCLUSIONS AND RECOMMENDATIONS

3.5.1 Conclusions

The preliminary test showed that the packet length that gives the best speech quality when packets are lost is about 8 ms, although 16 ms packets are not very much worse. It also showed that it is much better to fill in packet losses by repeating the last packet than to leave a gap. Consequently, 8 ms packets with filled-in gaps were used in the other tests.

The listening test showed that speech with losses up to 20% could be understood by most people with moderate effort. It also showed that the packet-loss distribution does not make a very big difference to opinion scores, although it did not prove that it makes no difference. The reason for this is that the large variability of opinion scores (which appears to be a feature of all tests on packet losses) means that two scores could only be proved to be the same if an extremely large sample of people were used.

The conversation test showed that an average of 2.5% loss was hardly noticeable, an average of 10% loss was acceptable to two thirds of the subjects and that an average of 25% loss was acceptable to one third of the subjects. ("Acceptable" here means that the link was rated at least "fair" and no difficulty was experienced.)

These results are very different from those of the formal tests described in references (60) and (59). However, they are similar to those of the informal tests described in references (61) and (5). This further emphasizes that the criterion of 'acceptability' in reference (60) (the only formal test to use repeated packets in gaps), was not suitable for a listening test.

Although packet loss was still noticed at 34 dB ORE, it is clear that it did not affect opinion scores until it reached 25%. Thus, there is an interaction between the effects of packet loss and loudness loss that completely masks the effect of packet loss until it reaches high levels. This means that, so far as interaction with loudness loss is concerned, it is not necessary to keep packet-loss levels lower when a network is connected to the public network. Unfortunately, it is not possible to say whether the same result would hold for other common degradations.

Finally, the scores for packet loss at 14 dB ORE can be compared to the zero packet loss score at 34 dB ORE. It can be seen that 25% packet loss is roughly equivalent to 20 dB loudness loss.

3.5.2 Some Recommendations

As people's opinions of packet losses change quite gradually as the loss increases, it does not make sense to set a threshold for the highest acceptable loss; presumably it would be acceptable to exceed any threshold occasionally. Thus, the designer should ensure that the normal level of loss is less than 2.5% (or perhaps 5%), and that 10% loss only occurs infrequently. Losses of around 25% ought to occur only rarely, perhaps being allowed only when an equipment fault (e.g. a failed link or a number of failed speech detectors) suddenly increases the load on a link. If this happened, 25% loss would certainly be better than disconnecting existing calls.

One interesting observation that follows from the result that high levels of loss are fairly acceptable concerns predictive and adaptive coders. As mentioned in Section 3.1.4, these will perform badly when packets are lost, due to loss of prediction and step-size information, although this information can be included in the packet header at the cost of increasing the overhead. However, if the amount of packets that can be lost when this information is included is considerably larger, (which it almost certainly is), the increase in overhead is worth having.

CHAPTER FOUR

OPTIMIZING THE PERFORMANCE OF PACKET-SWITCHED SPEECH NETWORKS

CHAPTER 4

OPTIMIZING THE PERFORMANCE OF PACKET-SWITCHED SPEECH NETWORKS

4.1 INTRODUCTION

Chapter 2 described a method of analysing a speech-packet queue that was both accurate and low on CPU time. Using this method, predictions of average packet loss (and queueing delay if required) can be made for any combination of parameters. Chapter 3 described tests which show how telephone users respond to various amounts and types of packet loss. From the results of these tests, certain recommendations were made as to how much loss should be allowed.

Together, these provide a powerful tool which is used in this chapter to show how the performance of a single packet-switched speech link can be optimized. In Section 4.9 it is shown how the principles established for a single link can be applied to networks.

The issues investigated in this chapter are; where and how packet loss should be controlled, how the queueing limit should be chosen and how it should vary with load, optimum packet length, receiver buffering, and how adequate performance for data can be maintained. An investigation of the trade-off between TASI advantage, delay and packet loss shows how much advantage packet switching has over switching systems that do not allow buffering or large amounts of speech loss.

4.2 WHERE PACKET LOSS SHOULD OCCUR

It has been shown that small amounts of packet loss (i.e. less than 2.5%) are barely noticable. This, however, is packet loss that occurs before transmission, when the loss can be evenly distributed between a number of channels. As mentioned in Section 1.2.4, many researchers have assumed that packet loss can take place after transmission where it is used to avoid having to wait for packets delayed beyond an acceptable limit[15,22,26,27,41-43]. In this case, as delays of successive packets are normally correlated, a number of packets will be lost in succession and the loss must be kept to much lower levels (probably the 0.5% used with TASI[4]). In other words, packet losses that occur after transmission are not as tolerable as those that occur before transmission.

There is also a considerable difference in the amount of loss that need take place before and after transmission. This is shown in Fig. 4.1 for three different link capacities with each link carrying twice its circuit-switched capacity (the speech activity is 42%). The dashed lines are the c.d.f. of delay with no pre-transmission loss. These can be interpreted as the fraction of packets that would have to be lost if the receiver were to discard all packets delayed more than the delay shown on the x-axis. By contrast, the chained lines are the fraction of packets lost if the switching node were to discard all packets delayed more than the delay shown on the x-axis. The difference between the two is about an order of magnitude.

As the number of calls carried on a link increases, the difference increases. Indeed, if no packets are discarded, the queueing delay increases exponentially with load, reaching ∞ at full load, thereby producing 100% packet loss. However, discarding packets at the switching node causes packet loss to increase only linearly with load, and congestion is avoided.

It may be concluded that, for efficient use of transmission capacity, packet loss must occur at switching nodes. Consequently, all the results of this chapter assume that this is where the queueing delay will be controlled.

4.3 THE USE OF PACKET LOSS

4.3.1 A Different Approach to Packet Loss

When considering how much packet loss or delay can be allowed, it is common practice to set a threshold of acceptability. Loss or delay below this threshold is considered acceptable, but anything above it is considered unacceptable.

This approach seems quite unreasonable. For both delay and packet loss there is a large "grey" area where the connection is most certainly usable, though some people may have a bit of difficulty from time to time (see Chapter 3 and reference (37)). Furthermore, no user is likely to complain if he occasionally finds the connection rather poor. For this reason it was suggested in Section 3.5.2 that levels of packet loss of 10% and above could be allowed to occur occasionally. This indicates that occasional demands for a large number of connections can be met by allowing packet loss to worsen, instead of making more channel capacity available or blocking (which would be necessary if a strict limit to packet loss existed).

Taking this approach, the capacity necessary to carry a certain quantity of traffic must be determined in a very different way from circuit switching. With circuit switching the important consideration is that less than 5% and preferably less than 1% of calls are blocked. However, with packet switching the important consideration would be that the speech quality should meet some given specification.

Evidently, average packet loss is not a suitable criterion as it does not indicate how bad the speech quality can get. A better specification would be the percentage of time a packet loss greater than 10% is encountered by a user. Blocking would still have to be used to prevent very bad losses. However, as even 20% or 25% loss allows communication, and as such large losses should occur only very occasionally, the blocking probability would be extremely small.

4.3.2 Packet Loss on a Single Link

To illustrate this approach to packet loss, a single link with no alternative connections will be considered. The variation in demand is given by the formula[81]

$$p_c(v) = \left(\frac{A^v}{v!}\right) \Big/ \sum_{r=0}^{N} \frac{A^r}{r!}$$

where $p_c(v)$ = probability of v calls

N = no. of calls at which blocking occurs

A = average traffic in Erlangs

The probability that a caller shares the channel with at least v-1 others, $p_u(v)$, is therefore given by:

$$p_u(v) = \sum_{r=v}^{N} p_c(r).\frac{r}{A} = \sum_{r=v}^{N} p_c(r-1)$$

Let v_L be the lowest value of v that gives a packet loss greater than L for some specified delay. If a link must provide a connection with losses greater than 10% less than 2% of the time, the capacity must be such that

$$p_u(v_{10}) < 0.02$$

Connections of this quality are illustrated in Fig. 4.2 for traffic levels of 4, 11 and 29 Erlangs and a delay of 0.5 μ^{-1}. On each loss v. maximum delay curve is the percentage of time a user will find himself on that curve or a worse one, assuming that call blocking takes place at about 20% loss. For a circuit-switched channel to carry the same traffic with less than 1% blocking, 10, 19 and 40 channels would be needed instead of 4, 8 and 16. Thus, in this case, packet-switched speech needs 60% less channel capacity than circuit-switched speech (ignoring the header overhead).

It is interesting to note that the following inequalities hold for all three graphs:

$$p_u(v_{10}) < 0.02$$

$$p_u(v_5) < 0.06$$

$$p_u(v_{2.5}) < 0.1$$

In the light of the conclusions of Chapter 3, this performance certainly seems to be perfectly adequate.

If alternative routing is available the blocking probability can be much higher, as blocked calls can take another route. In this case, a very low threshold for packet loss (about 2%) can be used on the link, which of course rather reduces the TASI advantage. However, the approach described here is still applicable, as even the alternative routes will be busy from time to time.

4.4 THE USE OF DELAY

4.4.1 Introduction

In Section 4.2 it was shown that, by setting a maximum queueing delay at the switching nodes and losing packets to ensure it is never exceeded, more efficient use can be made of a link than if all packets are queued and transmitted. The question then arises as to what this maximum delay should be. One approach to choosing the maximum delay would be to determine a threshold of acceptability for delay and set it to that. However, for each pair of values of v and h there must be a value of delay which minimises the combined degredations of packet loss and delay. In theory at least, this is the optimum value for the maximum delay.

In the following sections it will be assumed that the receiver buffers all packets to the maximum delay. So long as a timestamp is included in the header, this is certainly the simplest and safest method of operating the receiver buffer. The possible benefit of a different buffering strategy is discussed in Section 4.6. It should be noted that only the first packet of a talkspurt need carry a timestamp. Subsequent packets need carry only a short (perhaps two-bit) sequence number.

4.4.2 An Expression for the Effect of Delay

In order to find the optimum delay, an expression is required relating delay to the probability of a user having difficulty.

Klemmer's results[35] suggest that the probability of rejecting a call increases from zero at 600 ms roundtrip delay, through 0.13 at 1200 ms to 0.27 at 2400 ms roundtrip delay. (This is assuming that the "sensitized" results from the last ten weeks of his experiment are the most reliable since only then did people begin to associate confused situations with the connection). However, D L Richards [33, 34] has shown that confusions occur in proportion to delay even for small delays (600 ms roundtrip delay or less). This suggests that Klemmer's zero rejection rate for 600 ms roundtrip delay is likely to be due to the subjects still not associating the confusions with the line.

In view of D L Richard's result it will be assumed that the probability of difficulty rises linearly between 0 and 1200 ms roundtrip delay. If it is also assumed that rejection in Klemmer's test is equivalent to a user stating that difficulty was experienced, the probability of difficulty rises linearly from 0 right through to 2400 ms roundtrip delay.

The probability of difficulty with a one-way delay of D seconds (assuming the delay is encountered in both directions) is therefore approximately given by

$$p_D = 0.225\ D \qquad\qquad D < 1.2 \text{ sec} \qquad\qquad (4.1)$$

As the exact effect of delay is so uncertain (other papers cast doubt on both Klemmer's and D L Richards' results[35, 36], the effect of using two alternative expressions is considered in

Section 4.4.5.

4.4.3 An Expression for the Combined Effect of Delay and Packet Loss

From Figure 3.6, the probability of difficulty with a fractional packet loss of L, p_L, is given approximately by the formula

$$p_L = 3.23L - 3.3L^2 \qquad L < 0.25 \qquad (4.2)$$

Assuming that the effects of packet loss and delay are independent (there is no reason to suppose they are not), the probability of difficulty with a fractional loss L and D seconds delay, p_{LD}, is given by

$$p_{LD} = p_L + p_D - p_L p_D$$

$$= 3.23L - 3.3L^2 + 0.225D(1 - 3.23L + 3.3L^2) \qquad (4.3)$$

4.4.4 Optimum Delay

Fig. 4.3 shows how the probability of difficulty given by Eq. 4.3 varies with delay for a channel capacity of 4 and various activities. It has been assumed for these graphs that an average talkspurt length is 1.2 secs. From these graphs the following observations can be made.

Firstly, a delay of 312 ms gives the least (or very close to the least) difficulty when the difficulty rating is greater than about ten percent. This means that a link can operate very well with two values of delays - one large and one small. As the packet loss rises above a threshold the delay is increased, and when it falls below another threshold the delay is decreased. However, if speech quality is to be unaffected by the change in delay, the change must not take place until each talker has fallen into silence. This allows the receiver buffers to adjust to the new delay by changing the length of a silence. As it takes several seconds to allow each talker to stop speaking, the delay could not be changed very frequently. For this reason the two thresholds need to be well separated, as shown in Fig. 4.3.

Secondly, the optimum delay is independent of activity. This means that variations in activity (perhaps caused by one-sided conversations or noisy backgrounds) do not affect the optimum delay, though they evidently do affect the difficulty rating.

Thirdly, the range of delays that give difficulty scores close to the optimum is very wide. In Fig. 4.3 any delay from 200 ms to 500 ms would be reasonable.

Fig. 4.4 shows similar graphs for channels of capacity 8, 16 and 32. The optimum delay decreases as the channel capacity increases until, with a capacity of 32, a single delay of about 60 ms would almost be sufficient.

The feature these graphs do not show is the sharp rise in packet loss exhibited in Fig. 2.10 as delay drops below half the packet length. For this reason, the smaller delay ought not drop below about one third to one quarter the packet length, even if it appears from the analysis that there may be a small advantage in doing so.

4.4.5 Alternative Expressions for the Effect of Delay

As Eq. 4.1 is only approximate, some possible alternatives are considered in this section. This will show the sensitivity of the operating points found in the previous section to the particular formula used.

The first alternative is based on the results of call-back interviews on connections using echo cancellors[29]. The percentage of users who said that the connection was unacceptable was 9.5 for 540 ms and 21.3 for 1080 ms round-trip delay. Assuming that % difficulty and % unacceptable are equivalent, the expression for p_D becomes approximately

$$p_D = 0.375 \; D \qquad (4.4)$$

Strictly speaking, this expression is valid only for $D < 540$ ms. However, it will be assumed here that it applies for $D < 1.2$ sec. It will be clear from the results that this assumption is of little consequence. Comparing Eq. 4.4 with Eq. 4.1, Eq. 4.4 suggests that the difficulty with delay is 66% worse than that predicted by

Eq. 4.1. Eq. 4.4 is represented by line (a) in Fig. 4.5.

The second alternative is merely an attempt to get closer to Klemmer's "unsensitized" results, and it attributes very little degradation to the smaller delays.

$$p_D = 0.1875\ D^2 \tag{4.5}$$

This equation is represented by line (c) in Fig. 4.5.

Of the two alternatives, Eq. 4.4 is likely to be closer to the truth, especially as echo cancellors will have to be used until digital subscriber loops become widespread. However, it is hoped that these two expressions represent reasonable bounds on the probability of difficulty with delay.

Fig. 4.6 shows the difficulty v delay graphs obtained when each of the three expressions for p_D are used. In addition, the optimum delay when Eq. 4.1 is used is shown. Whilst the optimum delays are very different for the three alternatives, the use of the Eq. 4.1 optimum does not effect the difficulty rating very much. Thus, near-optimum operation can be obtained without an exact knowledge of the exact effect of delay.

4.5 OPTIMUM PACKET LENGTH

4.5.1 The Problems

Packet length is only an issue because of the conflict between the large header overhead of a short packet and the poor speech and conversational quality produced by a long packet. If the header overhead can be reduced to a few bits by grouping packets together into larger packets(21) or some other method, packet lengths can be very short without any significant header overhead. Only when protocols do not allow this is there a problem. Consequently, the discussion here concerns the case when conventional packet switching must be used.

The optimum packet length is very much dependent on both the header size and the bit rate of the speech coder. It is generally accepted that a header need be at most 32 bits(3, 21), unless X25 protocol must be adhered to, in which case 64 bits are probably needed(15). The header can be this short because packet-switched speech can work very well on a virtual-call basis, and because no error correction is needed (except for the header itself). With care it ought to be possible to use a still smaller header(3, 40).

The bit-rate of a speech coder may be anything from 2.4 to 32 kbits/sec. 64 kbits/sec PCM is not likely to be used with packet-switched speech since conversion to 32 kbits/sec ADPCM(82) would only slightly increase the processing already needed for echo cancellation and speech detection.

There are two problems with increasing packet length to reduce the header overhead. The first is that packetisation and transmission delays increase, and the second is that the acceptability of packet losses decreases. These two problems will be considered separately.

4.5.2 Packetisation Delay

To understand the effect that packetisation delay has on the choice of packet length, it will be assumed in this section that packet length does not affect the acceptability of packet losses. This would actually be true if embedded encoding were used, since packets would be shortened instead of lost altogether, and so packet length would not affect speech quality.

The shorter a packet, the larger is the header overhead. The larger the header overhead, the smaller is the useful channel capacity and the greater is the packet loss (if there is any). Thus, as packetisation delay decreases, packet loss is increased. This is a similar sort of trade-off to that between packet loss and queueing delay considered in Section 4.4; thus the optimum packetisation delay (and therefore length) can be found in the same way as in Section 4.3.4.

The problem is that the optimum length depends on the amount of packet loss and therefore on the number of calls carried by the link. It can easily be shown that the optimum length for a heavily-loaded link is much greater than the optimum for a lightly-loaded link. Therefore a compromise must be reached. One possible alternative is to change the packet length as the loading of the link changed, but this would probably be more trouble than it was worth.

The essence of the compromise is to have a packetisation delay that does not cause much difficulty when the link is not heavily loaded, and does not increase difficulty too much (by having too short a packet) when the link is heavily loaded. This is illustrated by the example considered in Table 4.1, where the predicted difficulties at both a light load and a heavy load are given as a function of packet length. Predicted difficulties based on both Eq. 4.1 and Eq. 4.4 are given to show the possible error in the results.

In the example considered, the channel capacity is 32. The "light load" columns represent the case where there are less than 50 talkers. The "heavy load" columns represent the case where there are 75 talkers. The queue limit for the light load is one quarter of the packet length, which should give negligible packet loss. The queue limit for the heavy load is 75 ms. From Fig. 4.4(c) this is roughly the optimum queueing limit for a link of capacity 32 under heavy load. (As we are mainly concerned with packets less than 300 ms in length, this limit has been used regardless of packet length). The packet loss for the heavy load is given in column 3. A 32-bit header is assumed, and both a 2.4 kbits/sec coder and a 32 kbits/sec coder are considered.

TABLE 4.1 Trade-off between Difficulty Rating and Packet Length

			Eq. 4.4		Eq. 4.1	
% Header overhead	Packet length (ms)	% Packet loss with 75 calls	p_{LD} light load	p_{LD} 75 calls	p_{LD} light load	p_{LD} 75 calls
1	1320	3.46	63.43	58.84	38.06	39.62
2	653	3.97	31.39	36.92	18.84	27.08
3	431	4.52	20.73	30.71	12.43	24.00
4	320	5.12	15.40	28.49	9.24	23.36
5	253	5.76	12.17	27.91	7.30	23.74
6	208	6.44	10.01	28.19	6.01	24.68
7	177	7.17	8.52	29.06	5.11	26.02
8	153	7.92	7.37	30.20	4.42	27.52
9	134	8.71	6.45	31.59	3.87	29.20
10	120	9.53	5.78	33.18	3.47	31.02

a) 2.4 kB/sec coder

			Eq. 4.4		Eq. 4.1	
% Header overhead	Packet length (ms)	% Packet loss with 75 calls	p_{LD} light load	p_{LD} 75 calls	p_{LD} light load	p_{LD} 75 calls
0.8	124	3.36	5.96	17.29	3.57	14.56
1	99	3.46	4.75	16.71	2.85	14.33
2	49	3.97	2.35	16.43	1.41	14.77
3	32	4.52	1.54	17.42	0.93	16.02
4	24	5.12	1.15	18.82	0.69	17.56
5	19	5.76	0.91	20.73	0.55	19.27

b) 32 kB/sec coder

Taking the 2.4 kbits/sec coder first, it can be seen that the optimum packet length at heavy load is 250 to 320 ms, depending on the difficulty formula used. This would give 10 to 12% difficulty under light load, which is high considering that this would be the normal operating condition of the link. A compromise in this case would be to use a packet length of 120 to 150 ms, which would give about 5% difficulty under light load at the expense of increasing difficulty by about 5% under heavy load. Thus, in either case, the link would be operating at about 5% difficulty more than its optimum.

The optimum packet length for the 32 kbits/sec coder at heavy load is about 50 to 100 ms. A compromise in this case would entail a packet length of about 32 to 50 ms where under both loads the link would operate at 1 to 2% difficulty more than the optimum. It can be seen that the increase in difficulty is much less for the higher bit-rate coder. This is consistent with the quality of the speech coder itself, which at 32 kbit/sec would be considerably better than at 2.4 kbits/sec.

If embedded encoding were used, the losses would probably be more acceptable than indicated by Eq. 4.2 (used in Table 4.1). Thus, the optimum value of packet length would be less than the figures given here. If embedded encoding were not used, the way the acceptability of packet losses varies with packet length would have to be taken into account. This is considered in the next section.

4.5.3 Acceptability of Losses

The results of the test described in 3.2 show that for packet lengths between 8 and 32 ms, speech containing packet losses is more intelligible and acceptable for shorter packets. For packet lengths above 32 ms, it has been shown that until about 250 ms the intelligibility decreases as the length increases[62, 63]. The same applies to the opinion score of packet losses with gaps filled by silence[59]. Although no tests have been published on the effect of packet length on opinion scores when gaps are filled by the previous packet, the same relationship must to some extent still apply.

However, the shorter a packet is, the larger is the header overhead and the smaller is the useful channel capacity. Thus for a heavily loaded link, the shorter the packet, the greater is the packet loss. Therefore there is a trade-off between the amount of packet loss and the acceptability of those losses.

Unfortunately, it is not possible to know the exact nature of this trade-off because of the lack of published results of tests on packet losses where gaps have been filled by the previous packet. It is most likely that it would cause packet lengths to be shorter than if packetisation delay were the only consideration. Further research is needed to find out precisely how much shorter the packets would have to be.

4.5.4 Conclusions on Optimum Packet Length

The example outlined in Table 4.1 has shown that if packetisation delay alone is taken into account, the optimum packet length to go with a 32-bit header is between 32 ms and 150 ms, depending on the bit rate of the speech coder. For smaller headers, the optimum packet length is less.

However, these lengths are an upper bound to the true optimum. If embedded encoding were used, the optimum length would be less because of the greater acceptability of losses. If embedded encoding were not used the optimum length would be less because of the poor speech quality produced when packets greater than 50 ms in length are lost.

4.6 THE RECEIVER BUFFER

4.6.1 The Problem

Fig. 4.7 demonstrates the basic dilemma of the receiver buffer design. Continuous speech can be guaranteed by buffering each packet to the maximum delay. However, the average delay is then greater than necessary. The average delay can be reduced by buffering to less than the maximum, but the quality of the output speech is then worse than necessary.

In the previous sections of this chapter it has been assumed that the packets would be buffered to the maximum, mainly because this is the simplest and safest method of buffering. However, it may be possible to reduce the difficulty rating by employing a buffering scheme that reduces delay to compensate more than enough for the poorer quality speech that will inevitably be produced.

The following sections discuss the loss of speech quality and the reduction in delay that occur through the use of what will be called a "variable" buffer, i.e. a buffer that buffers each talkspurt out to a different delay. This discussion is necessary since all previous work on receiver buffering[18, 24, 28, 43, 45, 48] has not taken into account the existence of a maximum delay. Thus the obvious method of buffering to the maximum delay has never been considered.

4.6.2 Variable Talkspurt Delay and the Addition of Hangover

One of the consequences of variable buffering is that consective talkspurts will be delayed by different amounts, and so the length of the silent interval between them will be altered. Gruber measured this effect, which he called variable talkspurt delay(59), and found it to be very much dependent on the amount of hangover used in the speech detector. Because of this, he recommends that 200 ms hangover be used when variable delays may occur.

It is quite likely that these results do not apply to the system considered here. This is because the correlation between successive packets ensures that no great change in delay takes place between two consecutive short talkspurts, whereas in the TMS proposed by Gruber(2) a large change in delay is very likely. However, whether or not hangover affects variable talkspurt delay, it is undesirable to add it if the speech detector does not require it, as it increases the speech activity.

This is illustrated in Fig. 4.8, where TASI advantage is plotted against packet loss for three different values of v. Two sets of curves are shown. The dashed lines represent an accurate speech detector, which uses very little hangover (32 ms) and produces an activity of 36%(66), operated on a link with very small delay (so variable talkspurt delay cannot be a problem). The continuous lines represent a detector with 200 ms hangover, which produces an activity of 42%(66), operated on a link with a delay of 1 μ^{-1} .

For v = 20 and a TASI advantage of 2.48, the two detectors both give 5% packet loss. For larger v, the accurate detector is better, for smaller v the detector with hangover is better. Thus, for v > 20, there is no conceivable advantage in adding hangover to reduce variable talkspurt delay, since even a very large delay cannot compensate for the increase in activity. For v = 10 there is still no real advantage, since the effect of the delay itself would tend to counteract the few percent reduction in packet loss.

If the time of transmission (i.e. a timestamp) is included in the first packet of a talkspurt, the buffer can reconstruct the small silences exactly, and only change the longer silences where the variation is not so noticeable. This would have the same effect as hangover without having to waste transmission capacity by transmitting silences. As most buffering strategies would require a timestamp anyway, this is surely a better way to guard against variable talkspurt delay than the unnecessary use of hangover.

4.6.3 Buffer Underflow

Another effect of variable buffering is buffer underflow, the name given by Forgie[58] to the effect of leaving gaps in the speech when there are no packets in the buffer to read out. Unless packets are buffered to the maximum, some buffer underflow is inevitable as the possibility always exists of a packet being delayed by more than the buffer allowed for.

The effect of buffer underflow, according to some informal tests[58], is similar to that of packet loss. A maximum bound to buffer underflow can be obtained by considering the case when no buffer at all is used. Then buffer underflow will occur whenever the queue is increasing (i.e. when $j > h$ and the queue is not on its limit). The ratio of gaps to speech is given by

$$\sum_{j=h}^{v} \frac{(j-h)}{j_{ave}} \int_0^m q_j(x)dx \tag{4.6}$$

However, packet loss is given by

$$\sum_{j=h}^{v} \frac{(j-h)}{j_{ave}} \pi_j$$

and $$b_j = \int_0^m q_j(x) + \pi_j \qquad j > h$$

Thus, it follows that with no buffer the total degredation due to loss and underflow (assuming their effects to be the same) is equivalent to a loss given by

$$\sum_{j=h}^{v} \frac{(j-h)}{j_{ave}} b_j$$

and is independent of delay. Thus, with no buffer, the buffer underflow increases as the packet loss decreases with delay.

Buffer underflow can be avoided by queueing talkspurts instead of packets, as in the TMS system proposed by Gruber[2] (see Section 1.2.5). As no analytical model exists that is accurate for very high efficiencies or substantial amounts of packet loss (see

Section 1.2.6 for the reasons for this), it is difficult to predict the effect TMS has on packet loss and delays. Because of this, it is not considered further here.

4.6.4 Delay

Variable talkspurt delay and buffer underflow are the disadvantages of variable buffering. The advantage is that the effect of a given maximum delay can be reduced, allowing larger delays to be used.

Although Gruber has conducted listening tests on variable delays[59], it appears that nothing has been published concerning the conversational effects of varying delay. The major studies[33-37] on the conversational effects of delay have all been concerned with a fixed delay. However, because the number of confusions is roughly proportional to delay[33, 34], it would be reasonable to assume that with a varying delay the average number of confusions due to delay is roughly proportional to the average delay. Thus, the conversational degradation can be considered to be proportional to average delay.

A minimum bound for average delay is the average delay with no buffering, given by Eq. 2.13. The more gap modulation and buffer underflow is reduced, the further above this bound will be the average delay.

4.6.5 Potential of Variable Buffering

The expressions giving buffer underflow and average delay when no buffer is used (i.e. Eqns. 4.6 and 2.13) can be used to show the potential of variable buffering in two ways. Firstly, they can be used to show what improvement on no buffer must be achieved to match the performance of buffering to the maximum delay. Secondly, using the average delay of no buffer but assuming buffer underflow to be negligible gives the best performance a variable buffer can possibly achieve.

In Fig. 4.9, the chained lines show the predicted difficulty when packets are buffered to the maximum delay (these are identical to those in Figs. 4.3 and 4.4). The continuous lines show the predicted difficulty for a buffer that, compared to using no buffer, reduces buffer underflow by 80%, but increases the average delay by only 20% of the maximum delay (labelled as "20% buffer"). The dashed lines show the best performance possible with a variable buffer, obtained using the average delay of no buffer given by Eq. 2.13 instead of maximum delay (labelled "lower limit").

It can be seen that a variable buffer must be able to improve considerably on the 20% buffer to have any advantage over buffering to the maximum. Whether this is possible or not can only be discovered by further research. It must also be remembered that whilst it is possible for people to adjust to a fixed delay in a conversation, it is rather difficult to adjust to a variable delay. Besides this, people who make occasional intercontinental telephone calls are to some extent used to the effect of a fixed delay. No-one is used to the effect of variable delays. Thus, however good the variable buffering strategy, it is unlikely that it can cause a very significant improvement to the performance of buffering to the maximum.

4.7 MAINTAINING ADEQUATE DATA TRAFFIC PERFORMANCE

If speech packets have priority over data packets (as has so far been assumed), a mechanism is needed whereby requests for speech calls can be blocked when the data queue is too big. As the response of the data queue to call blocking would be slow, it is better not to allow the data queue to grow too big. In other words, speech calls should be blocked before the data traffic exceeds the free capacity of the link.

Table 4.2 shows the amount of speech degredation that would have to be incurred for the free capacity of a link to drop to less than 20%, 10% and 5% of the total capacity, for the operating points shown in Figs. 4.3 (b) and 4.4 (a) and (b). With fairly large amounts of data traffic (e.g. 20% of link capacity) speech calls would be blocked when the degredation was quite low, especially for larger links. In this case, a better compromise between data and speech traffic performance can be reached by permanently allocating some of the capacity to data packets. This allows more reliable data transmission at the expense of raising the level of degradation at which speech calls are blocked. For example, 20% of the link can be given to data traffic either by giving speech absolute priority and blocking at about 80% utilisation, or allocating 16% of the link capacity to data traffic and blocking at about 95% utilisation. Table 4.2 shows the effect of this.

To find the best balance between speech and data performance for any given mix would involve an analysis of the data traffic

TABLE 4.2: The Degradation that Must be Incurred to Achieve Utilisations of 95%, 90% and 80% for the Operating Points Shown in Fig. 4.3 (b), and Fig. 4.4 (a) and (b).

	Channel Capacity	4	8	16
% Utilisation	% Free for data	Difficulty		
		%	%	%
95	5	45.0	28.5	16.3
90	10	32.5	16.8	8.3
80	20	15.9	7.9	1.9

performance, which is beyond the scope of this work. However, a reasonable approach would be to decide on the speech quality required, provide enough capacity to maintain this, and allocate the rest of the channel capacity to data.

4.8 THE TRADE-OFF BETWEEN TASI ADVANTAGE, DELAY AND PACKET LOSS, AND THE USEFULNESS OF DELAY

Whilst Fig. 4.2 shows how delay considerably reduces packet loss, it is not immediately obvious that a similar reduction could be achieved by a small increase in channel capacity. Thus there is a trade-off between delay and channel capacity as well as a trade-off between delay and packet loss. Fig. 4.10 shows this trade-off at three different values of loss, and a number of different values of v. In order to present curves for different values of v on the same graph, TASI advantage has been plotted instead of channel capacity.

It is clear that the gain in TASI advantage obtained through the use of delay decreases as the channel capacity increases. To a lesser extent it also decreases as the packet loss increases. For this reason, Weinstein's simulations(9), and more recently Janakiraman's analytical results(47) which worked to a loss of only 0.5%, slightly exaggerate the effect of delay.

To show more clearly the effect that packet loss has on TASI advantage, Fig. 4.11 shows similar curves to Fig. 4.10 but with packet loss instead of delay. The effect of packet loss is very similar to that of delay. The most important difference is that the limiting curve for loss as $v \rightarrow \infty$ has a positive gradient, whereas the limiting curve for delay is flat. This is simply because packet loss actually reduces the load on the channel, whereas delay can only distribute it more evenly.

As one of the prime differences between packet switching and other integrated switching methods is its use of delay, it is important to consider the gain in performance that delay gives. Although Figs. 4.10 and 4.11 show the trade-off between TASI advantage, delay and packet loss, they do not show the advantage of using delay because the degredation caused by the delay itself is not taken into account. To show the advantage of using delay, the increase in TASI advantage with delay for constant difficulty is needed. This can easily be found using Eq. 4.2.

There is no longer any need to plot a wide range of delays, since for any given channel capacity and loading, the delay which minimizes the difficulty (i.e. optimum delay) would usually be used. Plotting the difficulty at this optimum delay against TASI advantage gives the continuous lines of the graph shown in Fig. 4.12 (b). The dashed lines show the TASI advantage obtained with no delay, and the difference between the two lines is the gain in TASI advantage that delay gives at any particular level of difficulty. For the sake of clarity only three values of v have been used.

The graphs in Fig. 4.12 (a) and (c) show the gain in TASI advantage with delay when the alternative formulas for difficulty with delay, given by Eq. 4.4 and Eq. 4.5, are used. Unfortunately, the particular formula used makes a very significant difference, and so it is difficult to draw any definite conclusions concerning the usefulness of delay. However, it can be seen that delay is always of some use, although for large capacity channels the increase in TASI advantage is very slight. Assuming that Fig. 4.12 (b) is the more reliable of the three graphs, the increase in TASI advantage is only very significant in a network which carries small amounts of traffic.

4.9 SUMMARY OF PRINCIPLES AND THEIR APPLICATION TO A NETWORK

Section 4.2 shows that every queue must have a mechanism to lose packets when it grows too big. Later sections showed that this should not be regarded as an emergency action to combat severe congestion, but as an integral part of the design and dimensioning of a network. Since quite high levels of loss can be allowed at times, it is most important that the losses are distributed evenly over all the channels. In a network, this will be particularly important when a node is switching channels that have already suffered some packet loss. It needs to read the sequence numbers of all packets that are queued to ensure that the packet losses, including those that occurred at previous nodes, are evenly distributed. With embedded encoding, the packet length would make it obvious which packets had already been shortened.

From Sections 4.3 and 4.4 emerged the principle that a link needs to operate in two modes. The "light-traffic" mode uses very little queueing delay and has less than 2.5% loss; this would be the normal mode of operation. The "heavy-traffic" mode uses queueing delay and may have quite large levels of loss; this would only be used during peaks in the demand for speech connections. In a network, the route chosen during the call set-up period would have to avoid links that were in the heavy-traffic mode whenever possible. It may be preferable to block a call rather than route it through more than two links in heavy-traffic mode.

It was seen in Section 4.4 that the optimum queueing limit is a function of channel size, with channels of capacity greater than 32 not really needing queueing at all. The range of near-optimum values was very wide; the exact value of queueing limit is therefore not critical.

Section 4.5 showed that the best packet length to use is one which ensures that near to optimum performance is obtained for all traffic levels. How near to optimum this can be depends on the bit rate of the speech coder and the size of the header. The packet lengths which were obtained by considering packetisation delay alone are only an upper bound to the true optimum. This is because the exact effects that embedded encoding and different packet lengths have on the acceptability of losses are unknown.

Receiver buffering is not a problem when all links are operating in light-traffic mode as then very little queueing is allowed. When in heavy-traffic mode there is some doubt as to whether or not it is best to buffer all packets to the maximum possible delay, or to something less. Section 4.6 did not remove this doubt, but showed that it is not likely that buffering to less than the maximum delay will make a very significant difference.

In Section 4.7 it was pointed out that speech calls would have to be blocked if the data traffic could not all be sent in idle channel time. It was also shown that giving speech packets priority over data packets may mean that speech packets are transmitted with very little delay and loss whilst data packets have to queue for long periods of time. This imbalance in quality of transmission can be set right by allocating some of the channel permanently to data.

Finally, Section 4.8 showed how allowing queueing delay and packet loss increases the TASI advantage for different-sized channels. For both delay and packet loss the increase is largest for small channels. However, whilst there is no advantage in allowing large amounts of queueing with large channels, there is always some advantage in allowing large levels of loss.

The advantage of packet switching over systems that do not use delay was found by comparing the TASI advantage obtained using delay with that obtained with no delay for a wide range of difficulties. It was found that packet switching is most advantageous when the amount of speech traffic to be carried in the network is small. For networks carrying large amounts of speech traffic, packet switching does not need significantly less channel capacity than other SI systems (e.g. DSI). However, the flexibility of packet switching may well make it the best choice, even for large capacity networks.

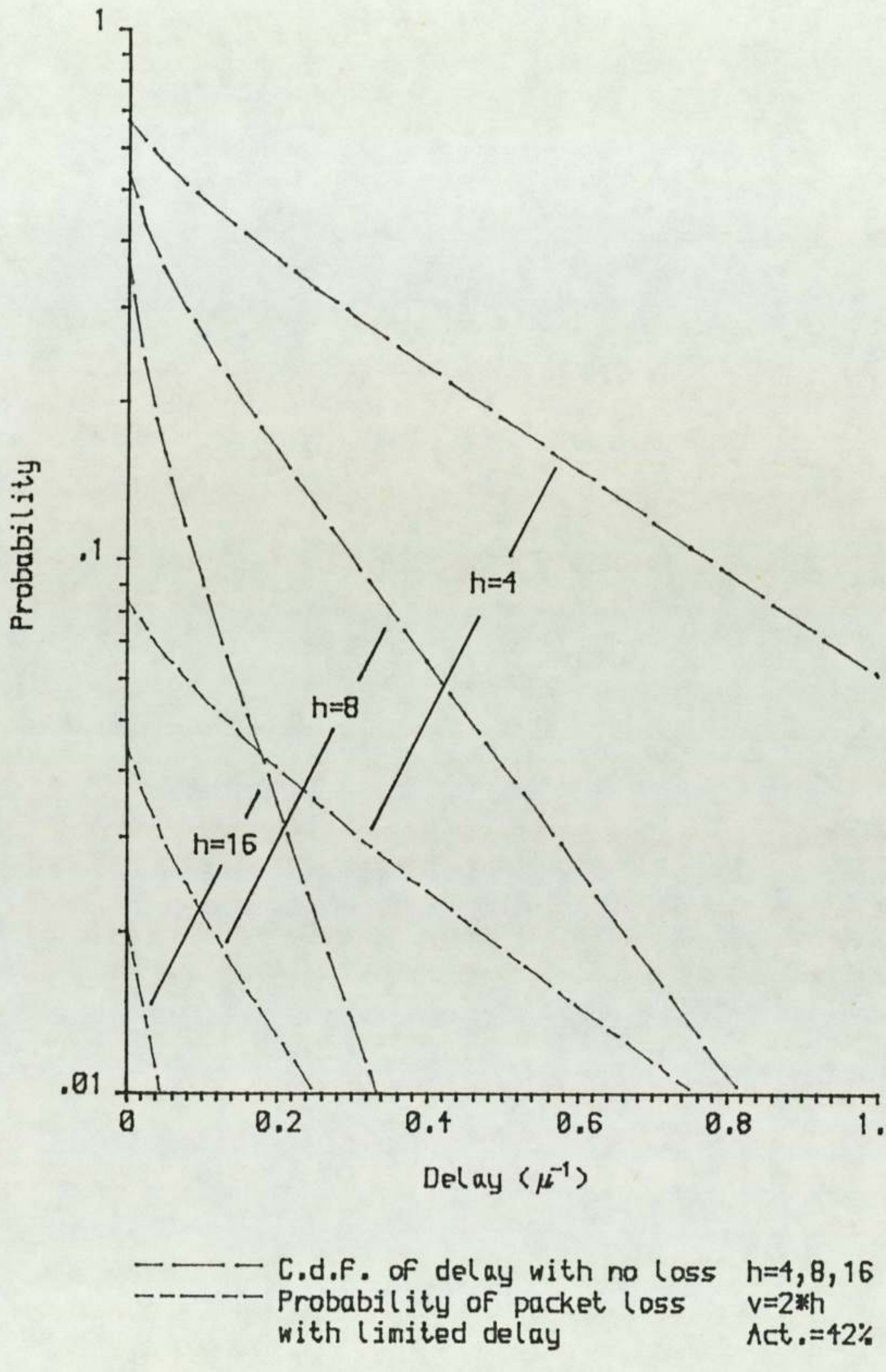


Fig. 4.1 Comparison of the performance of a queue with and without limited delay and packet loss

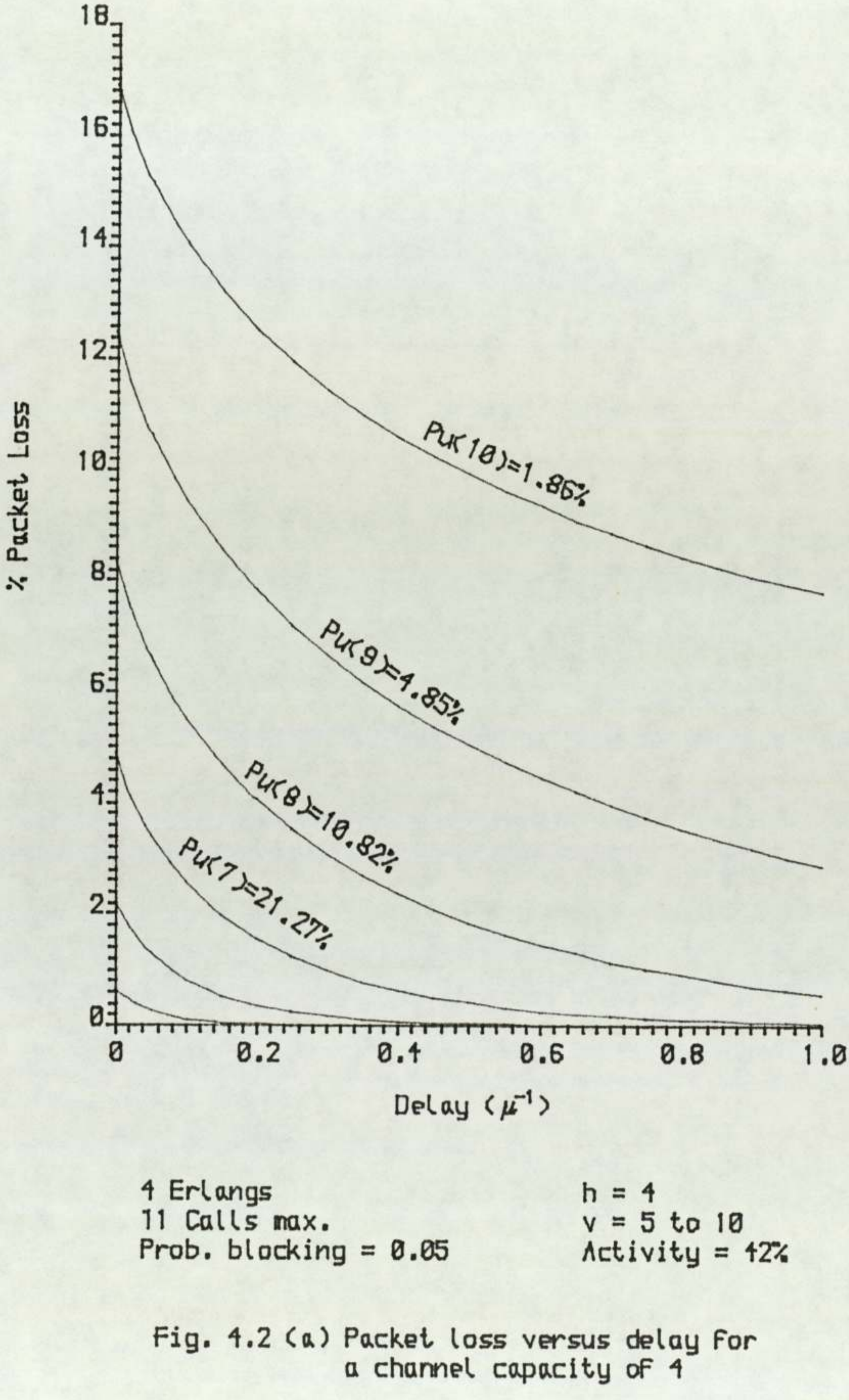
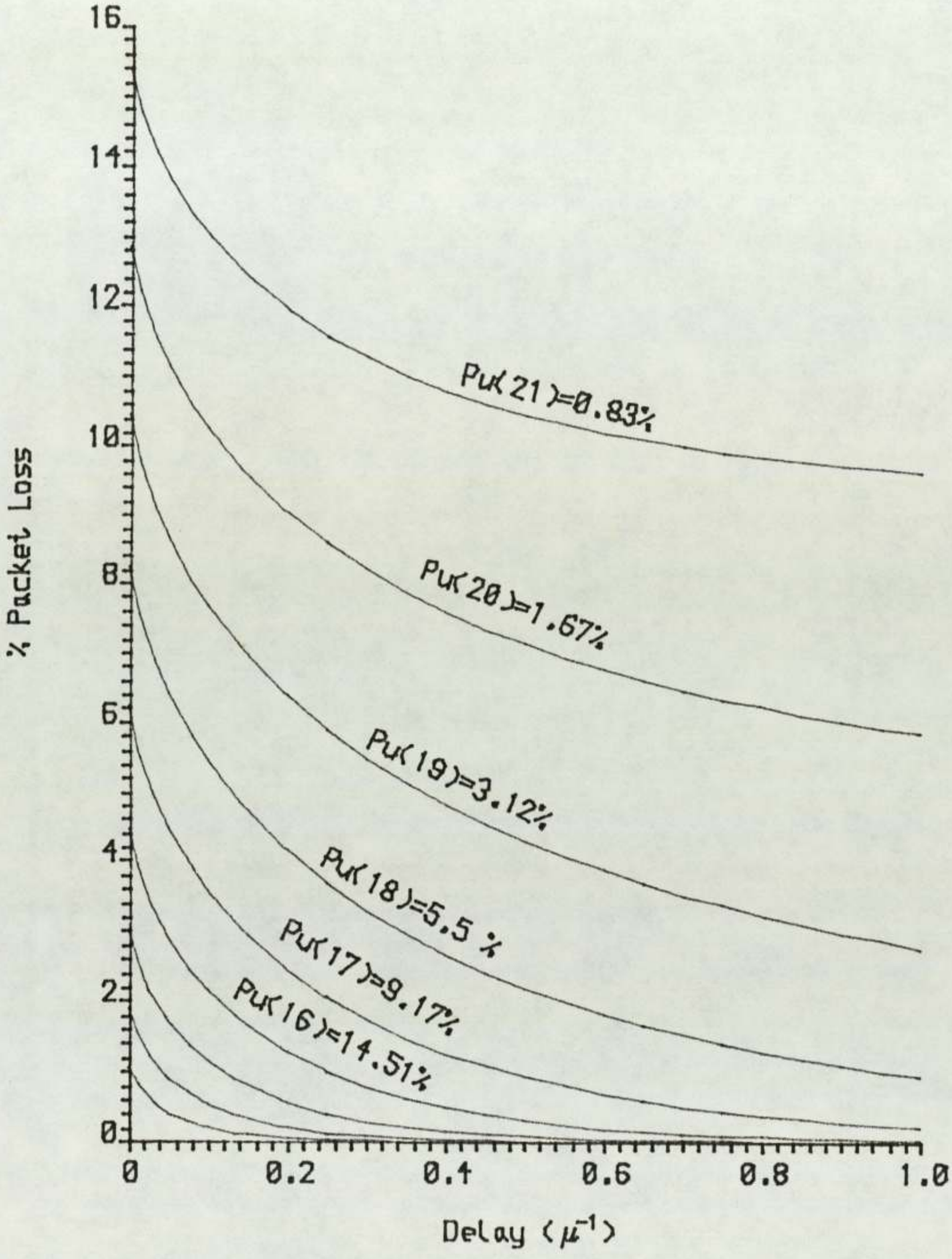


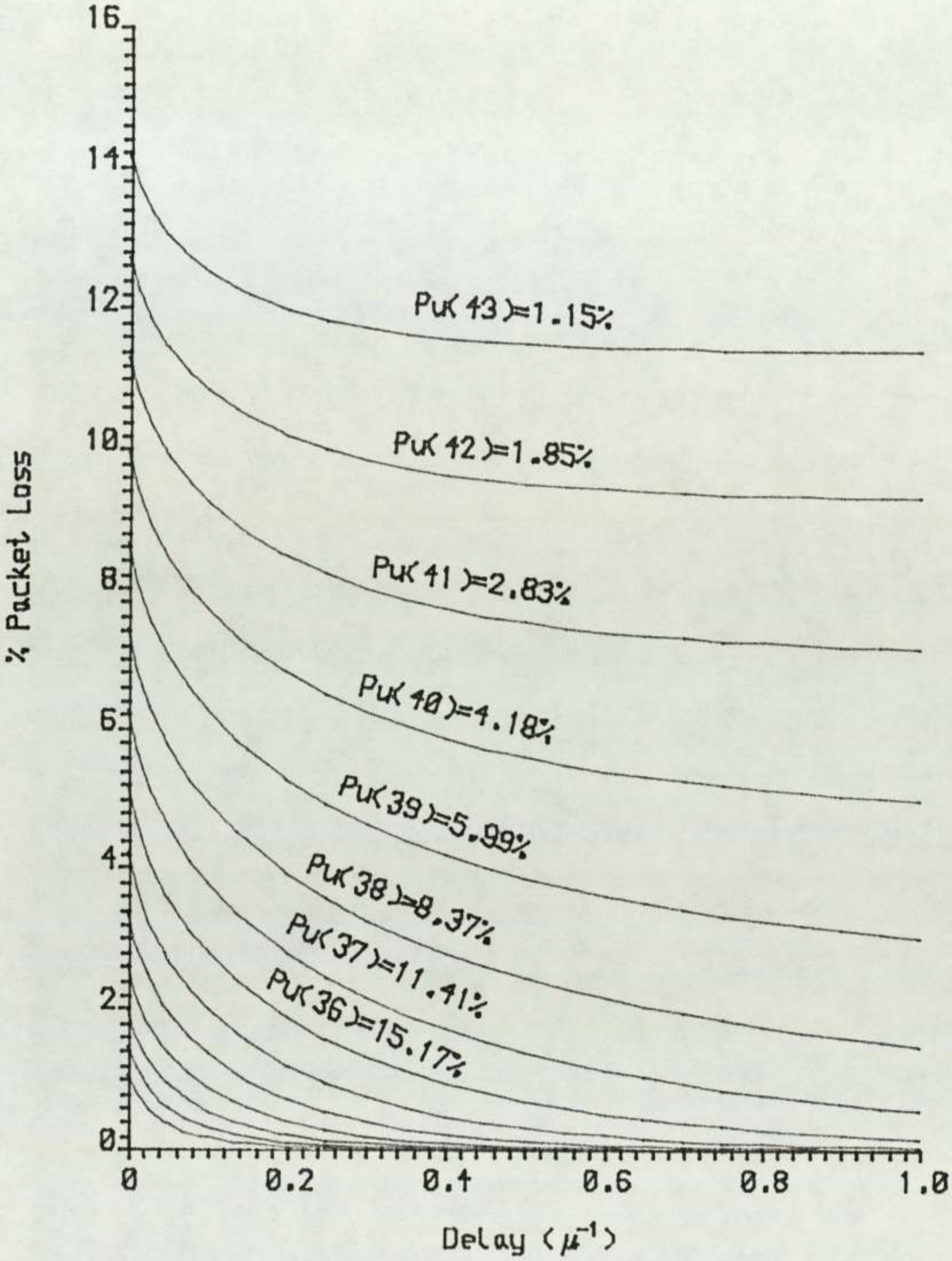
Fig. 4.2 (a) Packet loss versus delay for a channel capacity of 4



11 Erlangs
24 Calls max.
Prob. blocking = 0.05

h = 8
v = 13 to 21
Activity = 42%

Fig. 4.2 (b) Packet loss versus delay for a channel capacity of 8



29 Erlangs
46 Calls max.
Prob. blocking = 0.08

h = 16
v = 30 to 43
Activity = 42%

Fig. 4.2 (c) Packet loss versus delay for a channel capacity of 16

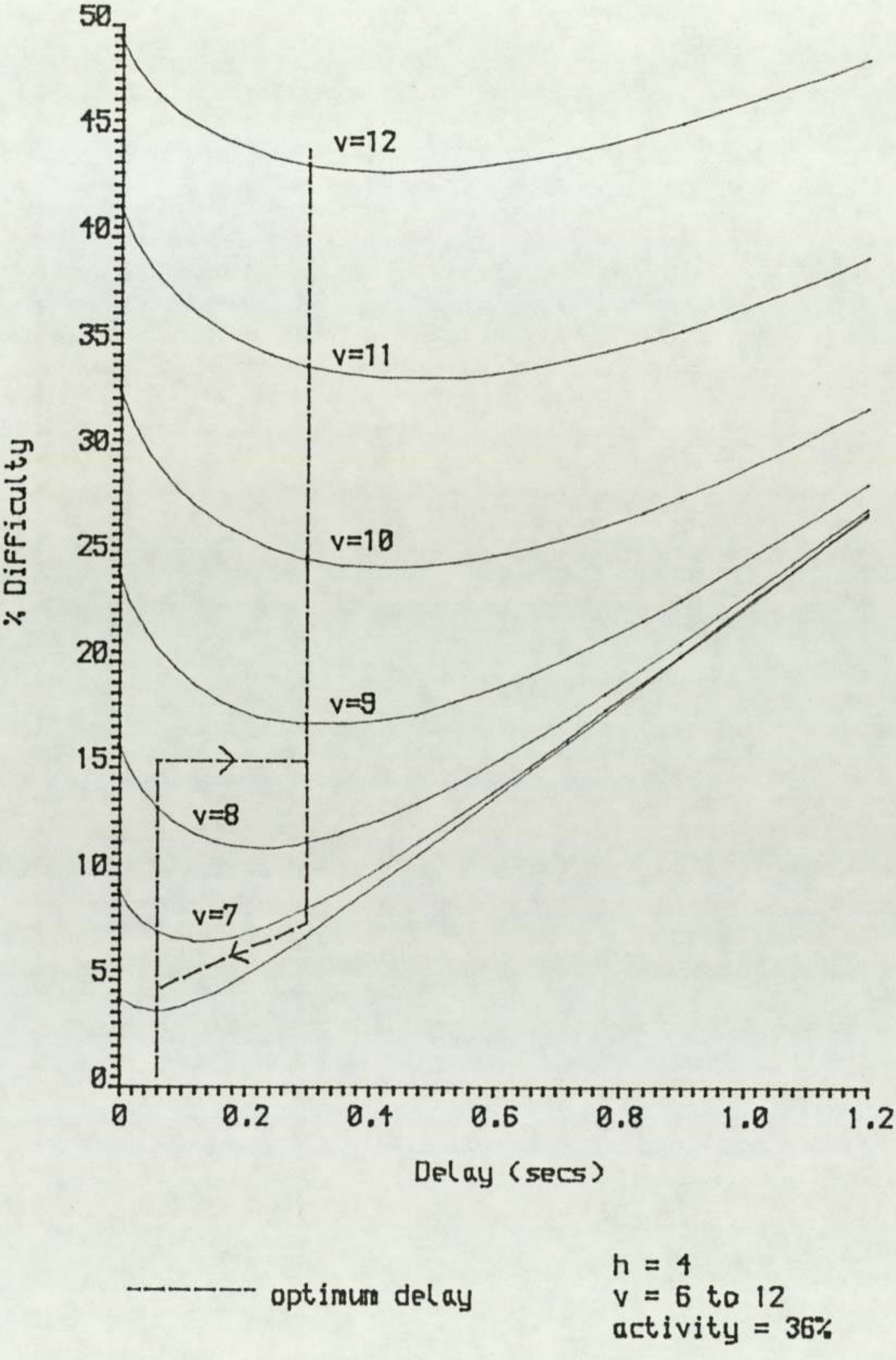


Fig. 4.3 (a) % difficulty versus delay for an activity of 36%

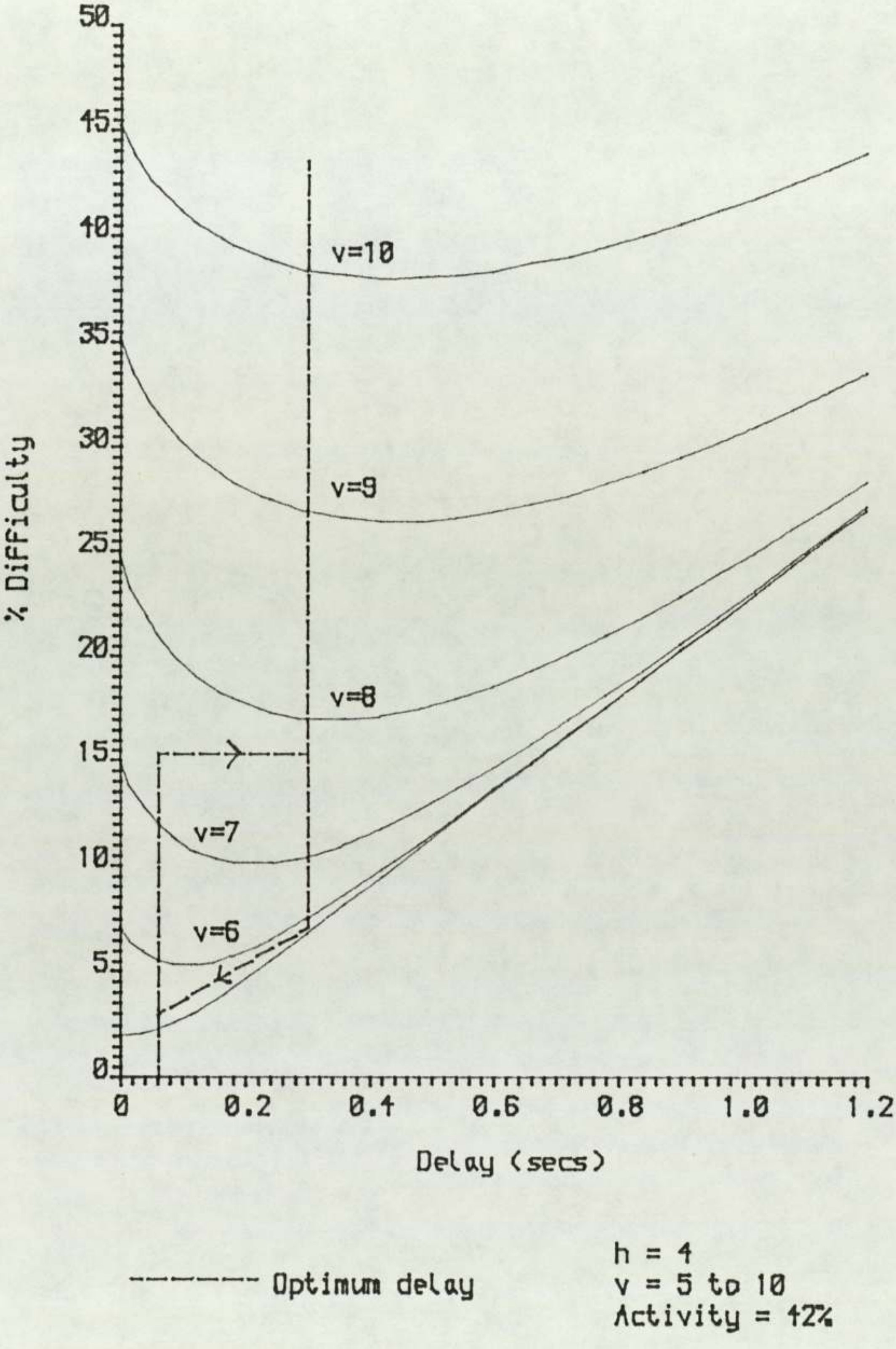


Fig. 4.3 (b) % difficulty versus delay for an activity of 42%

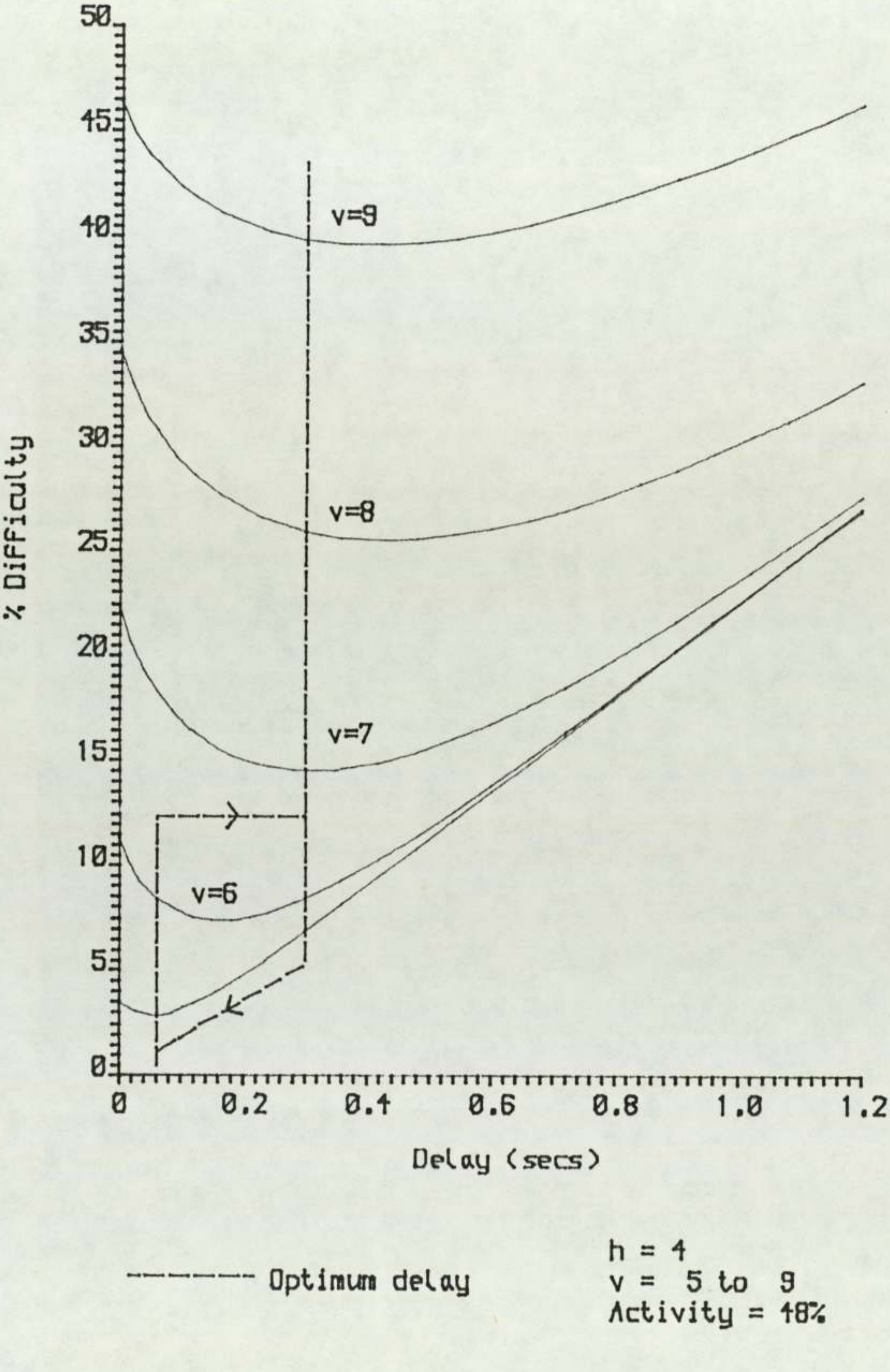
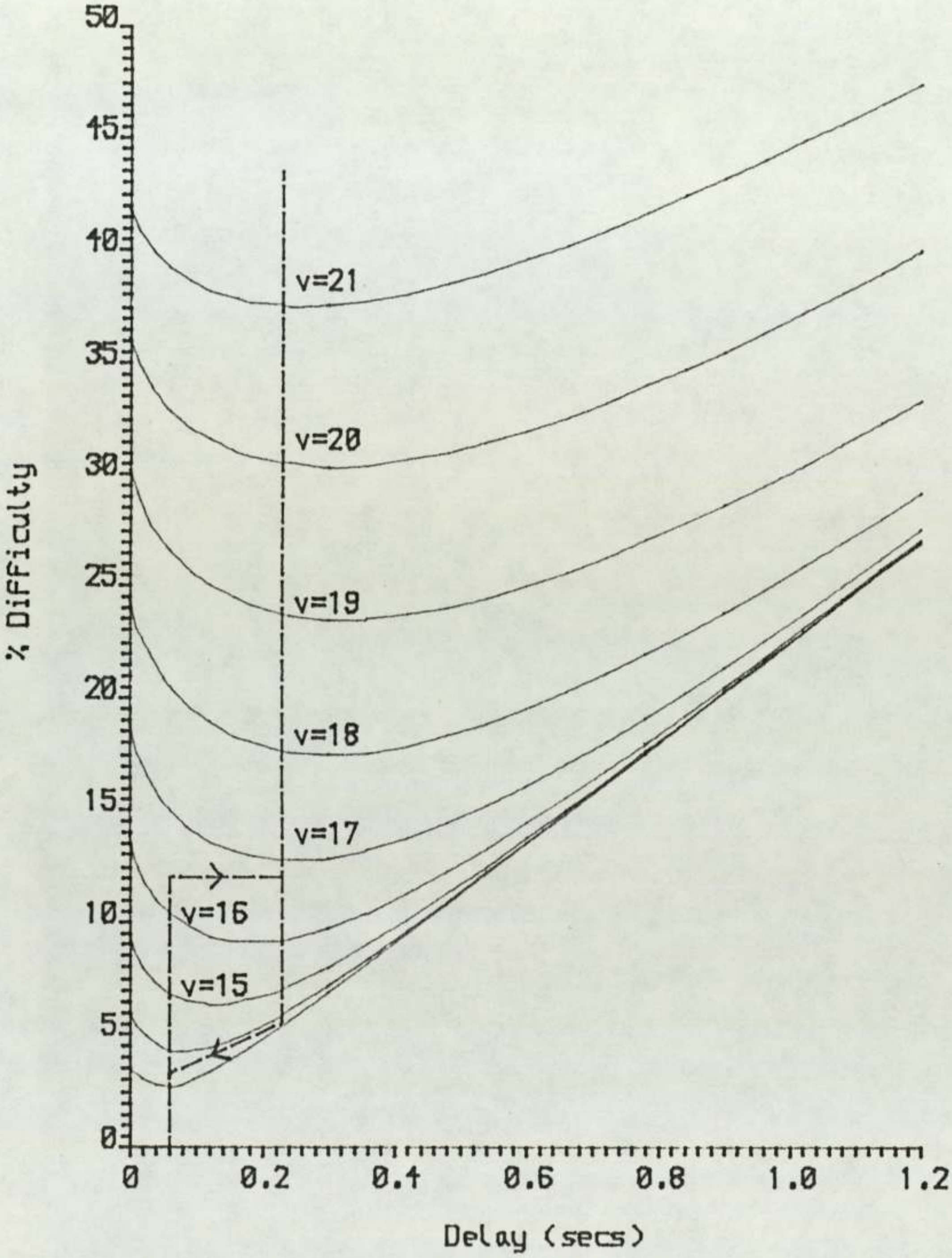


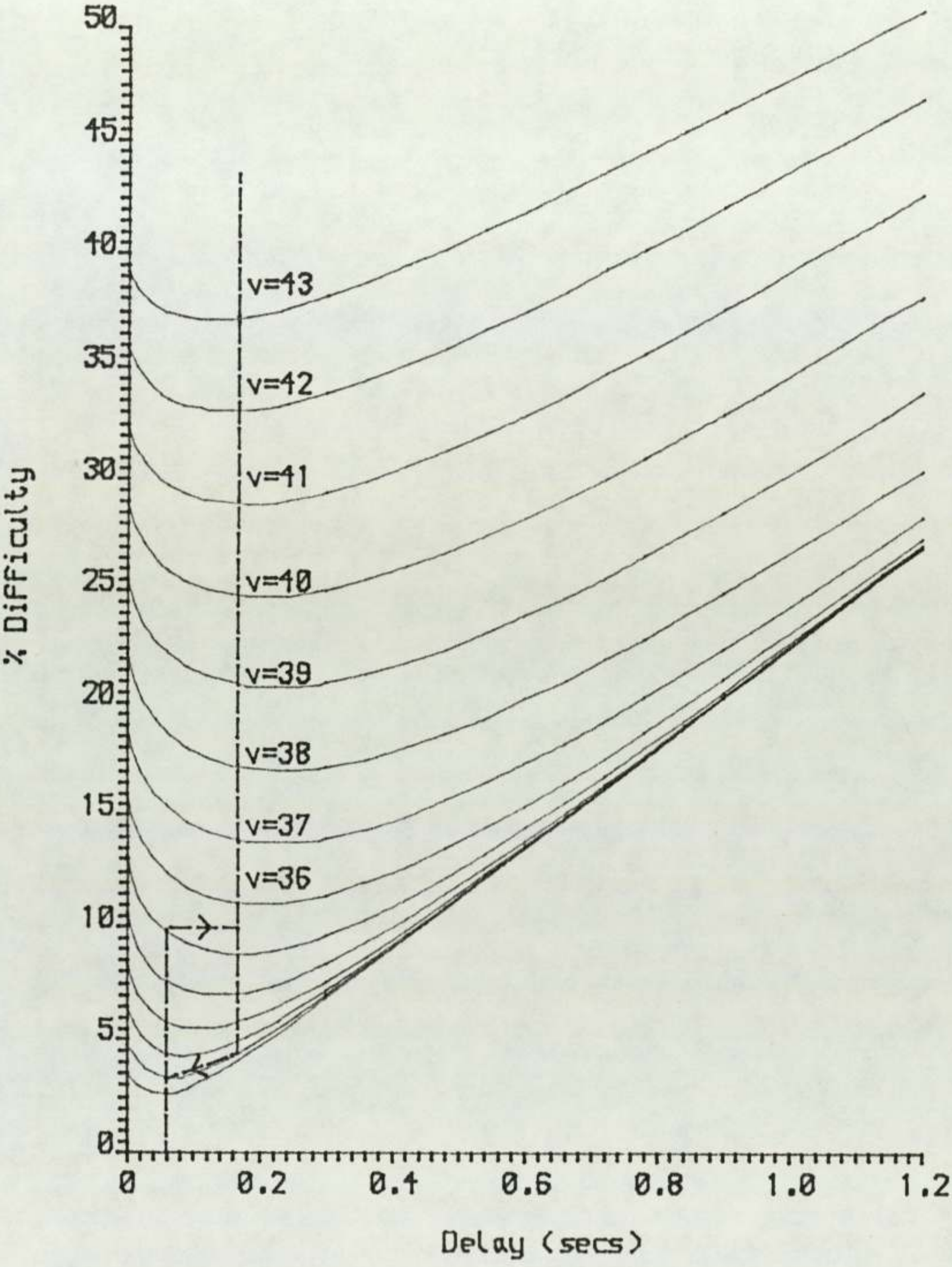
Fig. 4.3 (c) % difficulty versus delay for an activity of 48%



------ Optimum delay

h = 8
v = 13 to 21
Activity = 42%

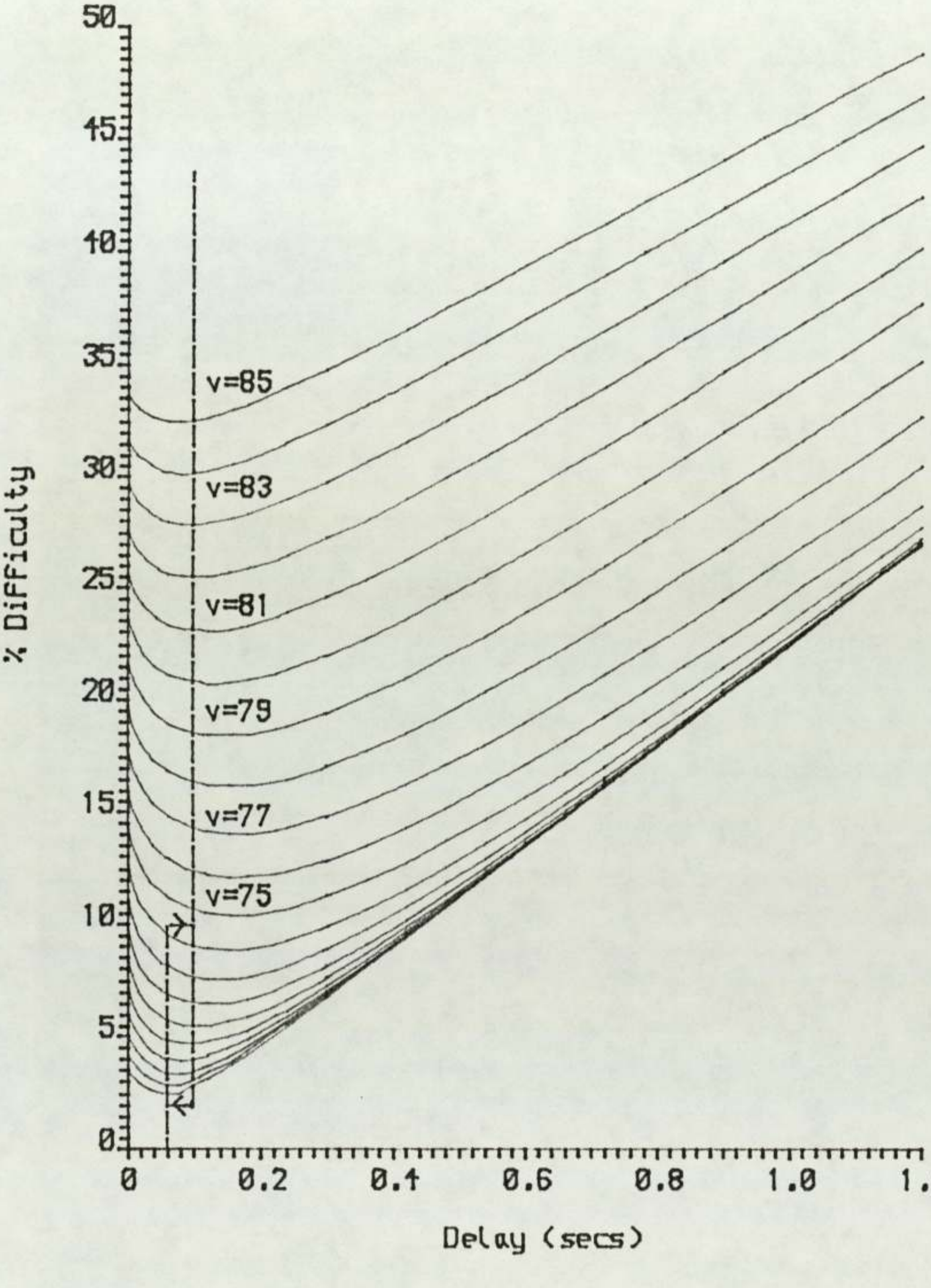
Fig. 4.4 (a) % difficulty versus delay for a channel capacity of 8



------ Optimum delay

h = 16
v = 30 to 43
activity = 42%

Fig. 4.4 (b) % difficulty versus delay for a channel capacity of 16



------ Optimum delay

h = 32
v = 66 to 85
Activity = 42%

Fig. 4.4 (c) % difficulty versus delay for a channel capacity of 32

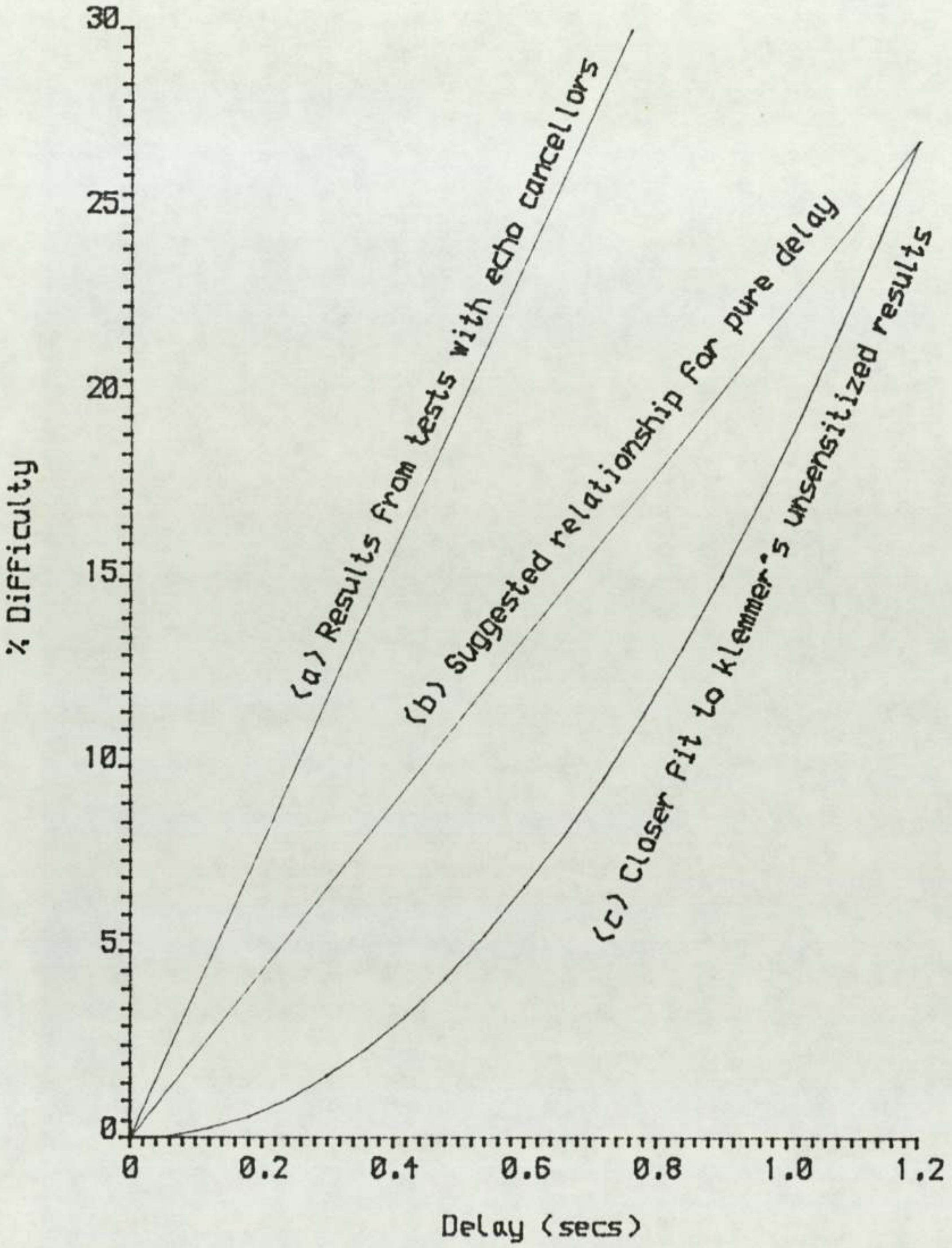


Fig. 4.5 Alternative relationships between difficulty and delay

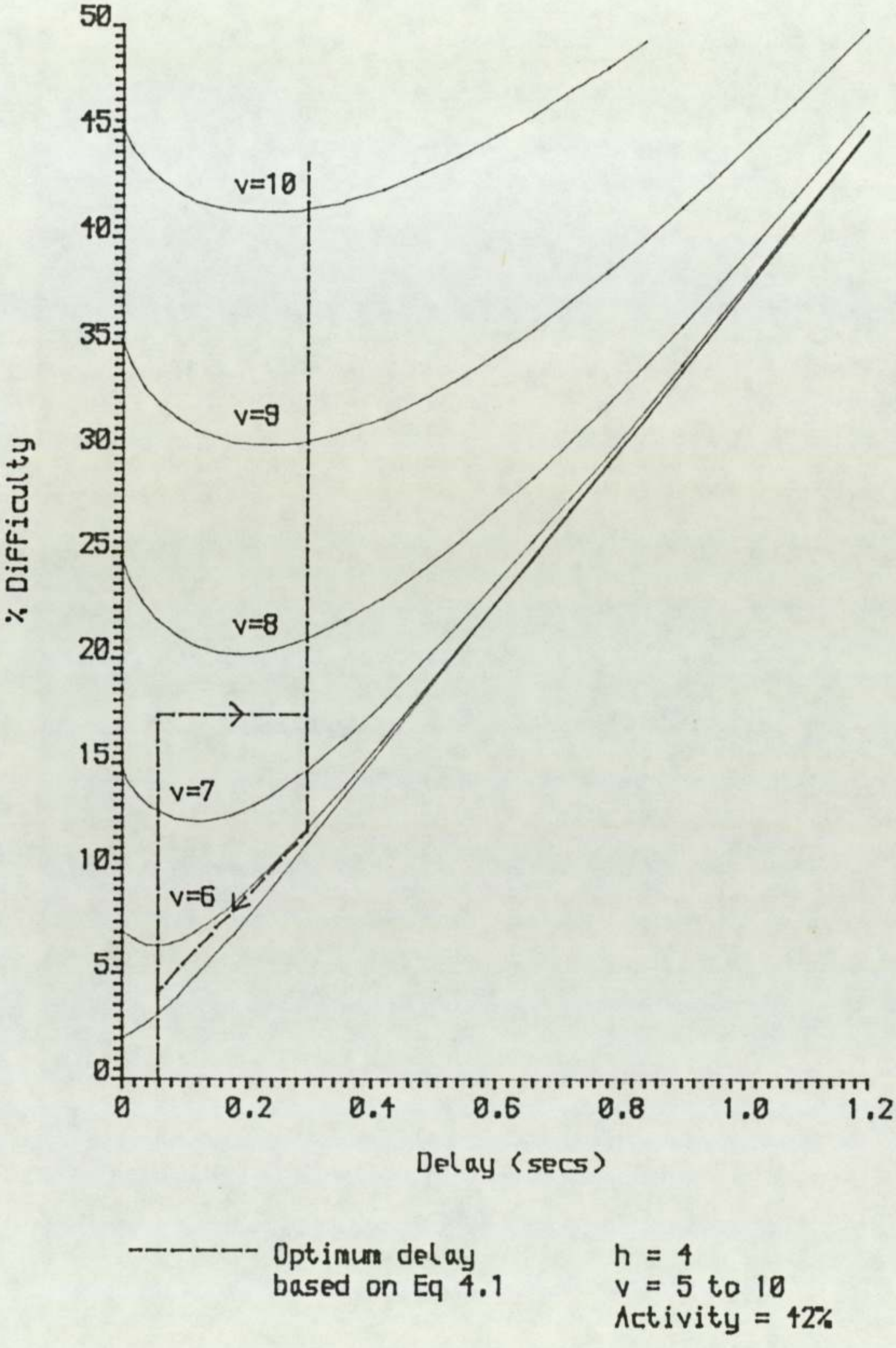


Fig. 4.6 (a) % difficulty versus delay based on Eq 4.4

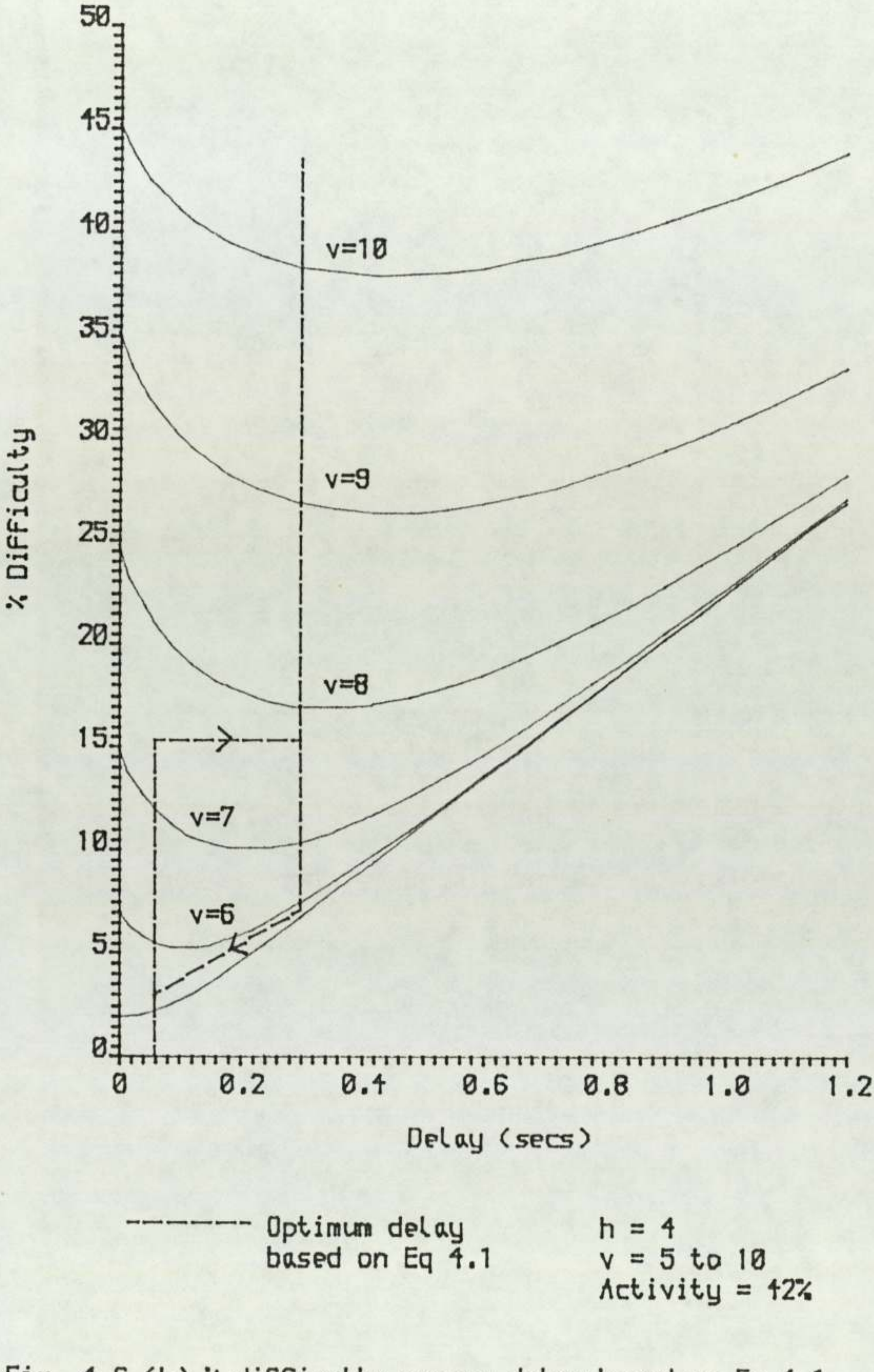


Fig. 4.6 (b) % difficulty versus delay based on Eq 4.1

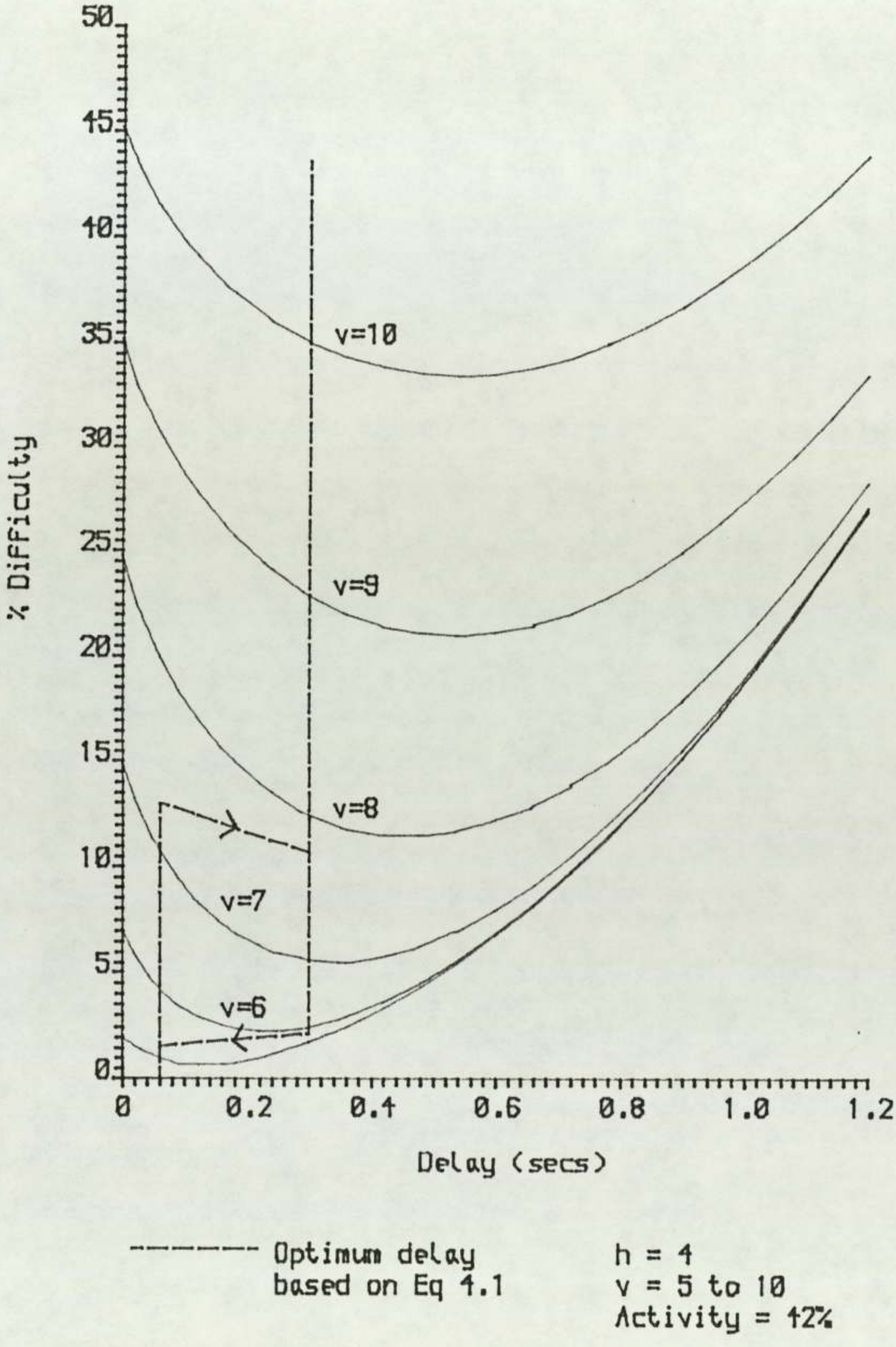


Fig. 4.6 (c) % difficulty versus delay based on Eq 4.5

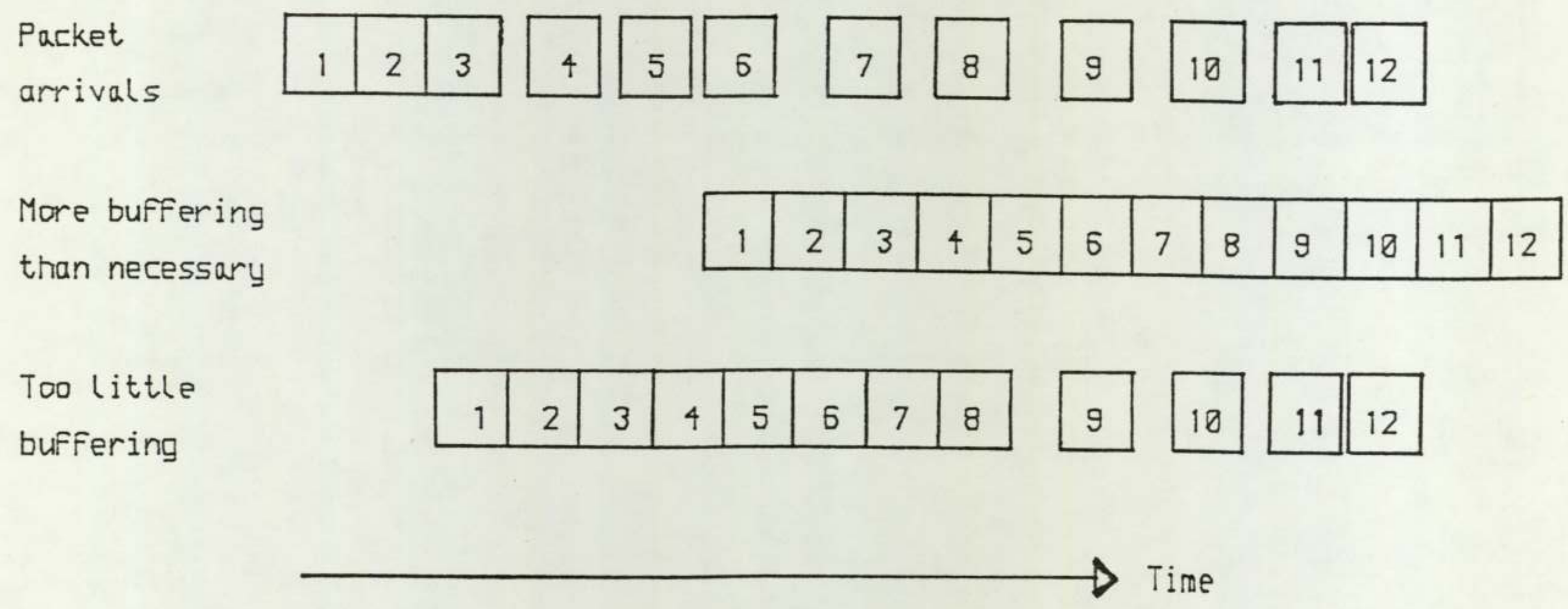


Fig. 4.7 An illustration of the receiver buffering problem

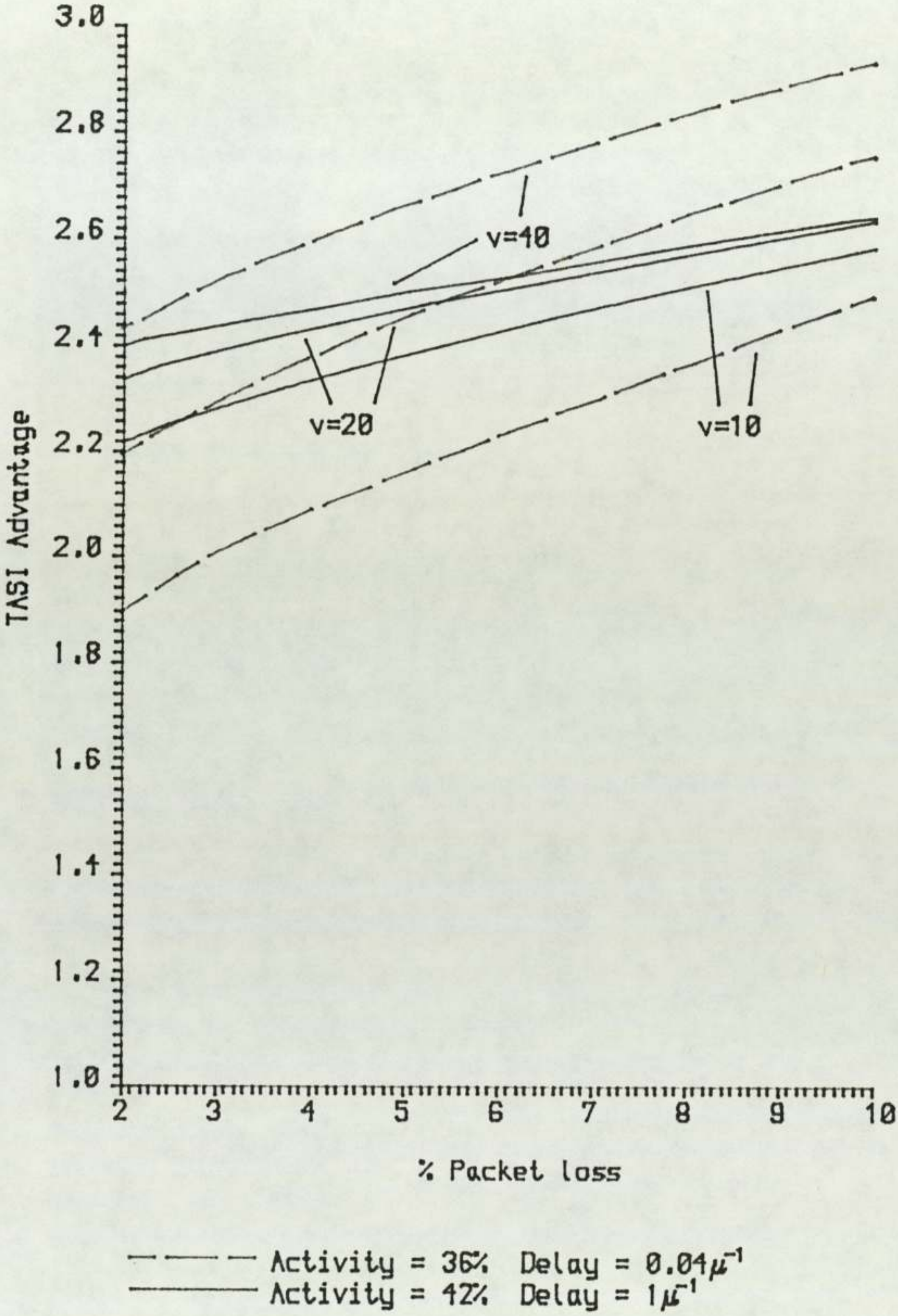


Fig. 4.8 Performance comparison between an accurate speech detector used on a link with very little delay, and a speech detector using hangover used on a link with a large delay

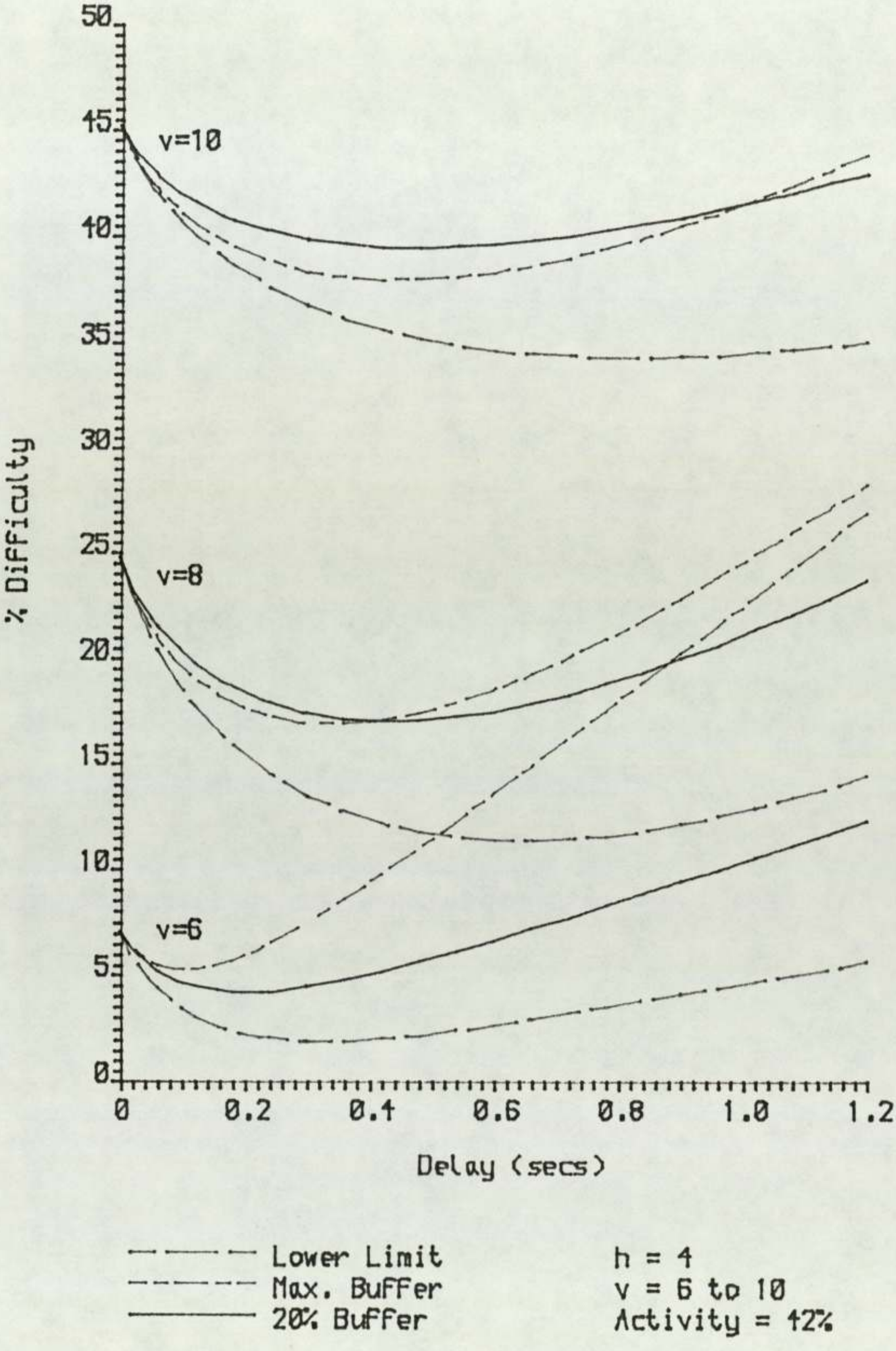


Fig. 4.9 (a) Receiver buffer performance for a channel capacity of 4

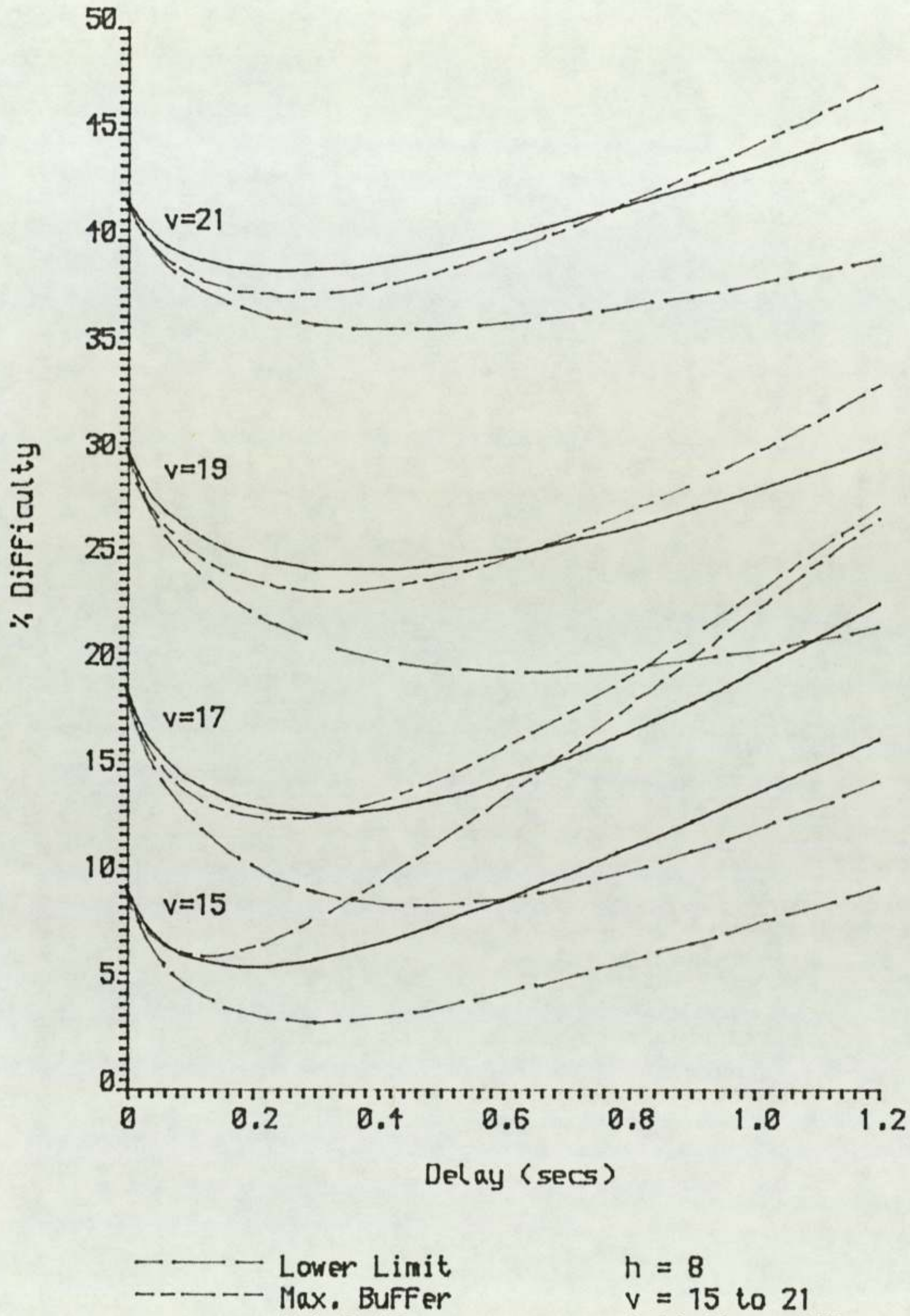


Fig. 4.9 (b) Receiver buffer performance for a channel capacity of 8

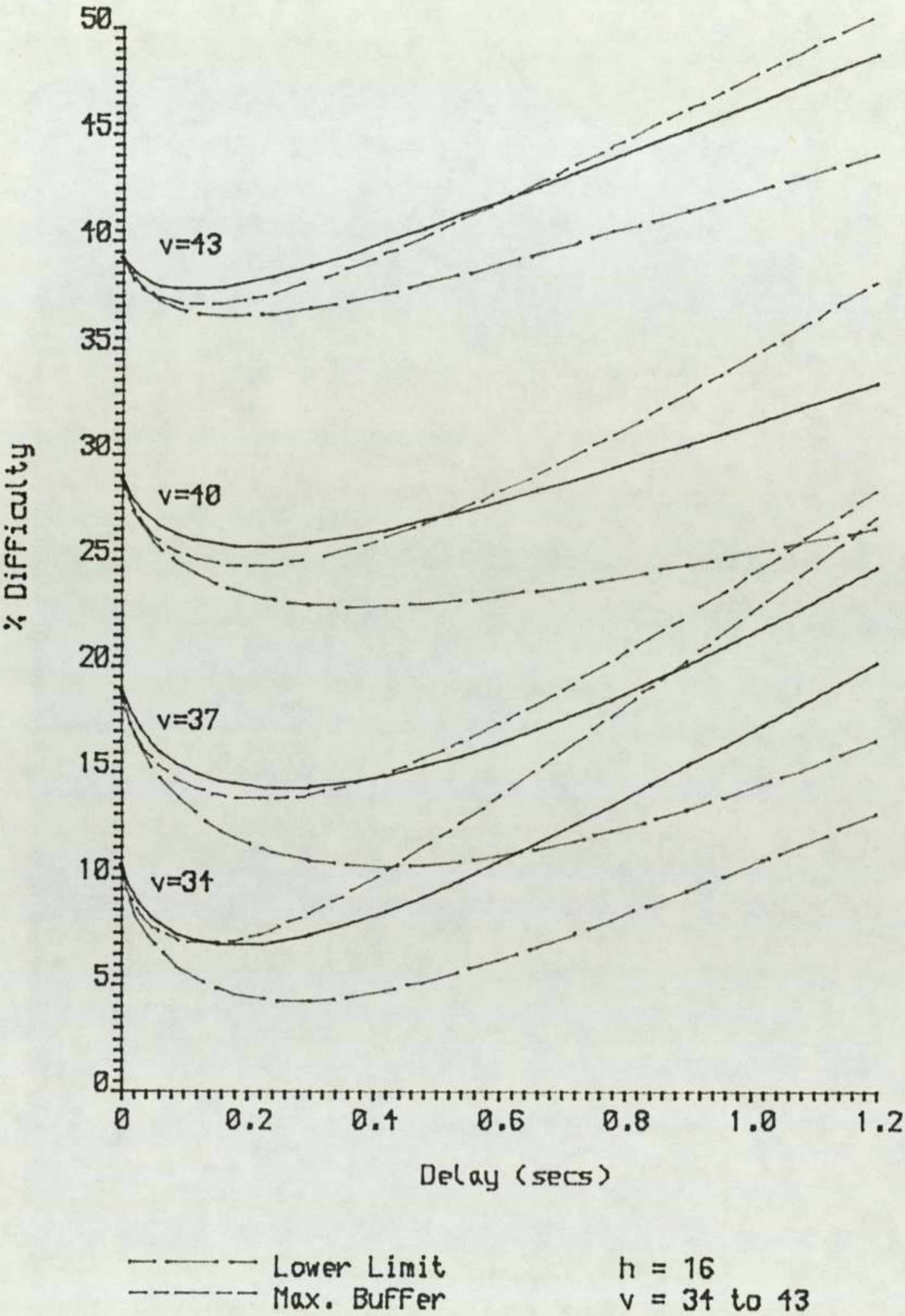


Fig. 4.9 (c) Receiver buffer performance for a channel capacity of 16

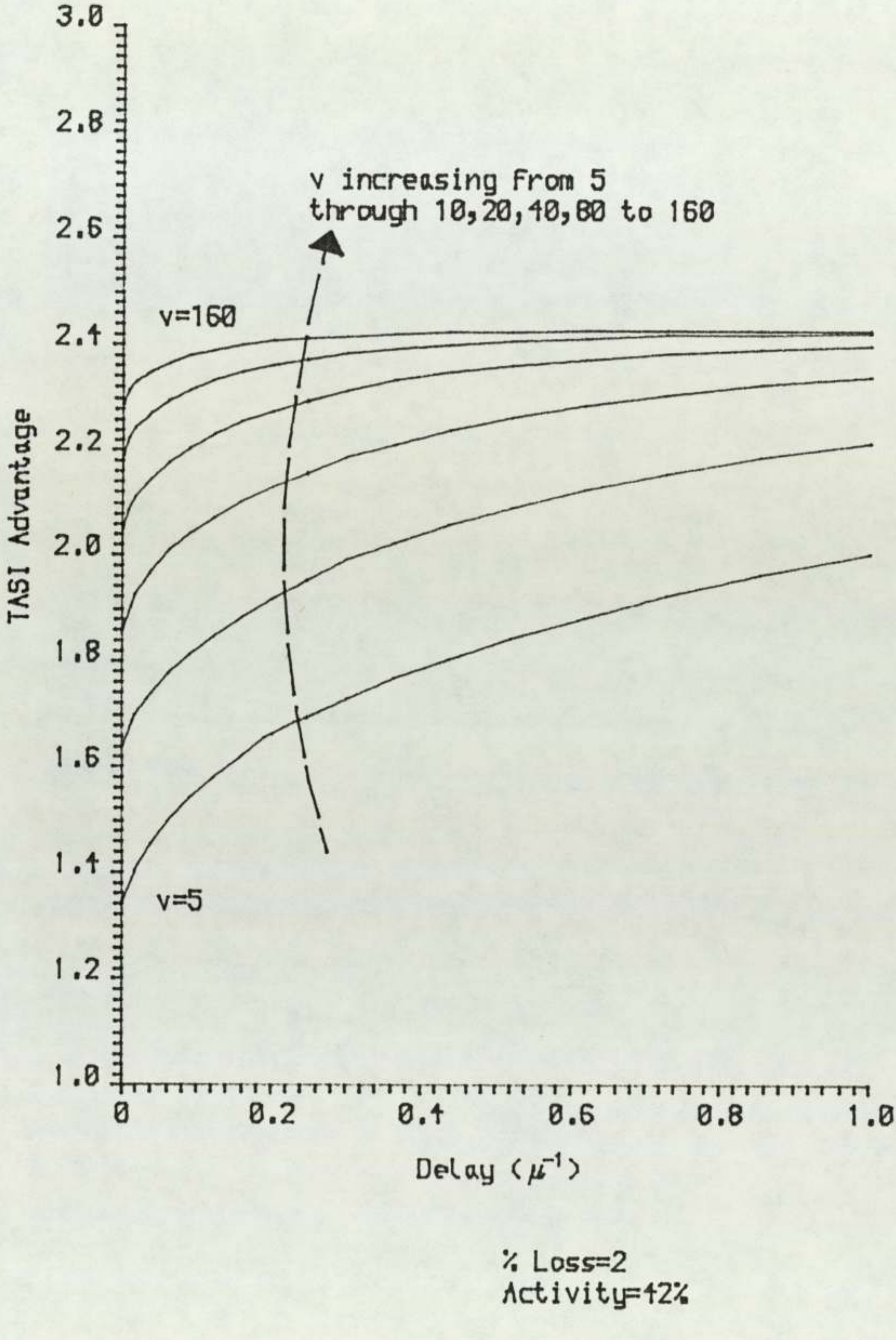


Fig. 4.10 (a) The trade-off between TASI advantage and delay at 2% packet loss

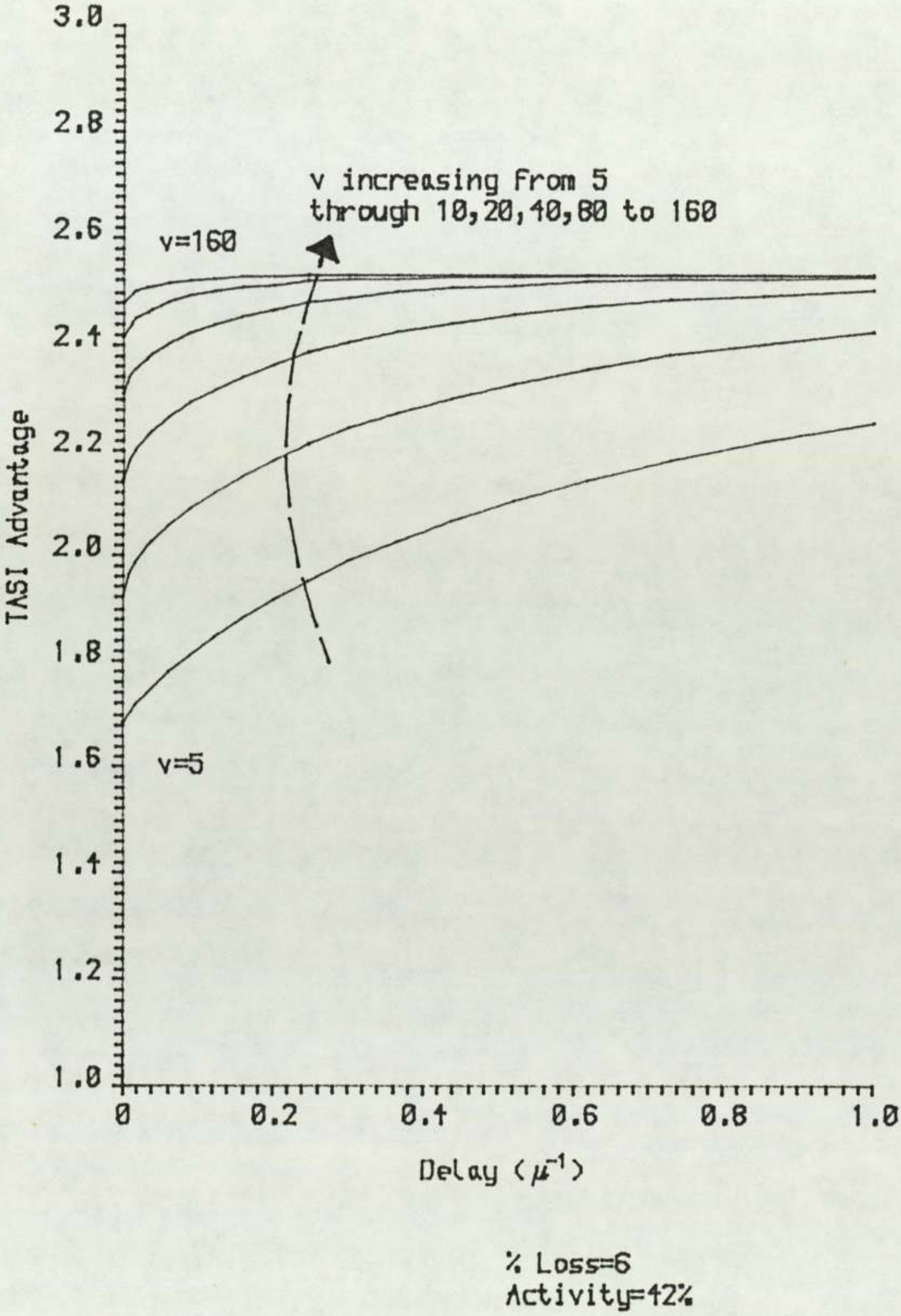


Fig. 4.10 (b) The trade-off between TASI advantage and delay at 6% packet loss

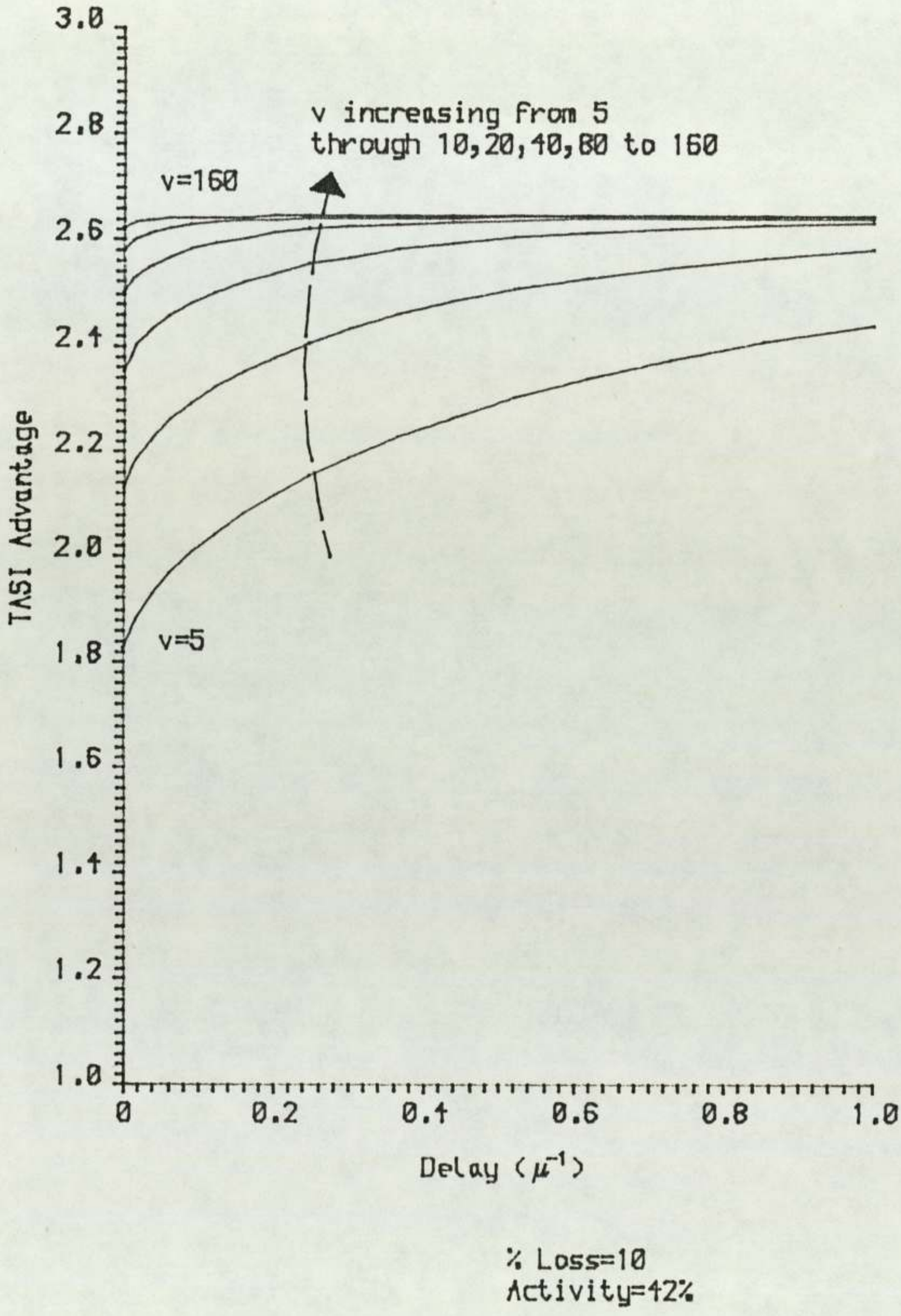


Fig. 4.10 (c) The trade-off between TASI advantage and delay at 10% packet loss

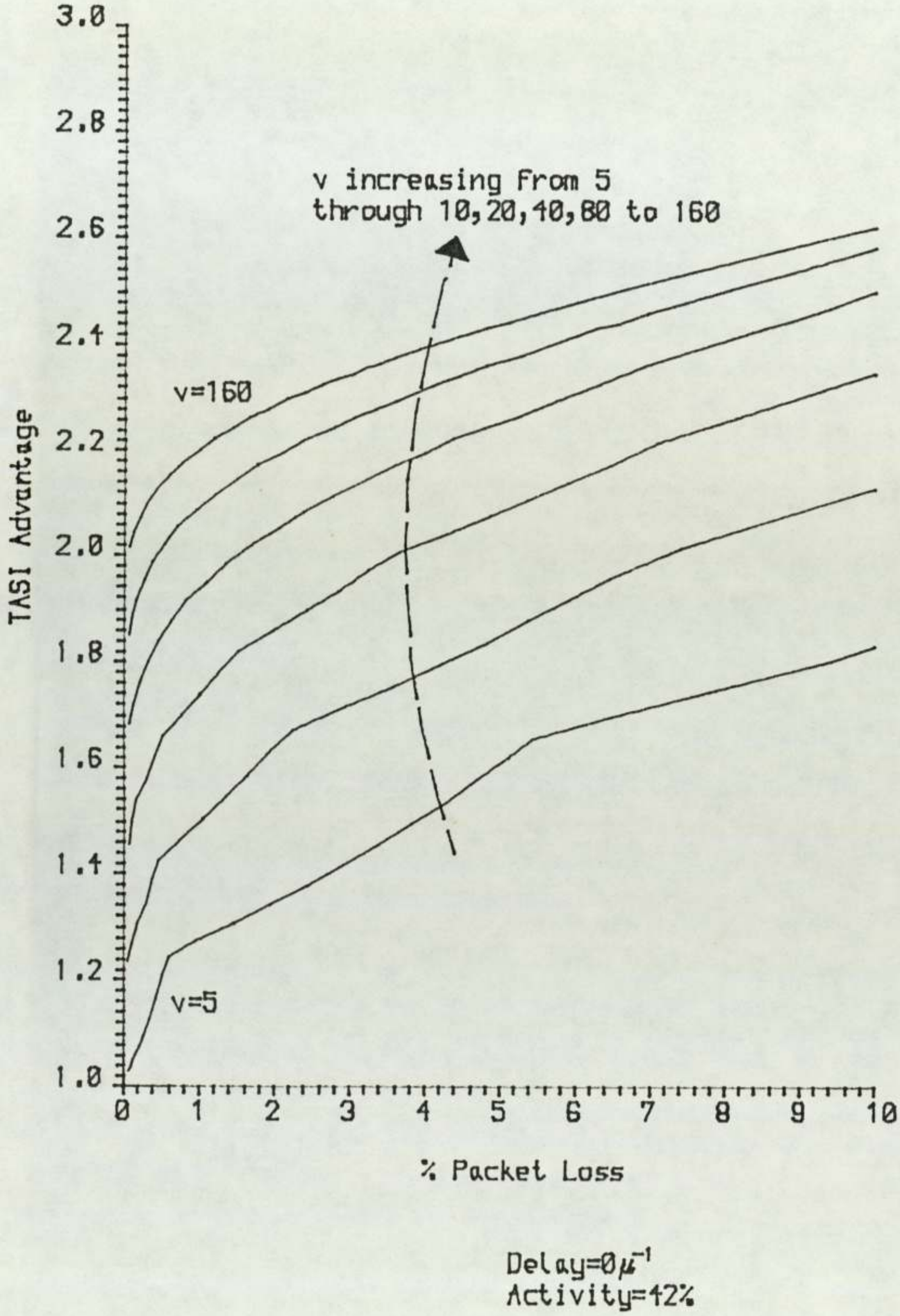


Fig. 4.11 (a) The trade-off between TASI advantage and packet loss at no delay

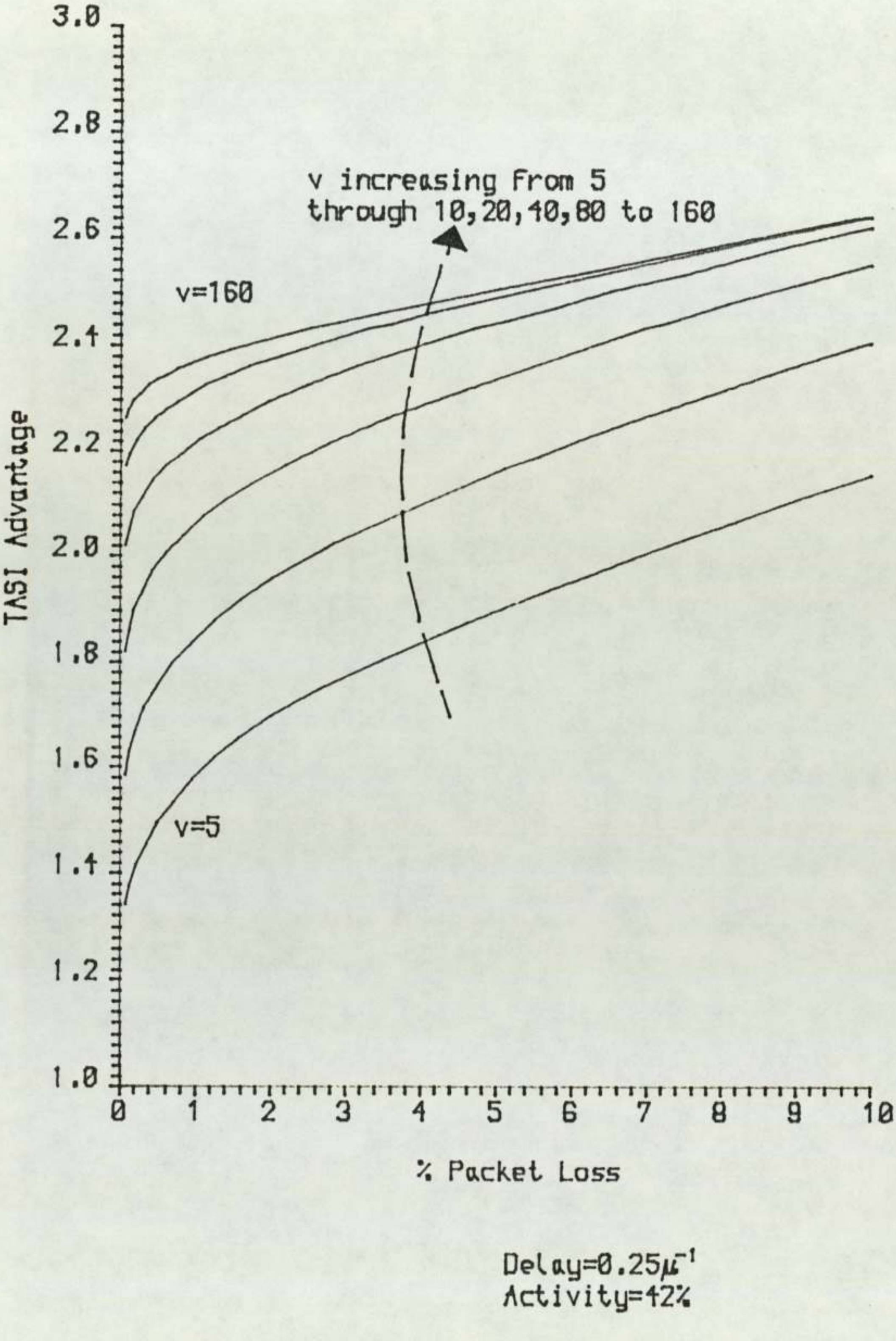


Fig. 4.11 (b) The trade-off between TASI advantage and packet loss at $0.25\mu^{-1}$ delay

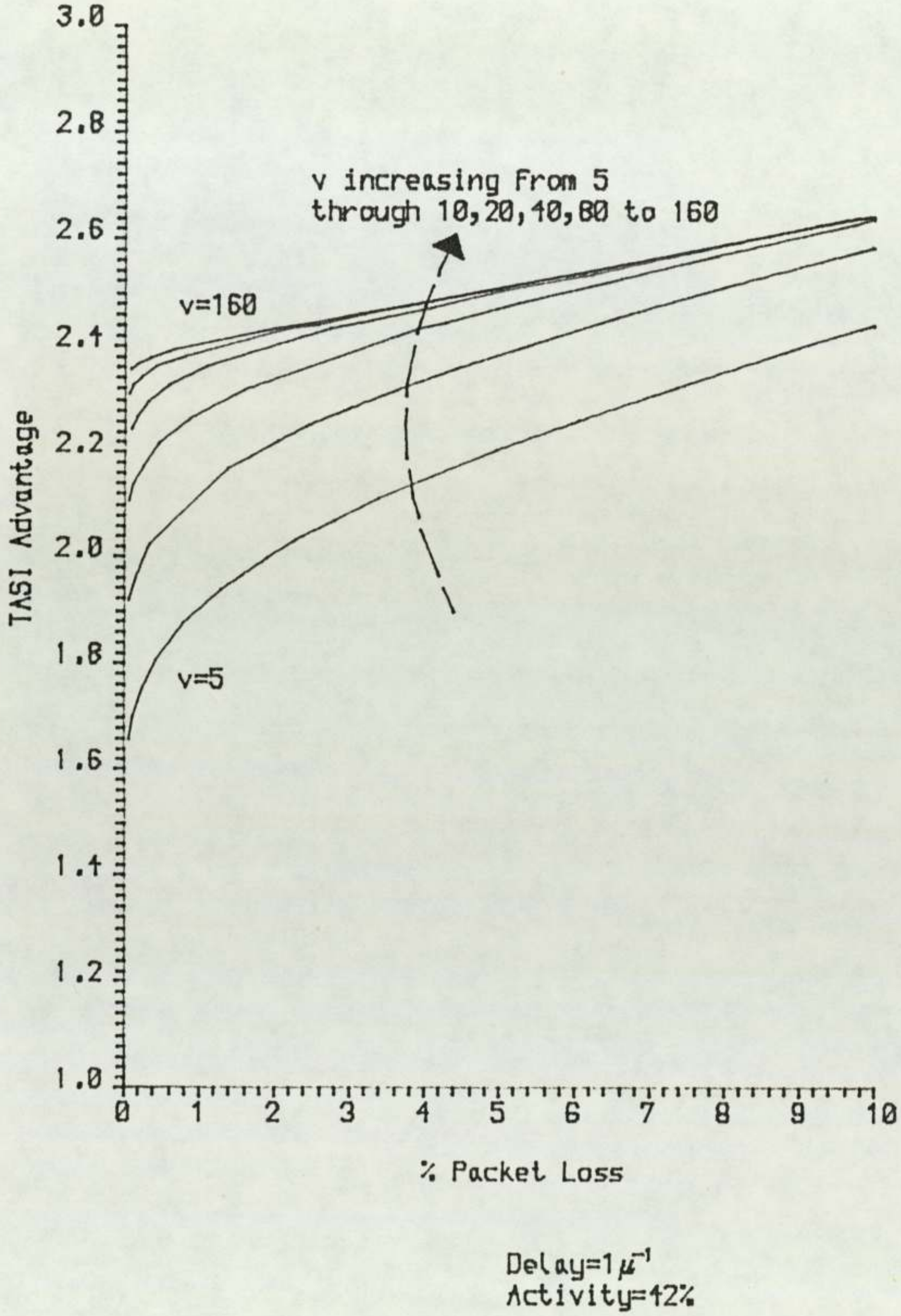


Fig. 4.11 (c) The trade-off between TASI advantage and packet loss at $1\mu^{-1}$ delay

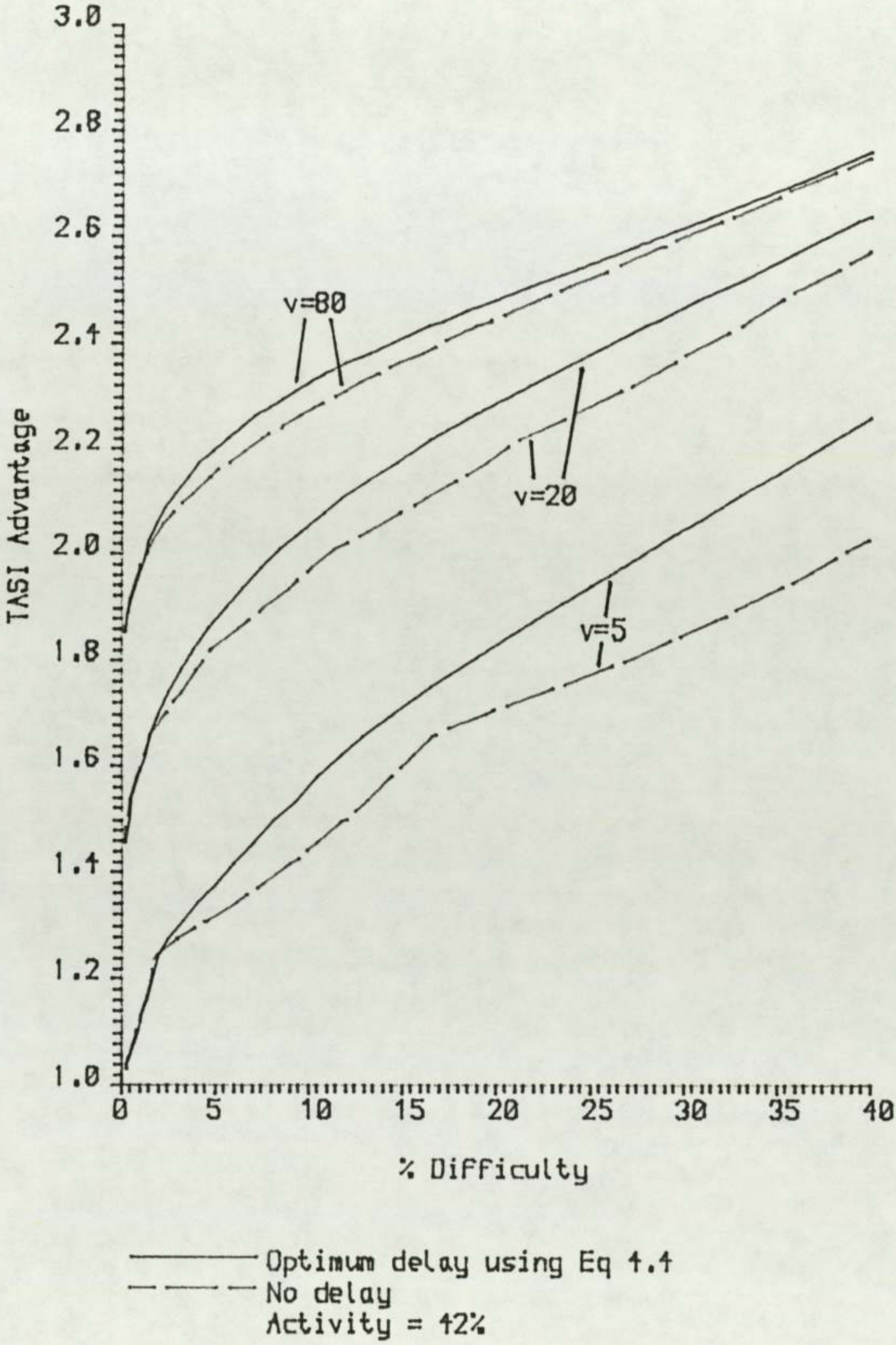


Fig. 4.12 (a) The TASI advantage obtained using the optimum delay based on Eq 4.4 and no delay

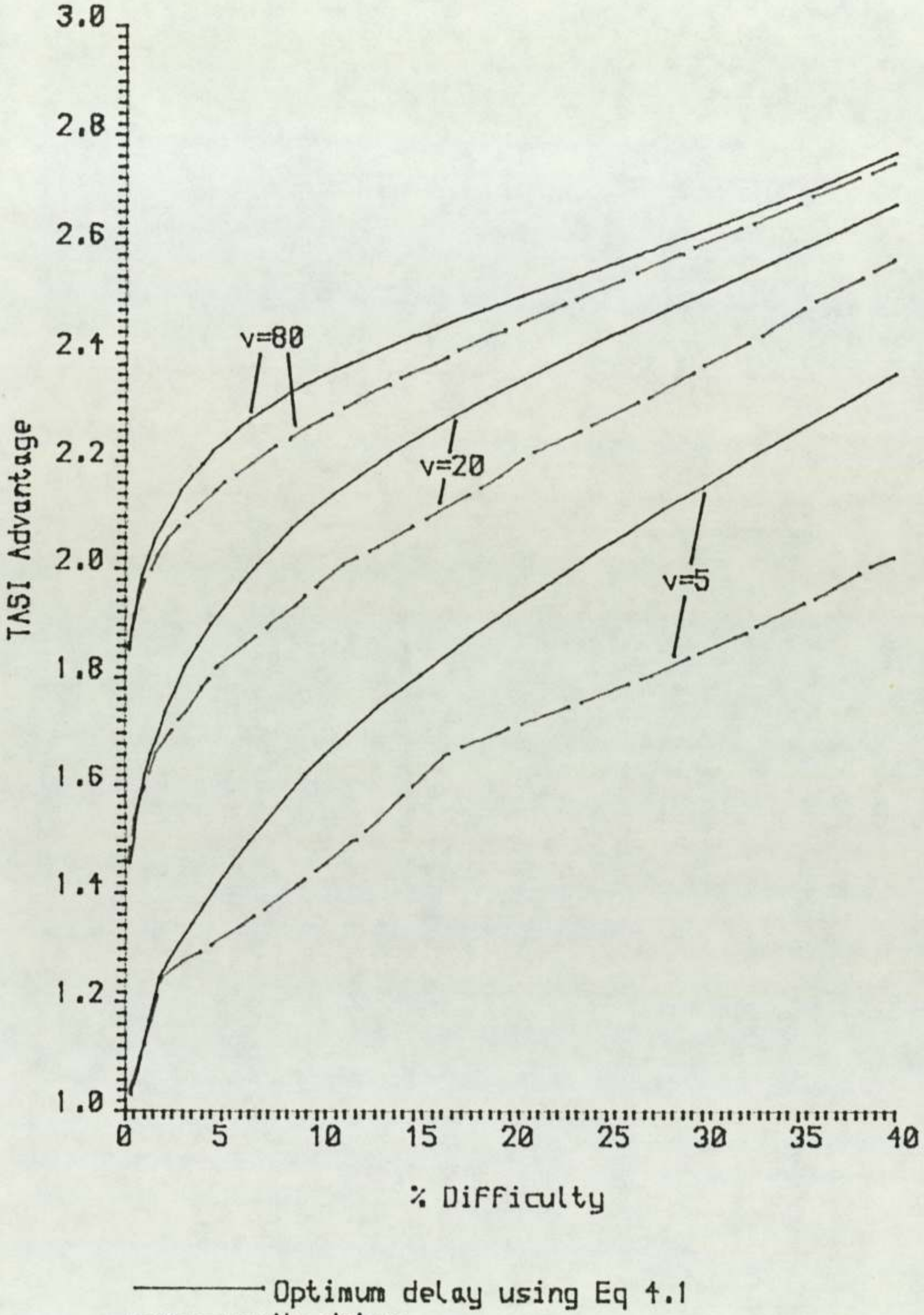


Fig. 4.12 (b) The TASI advantage obtained using the optimum delay based on Eq 4.1 and no delay

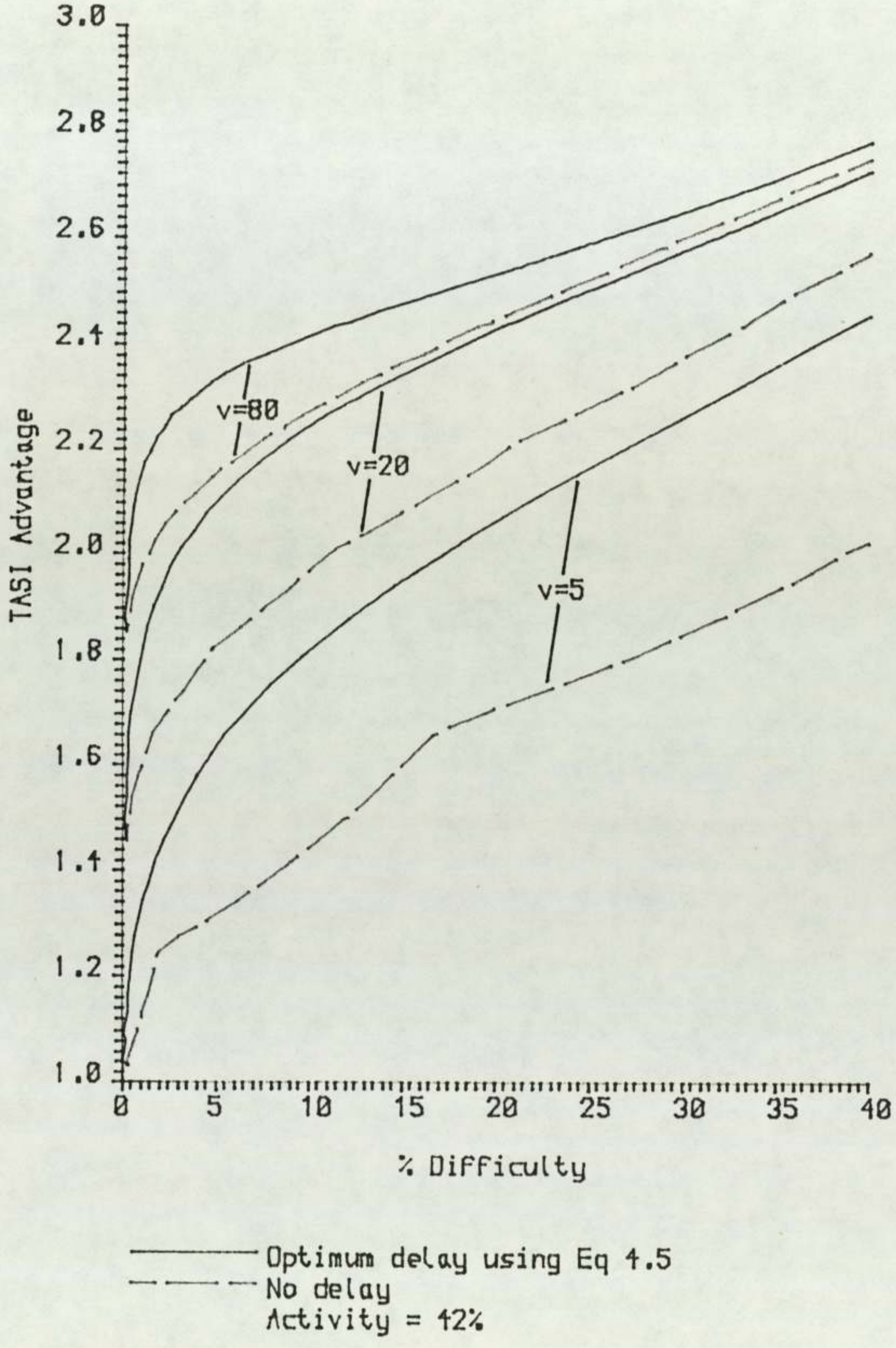


Fig. 4.12 (c) The TASI advantage obtained using the optimum delay based on Eq 4.5 and no delay

CHAPTER FIVE

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

5.1 CONCLUSIONS

The ways in which channel capacity can be used efficiently with packet-switched speech have been investigated. In order to do this, an analytical model of the queueing process has been developed, and extensive tests on the effects of packet losses have been carried out. Together, these provide a powerful tool for the performance assessment of a packet-switched speech link.

The analytical model is based on a two-dimensional Markov chain. The main difference between this and other models of a speech-packet queue is that the queue length is modelled as a continuous variable. This considerably reduces the number of state equations, which considerably reduces the CPU time needed for each calculation. In practice, this has meant that computer time has not restricted the range of parameters investigated.

Three different types of queue have been modelled. The first is an unlimited queue, which, although a practical impossibility, is useful as an approximation to a queue with an extremely large amount of buffering available. This particular queue cannot be analysed with a discrete queueing model.

The second type of queue modelled is one where packets are lost when the queue reaches a certain pre-determined limit. The important parameter to be calculated for this queue is the percentage of packets that are lost. It is this queue that has been used for both the simulation of packet losses and the investigation of how a packet-switched speech link should be operated.

The third type of queue modelled was one where packets are shortened before the queue reaches its limit. This can only happen if packets are coded using embedded encoding. This queue was not considered in the subsequent chapters simply because its operation is very dependent on the type of embedded encoding, and speech coding has not been considered in this work.

Simulations using data from real conversations proved that the analysis is accurate for a wide range of parameters. The continuous approximation only affects results when the queue limit is very much less than the packet length.

Approximate empirical relationships for the delay distribution of an unlimited queue and packet loss in a limited queue were found. These give a quick method of finding approximate values of delay and loss, and give an appreciation of how various factors affect the queueing delay and packet loss.

Subjective tests on the effects of packet losses attempted to clarify inconsistencies in the literature, and to obtain reliable results on the acceptability of losses. Due to the difficulty of taking into account the way the losses occur (i.e. in bursts), only a limited number and range of parameters could be tested.

A combined intelligibility/opinion test showed that it is better to fill a gap by repeating the previous packet rather than by inserting silence. It also showed that the shorter the packet the better, although the improvement between 16 ms and 8 ms was not very significant.

A listening test showed that packet losses up to 20% allowed speech to be understood with only moderate effort. It also showed that the way that losses are distributed does not have a large effect on opinion scores.

Finally, a conversation test on packet losses up to 25% produced a reliable assessment of difficulty with packet losses and reliable opinion scores. It was found that 2.5% loss was acceptable to almost everyone, 10% loss was acceptable two thirds of the subjects and 25% loss was acceptable to one third of the subjects. It was also found that 20dB loudness loss tends to mask the effect of losses up to 10%, and that 20dB loudness loss and 25% packet loss are roughly equivalent degradations.

Using the analysis and the results of the tests, several principles regarding the operation of a packet-switched speech network were discovered. It was found that the mechanism for losing packets once a queue has reached its limit is a very important part of a packet-switched speech system. The mechanism has to be capable of ensuring that losses are spread over all channels as evenly as possible. Without such a mechanism the capacity of a link is considerably reduced.

It appears that two modes of queue operation are required. In the "light-traffic" mode very little packet loss is needed. As this would be the normal mode, the quality of most connections would be very good. Peaks in the number of calls can be handled by a queue entering a "heavy-traffic" mode, where some queueing delay and packet loss is permitted. To some extent, call blocking is not needed in the heavy-traffic mode, since speech quality can be reduced instead. However, if the packet loss rose above 20% or so, further call requests would have to be blocked. The best queueing limit for the heavy-traffic mode is a function of channel size, but a wide range of limits will give quality very close to optimum.

The question of optimum packet length was considered. Packet length is not a problem when speech packets from different channels are grouped together into a large packet, sharing a common header, since the individual packets can be short without a large header overhead. However, when each packet has to have its own header, the optimum packet length depends on the load on each link and, in practice, a compromise must be reached between the optima at light and at heavy loads. If embedded encoding is not used, the variation in the acceptability of losses as packet length changes must also be taken into account.

The problems of how best to buffer packets that have passed through a link in heavy-traffic mode were investigated. It appears that buffering each packet to the maximum delay, which is the simplest buffering technique, is unlikely to be improved on by more sophisticated techniques. However, a detailed study of different techniques is required to verify this.

A short consideration of the needs of data traffic showed that, in many cases, it would be best to allocate some of the capacity of each link to data traffic. This would ensure a sensible balance between the performance of the network for speech and for data. Also, speech calls would have to be blocked if the data traffic was already using most of the idle capacity of a link.

Finally, it was seen that for networks carrying small amounts of speech traffic (up to about 20 channels per link) significantly less channel capacity is needed with packet-switching than with other SI systems. For larger-capacity networks the difference is not very significant, because the effect of buffering is very small and the extra speech loss allowed makes only a small difference. However, if the potential advantge of using variable-rate coding is taken into account, the difference may be quite significant even for large-capacity networks.

5.2 SUGGESTIONS FOR FURTHER RESEARCH

A few areas that need further research have emerged. The variation in acceptability of packet losses with packet length needs to be known before definite conclusions on optimum packet length can be drawn. The problems of how best to buffer packets at the receiver need to be investigated in more detail. Also, the performance of a buffered-talkspurt queue employing packet loss needs to be found, to see if it has significant advantages over the usual first-come-first-served queue.

Perhaps the greatest need for further research lies in the area of speech coding. Although the usefulness of variable-rate coding has been shown for both waveform coders (10, 12, 13) and vocoders (11, 14), its full potential has not been exploited. The ideal speech coder would maintain a constant speech quality whilst varying the bit rate, keeping the average bit rate as low as possible. To do this would require a very flexible coder which might even switch between waveform coding and vocoding techniques as the signal changed.

Another technique that could be put to good use is that of priority queueing. This would enable hangover to be used at the beginning of a talkspurt as well as at the end, since extra packets can be sent in a burst once speech is detected, their high priority ensuring they reach the receiver in time to be played out. It could also be used to reduce delays for any packets that had to pass through more than one queue operating in heavy-traffic mode.

This work has only considered issues concerning the efficient use of transmission capacity. No study has been made of reliability, which is particularly important for military networks, nor of ease of multiplexing, which is particularly important for local area networks (LANs). Neither has any study been made of non-real time applications of packet-switched speech, such as voice messaging and voice annotation of text. Much work therefore remains to be done before the full potential of packet-switched speech is realized.

APPENDIX 1 DETAILS OF THE SUBJECTIVE TESTS

A1.1 Introduction

This appendix supplies details about the subjective tests that are left out of chapter 3 for the sake of brevity. These details mainly concern the preparation of each test, the subjects, the test structure, the results and an analysis of variance of the results using the statistical language GENSTAT(83). Aspects of the results that relate to the design of the test are discussed here, but aspects relating to the factors being tested are discussed in chapter 3.

A1.2 Preliminary Test

A1.2.1 Preparation of Test

108 sentences were used for the test, 54 read by a male and 54 by a female. Each sentence was stored in the computer, and losses carefully inserted into the four main words of the sentence. The losses were spaced as evenly as possible; e.g. 20% loss would mean the first of every five packets was lost. The sentence was then output onto tape 6 times, each time with a different packet length or fill-in technique. After 18 sentences, 9 male spoken and 9 female spoken, the pattern of losses given in table A1.1 was repeated. In all, 648 sentences were recorded.

These sentences were then divided into 6 tests each with 6 groups of 18 sentences as shown in table A1.2. The groups were arranged in pairs so that the subject could make a comparison between the loss types in each group of the pair. Each pair occurs twice, the second time the opposite way round to the first time. In each pair either the fill-in strategy or the packet length is kept constant.

The subjects were chosen to minimize the variation of the intelligibility scores and were all females aged 16 to 50 who did not use the telephone for business purposes (see table A1.3). It was assumed that this restriction would not affect the outcome of the test.

A1.2.2 Analysis of Results

The intelligibility scores, in terms of the number of words repeated incorrectly, and opinion scores for each half-group of sentences are shown in Table A1.4. The intelligibility scores are out of a total of 36. The opinion scores were only introduced into the test to help the subjects make their comparison, but have proved to be very useful.

Because of the way the tests are arranged, a separate analysis of variance is needed for each pair of loss types. The analysis for the length comparisons uses only the results when the two lengths are compared. The analysis for the fill-in strategy is able to use results from other comparisons without unbalancing the test, so twice as many scores are used.

Since an analysis of variance was performed on intelligibility and opinion scores for each of the nine comparisons, only a summary of the results is given here. Tables A1.5 and A1.6 give the significance (rounded up to not significant (NS), 25%, 20%, 5%, 2%, 1% and 0.1%) of each independent variable and interactions between variables in each test. The variables related to the test structure are labelled subjects, sex (of the speaker) and sex.sentence. Sentence cannot be used on its own since the sentences spoken by the male and female are different. The variable which is being tested is denoted "fill-in/length". The tables also give the mean scores and the standard error of difference for the two means. Using the degrees of freedom of the residual, the significance of the difference between the two means can be found, and is given in the bottom row. It ought to be the same as the fill-in/length row of the analysis of variance, but because of small differences it is sometimes one significance level lower.

For the analysis of variance of intelligibility scores, a square-root transform (i.e. score = $\sqrt{(\text{score} + 0.5)}$) was used since the scores are probably distributed with a Poisson distribution. It can be seen from Table A1.5 that, in spite of the efforts to eliminate variation due to subjects and sentences, in the majority of tests these variations were significant. It is interesting to note that the sex of the speaker had a significant effect in two of the fill-in tests, and a small effect in the length tests. (If more results had been available, the effect would probably have been more significant). Fig. A1.1 shows plots of the residual error against fitted values for each of the nine intelligibility analyses.

Generally, the error is reasonably constant over the range of scores, and so the results of the analysis are valid. No transform was used for the opinion scores, since the plots of residual error shown in Fig. A1.2 suggest that the error is already reasonably constant over the range of scores. The analysis of variance outlined in Table A1.6 shows that the effect of the subjects is now marginal, and the sex of the speaker and sentences themselves have no effect at all.

More significant results would have been obtained if a different test structure had been used. For instance, one alternative would have been to use a Latin square design, and make a comparison of the means of opinion scores rather than asking the subjects to make the comparison. Latin square designs were used in all subsequent tests.

A1.3 Listening Test

A.1.3.1 Preparation of test

For this test 120 sentences, taken from novels and magazines, were used. 60 were spoken by a male and 60 by a female. To produce realistic losses, a simulation of a packet-switched speech link was developed on a PDP 11/03 minicomputer. The program used the data collected for the simulations described in Section 2.8 to simulate a 21 channel link. 20 of the channels came from the recorded data and the 21st channel was real speech. The taped sentences were then played over the simulated link, and the output recorded. Six different links were used, details of which are given in Table 3.3.

The sentences were divided into groups of ten. These were arranged into two Latin squares, one for the male-voice sentences and one for the female-voice sentences, as shown in Table A1.7. The variation in amount of packet loss between groups of sentences played over the same link is shown in Table A1.8. The variations are very large; however, if more than ten sentences had been used, it would have been difficult for the subject to make an overall assessment of quality. Because the variations are caused by fluctuations in the amount of speech traffic (even though the number of callers remains fixed), peaks and troughs of loss level occur with the same group of sentences for all the links (groups 3 and 6 for the male voice, 4 and 3 for the female voice). It is therefore to be expected that the sentences will add a significant variability to the results as these also represent the variations in losses.

For this test the subjects were chosen from as wide a range of ages and professions as possible, and were of both sexes (15 females, 21 males). These details are given in Table A1.9.

A1.3.2 Analysis of results

Table A1.10 contains a complete listing of the results, and Table A1.11 gives an analysis of variance of them. The independent variables relating to the test are the same as in the preliminary test (SENT stands for "sentence"). The variable being tested is now labelled "LOSS". The final column indicates the significance level of the variance ratio, rounded up as in tables A1.5 and A1.6. From this it can be seen that all the test factors are significant.

The significance of the subjects can be explained by the fact that every person gives a different judgement of the same thing. However, it may also indicate that the different conditions under which the test was done affected the results. More care was taken with the test conditions in the conversation test.

The sex of the speaker is very significant. However, no hard and fast conclusions can be drawn, as the sentences used for each sex were different and the male voice always came first in the test. In the preliminary test the sex of the speaker did not significantly contribute to the variation in opinion scores (see Table A1.6); thus it is possible that the variation is caused solely by the test structure. However, a good balance of males and females was used in the conversation test to ensure that a dominance of one sex did not bias the results. The significant effect of the sentences is most likely caused by the variations in loss levels rather than the sentences themselves.

The plot of residual error versus fitted values in Fig. A1.3 shows the marked diamond shape that must be expected from a large number of opinion scores (because of the limited range of scores). The arcsin transform and various sigmoid transforms were tried, but the diamond shape always remained. This casts doubt on the validity of the analysis of variance and probably means that the confidence limits are even wider than given by the analysis.

A1.4 Conversation Test

A1.4.1 Preparation of Test

For this test a two-way real-time simulation of a packet-switched speech link was developed. Telephone sets were placed in two fairly quiet rooms at opposite ends of the building and each set was given the facility to ring the other one when permitted to by a central control switch. Each set was given sidetone at 1.5 dB below the normal listening level of the telephone. White noise bandlimited to 3.5 kHz with a level of -77 dBmp was injected just before the receiver to mask room, circuit and quantisation noise. It also masked the effect of the speech detector.

Each earpiece and microphone was connected to the speech interface using screened cable. Dynamic microphones with the same frequency response as carbon microphones, but with much greater reliability, were used. A switchable 20 dB attenuator was connected into each speech path between the computer interface and each earpiece, but before the noise was injected.

The sensitivity of the speech path was measured using standard Bruel and Kjaer equipment as follows. First, a 94 dB (relative to 20 $\mu N/m^2$) sound level calibrator Type 4230 was used to calibrate a precision sound level meter Type 2203. Then an artificial voice Type 4219 driven by a sine generator Type 1023 was positioned in a telephone test head Type 4905. The microphone in the mouth opening of the artificial voice was connected to a measuring amplifier Type

2607, which sent a feedback signal to the sine generator. This arrangement ensured that the sound pressure generated by the artificial voice was constant for all frequencies.

Positioning the microphone of the precision sound level meter at 25 mm from the lip ring of the artificial voice, the output of the sine generator was adjusted until a pressure of 89.3 dB was generated by the voice. A telephone headset was then placed in the test head, in the British modal speaking position (BMSP)[34] but not at an angle, since a carbon microphone was not used and the angle was therefore irrelevant. The frequency of the sine generator was changed from 100 Hz to 1 KHz in steps of 100 Hz and then from 1 KHz to 4 KHz in steps of 200 Hz, and the signal going to the earpiece of the telephone was measured. This gave the gain/frequency plot shown in Fig. A1.4 (a).

The earpiece of the telephone handset was then sealed against an artificial ear type 4153 and the output of the artificial ear connected to the precision sound level meter. Using a sine wave input to the earpiece, the gain/frequency characteristic of the earpiece was measured, and is shown in Fig. A1.4 (b). The two frequency responses can be combined to give the overall response shown in Fig. A1.4 (c).

From these gain/frequency responses, the overall reference equivalent (ORE) of the speech circuit was found (using the procedure outlined in reference (84)) to be 14 dB. With the 20 dB attenuator inserted, the ORE drops to 34 dB. These two OREs are approximately the ideal and the worst connection likely to be encountered in the public network[34].

The simulation program simulated two links carrying 21 calls each and, as before, had a variable queue limit and channel capacity. As 200 ms hangover was needed for the speech detector (which was just a simple peak-level detector), 200 ms hangover was added to all the stored data. This meant that link capacities of 7, 9, 11 and 21 were needed to give average losses of 25%, 10%, 2.5% and 0 with 10 ms queueing limit. These four link capacities, each at two different levels of loudness loss, were then formed into two Latin squares as shown in table A1.12.

32 subjects were used in this test, 17 males and 15 females. 6 of the pairs were both males, 5 both females and 5 mixed. Details of the ages and professions of the subjects are given in table A1.13. On arrival, the subjects were taken straight to their respective rooms and given the brief shown in table A1.14. They were each supplied with 8 envelopes containing 6 postcard size reproductions of paintings. The pictures were the same for both subjects. They then filled in section 1 of the opinion form given in
table A1.15, and one subject rang the other. After arriving at a mutual order of preference they terminated the call and completed the rest of the form and then moved on to the next set of cards. They took it in turns to ring each other. Subjects from the university did the test in two halves on two separate days. Those who came from outside had just a few minutes break between the two

halves.

A1.4.2 Analysis of Results

The results of the conversation test are given in table A1.16. A few of the opinion scores are missing because of a fault in the speech interface. In the analysis of variance, GENSTAT estimates the missing values using an iterative procedure which reduces the residual for the missing value to zero. However, the results shown in Figs. 3.6 to 3.9 have been calculated only from the valid opinion scores.

The analysis of variance of these results is shown in table A1.17. The independent variables related to the test are the order of the conversations and the picture cards, labelled TEST, the room labelled ROOM, and the individual subjects, labelled ROOM.CONV. Subject pairs, labelled CONV (short for "conversation"), are not included in the model, as there is no reason to think that the opinion scores of the two subjects would be in any way related.

The percentage sum of squares column does not add up to 100% because the sum of squares are calculated using estimates of the missing values but are expressed as percentages of the total sum of squares before these estimates were made. The 4.2% discrepancy suggests that the analysis has not been very badly distorted by the presence of missing values. The Standard Error of Differences have not been adjusted for the missing values as this would be very complicated, and the adjustements would probably make very little difference to the SED[83].

The plot of residual error versus fitted values in Fig. A1.5 shows the diamond shape that is characteristic of opinion scores. Again, no transformation that was tried was successful in removing the diamond shape. The analysis of variance can therefore only be viewed as approximate.

From table A1.17 it is clear that neither the test structure nor the rooms affected the test. The sex of the speaker and variations in loss levels cannot be separated out as factors, but instead add to the significance of the subjects (ROOM.CONV). However, as these are variations that would be encountered in practice, this is reasonable.

The good cross-section of people used for subjects, the fact that test conditions did not affect results, and the use of a realistic simulation all indicate that the results of this final test, presented in Figs. 3.6 to 3.9, are reliable.

TABLE A1.1 Distribution of Losses for Preliminary Test

Sentence	Word 1		Word 2		Word 3		Word 4	
	% loss	burst length	% loss	burst length	% loss	burst length	% loss	burst length
1	40	s	30	w	20	s	30	s
2	40	w	40	s	30	s	20	w
3	30	w	30	s	20	w	20	s
4	30	s	40	w	40	s	20	w
5	20	s	40	w	30	w	40	s
6	30	w	20	s	40	w	30	s
7	40	s	20	w	30	s	40	w
8	20	w	40	s	20	s	30	w
9	40	w	30	w	20	w	20	s
10	20	s	30	s	30	w	40	w
11	30	s	20	s	40	s	30	w
12	40	w	30	w	20	w	40	s
13	40	s	40	w	20	s	30	s
14	30	w	40	s	30	s	20	w
15	20	w	30	s	40	w	20	s
16	30	s	30	w	40	s	40	w
17	20	s	20	w	30	w	40	s
18	20	w	20	s	40	w	20	w

s 160 ms randomly placed in the word
w loss lasts for the entire word

TABLE A1.2 Structure of Preliminary Test

Test no.	Subjects	Pair 1		Pair 2		Pair 3	
		1	2	1	2	1	2
1	1 & 2	8s	8r	16s	8s	32r	8r
2	3 & 4	8r	8s	16r	8r	32s	8s
3	5 & 6	16s	32s	8r	16r	16s	16r
4	7 & 8	32s	16s	32r	32s	8r	32r
5	9 & 10	16r	32r	8s	16s	16r	16s
6	11 & 12	32r	16r	32s	32r	8s	32s

8s - 8 ms packets, silence in gaps
16s - 16 ms "
32s - 32 ms "
8r - 8 ms packets, repeated packets in gaps
16r - 16 ms "
32r - 32 ms "

TABLE A1.3 Details of Subjects of Preliminary Test

Subject	Sex	Profession	Age
1	F	Housewife	25
2	F	Housewife	40
3	F	Dressmaker	42
4	F	Clerical	19
5	F	Housewife	40
6	F	Housewife	50
7	F	Housewife	23
8	F	Nurse	21
9	F	Housewife	26
10	F	Housewife	45
11	F	Housewife	38
12	F	Clerical	16

TABLE A1.4 Results of Preliminary Test

Test 1

loss type / Subject	sex	8s		8r		16s		8s		32r		8r	
		M	F	M	F	M	F	M	F	M	F	M	F
1	int	2	0	4	1	4	3	2	6	4	5	0	0
2	int	6	6	6	7	9	5	5	14	7	7	1	5
1	op	3	4	4	3	2	2	3	3	4	3	4	4
2	op	2	3	3	3	2	3	4	2	3	4	4	4

Test 2

loss type / Subject	sex	8r		8s		16r		8r		32s		8s	
		M	F	M	F	M	F	M	F	M	F	M	F
3	int	1	1	5	1	4	3	3	4	6	7	5	5
4	int	1	6	5	7	3	3	4	3	7	8	5	2
3	op	4	4	3	3	3.5	3.5	3.5	3	1	2	2	2
4	op	3	2	2	2	3	3	3	3	2	2.5	2.5	2.5

Test 3

loss type / Subject \ sex	16s		32s		8r		16r		16s		16r	
	M	F	M	F	M	F	M	F	M	F	M	F
5 int	5	11	8	5	4	1	3	4	4	7	0	5
6 int	7	8	11	12	6	4	6	9	6	5	0	8
5 op	3	2	2	1	4	3	3	3	2	2	4	4
6 op	2.5	2	2	2	3.5	4	4	4	3	3.5	4.5	4

Test 4

loss type / Subject \ sex	32s		16s		32r		32s		8r		32r	
	M	F	M	F	M	F	M	F	M	F	M	F
7 int	10	15	11	5	9	7	11	15	5	5	3	6
8 int	13	14	11	2	3	6	9	7	2	3	3	4
7 op	2	3	2	3	3	3	2	2	4	3	3	3
8 op	2	2.5	3	3	4	3.5	3	3	4.5	4.5	5	4

Test 5

loss type		16r		32r		8s		16s		16r		16s	
Subject	sex	M	F	M	F	M	F	M	F	M	F	M	F
9	int	0	3	3	3	3	0	3	5	1	0	3	5
10	int	4	7	6	3	8	4	5	10	5	6	10	8
9	op	5	4	2	2	2	3	2	2	3	3.5	2	1.5
10	op	2	3	3	3.5	2	3	3	1.5	4.5	3	3	2.5

Test 6

loss type		32r		16r		32s		32r		8s		32s	
Subject	sex	M	F	M	F	M	F	M	F	M	F	M	F
11	int	7	10	6	2	12	7	5	9	5	4	4	12
12	int	8	13	3	3	12	11	6	10	5	5	4	7
11	op	3	3	4	3	1	2	2	2	3	3	2	2
12	op	3	2	2.5	2	1	1.5	3	2.5	2.5	2	2	2

TABLE A1.5 Summary of Analysis of Variance of Intelligibility Scores from Preliminary Test

Source of variation \ 1 2	8s 8r	16s 16r	32s 32r	16s 32s	16r 32r	8s 16s	8r 16r	8s 32s	8r 32r	No. tests sig <5%
Subject	0.1	1	NS	NS	5	1	5	NS	5	6
Sex	NS	1	1	20	NS	NS	NS	NS	20	2
Subject.sex	5	NS	NS	NS	NS	NS	NS	NS	NS	1
Sex.sentence	5	NS	1	20	5	1	20	NS	5	5
Fill-in/length	5	1	0.1	NS	5	NS	NS	20	5	5
Mean 1	2.184	2.546	3.251	2.88	1.914	2.237	2.031	2.22	1.641	
Mean 2	1.755	1.922	2.652	3.24	2.555	2.403	2.193	2.68	2.295	
SED	0.156	0.182	0.133	0.243	0.184	0.125	0.125	0.246	0.164	
DF of Residual	17	17	17	5	5	5	5	5	5	
Significance of the difference of means	2	1	0.1	20	2	25	25	20	2	

TABLE A1.6 Summary of Analysis of Variance of Opinion Scores from Preliminary Test

Source of variation \ 1 2	8s 8r	16s 16r	32s 32r	16s 32s	16r 32r	8s 16s	8r 16r	8s 32s	8r 32r	No. tests sig <5%
Subject	20	NS	1	20	NS	NS	20	20	20	1
Sex	NS	NS	NS	NS	NS	NS	NS	NS	NS	–
Subject.sex	NS	NS	NS	20	NS	NS	NS	NS	NS	–
Sex.sentence	NS	NS	NS	NS	NS	20	NS	NS	NS	–
Fill-in/length	1	0.1	0.1	5	NS	20	NS	5	20	5
Mean 1	2.69	2.34	2.062	2.562	3.19	2.75	3.375	2.437	4.0	
Mean 2	3.41	3.66	3.062	2.062	2.69	2.19	3.375	1.937	3.62	
SED	0.227	0.276	0.182	0.185	0.556	0.281	0.185	0.177	0.237	
DF of Residual	17	17	17	5	5	5	5	5	5	
Significance of the difference of means	1	0.1	0.1	5	NS	20	NS	5	20	

TABLE A1.7 Design of Listening Test

MALE VOICE

Subjects Group no	1 to 6	7 to 12	13 to 18	19 to 24	25 to 30	31 to 36
1	C	A	E	B	D	F
2	A	B	D	C	F	E
3	D	E	F	A	B	C
4	B	C	A	F	E	D
5	E	F	C	D	A	B
6	F	D	B	E	C	A

FEMALE VOICE

Subjects Group no	1 to 6	7 to 12	13 to 18	19 to 24	25 to 30	31 to 36
1	F	B	D	E	A	C
2	E	C	F	D	B	A
3	C	A	B	F	E	D
4	D	F	E	A	C	B
5	B	D	A	C	F	E
6	A	E	C	B	D	F

A	9	channels,	8 ms	queuing limit
B	8	" ,	256 ms	"
C	8	" ,	8 ms	"
D	7	" ,	768 ms	"
E	7	" ,	8 ms	"
F	6	" ,	2 sec	"

TABLE A1.8 Percentage Packet Losses for each Group of Sentences

Group no. / Link	MALE VOICE						FEMALE VOICE					
	1	2	3	4	5	6	1	2	3	4	5	6
A	3.7	4.2	4.5	3.7	3.2	1.2	3.8	3.7	2.4	5.0	3.1	3.6
B	4.1	3.9	4.9	4.1	2.9	0.8	3.5	3.3	2.3	6.2	3.6	2.9
C	8.8	8.6	10.1	8.8	7.3	3.3	7.6	7.5	5.6	9.8	7.2	7.0
D	10.5	9.0	12.3	8.1	4.5	3.2	5.5	5.8	3.9	11.9	7.1	7.1
E	15.9	15.2	17.8	15.9	13.7	7.9	13.4	13.9	11.2	17.6	13.3	12.3
F	20.8	21.9	26.4	21.7	18.6	10.7	16.0	18.3	13.3	24.0	17.9	19.7

TABLE A1.9 Details of subjects for Listening Test

Subject no.	Sex	Profession	Age
1*	F	Housewife	26
2*	F	Housewife	38
3*	F	Housewife	50
4	F	Student	20
5	M	Technician	50
6	M	Technician	45
7	F	Housewife	70
8	M	At school	15
9	F	Housewife	35
10	M	Student	21
11	M	At school	17
12	M	Social worker	48
13	F	Housewife	68
14	M	Accountant	27
15*	F	Housewife	40
16	M	Schoolteacher	28
17	M	Professor	68
18	M	Student	20
19	F	Actress	21
20	F	Hairdresser	20
21*	F	Housewife	45
22	M	Graphic designer	35
23	M	Head teacher	40
24	M	Researcher	24
25*	F	Housewife	25
26	M	Student	20
27	M	Schoolteacher	40
28	M	Christian worker	22
29	M	Dentist	35
30	M	Postgraduate	24
31	M	Computer operator	25
32	F	Housewife	30
33*	F	Dressmaker	42
34	F	Student	20
35	M	Tax manager	45
36	M	Postgraduate	25

* Also took part in Preliminary Test

TABLE A1.10 Results of Listening Test

Subject No.	MALE VOICE						FEMALE VOICE					
	A	B	C	D	E	F	A	B	C	D	E	F
1	3	2	2	2	2	1	2	2	2	1	2	1
2	3.5	4	3	3	3	3	4	4	3	4	2.5	4
3	3	3.5	3	2.5	2.5	3	3.9	3.7	3.7	2.5	3.5	2.5
4	4	3	3	4	2	2	1.5	1.5	1.5	1.5	2	2
5	3.5	3.5	3	3	3	2.8	3.5	3.2	2.5	2.5	2.7	3
6	3	3	2	2	2.5	1.5	4	3.5	2	2.5	2	3
7	3	3	3	3	3	3	3	3	3	2	3	2
8	2.5	3	2.5	1.7	3	2	2.5	2.5	3	2	2	2
9	3	4	3	4	3	3	4	4	3.5	2.5	2.7	2
10	3	2	2	2	3	2.5	2	4	3	2.5	2.5	1.5
11	3	2.5	2	3	2	2	2	3	2	2	1.5	2
12	3	3	3	3	3	3	3	3	3	2	3	2
13	1	3	2	3	2	1	3	4	2.5	3	3	1
14	3	3	2.5	2.5	2	2	2.5	2	2	2	1.5	1
15	4	4	3	4	3	2	3	3	3	4	3	2
16	3	4	3	3	2	2	3	4	3	3	2.5	2
17	4	4	4	3	3	3.5	4	3.5	4	3.5	3.5	3.5
18	3	3	3	3	3	3	3	3	3	3	3	3
19	3	3	3.5	3.5	2.7	2	3.5	3	3	2.7	2	1.7
20	4	2	2	2	3	4	4	2	2	3	1	2
21	4	2	4	4	3	4	2	3	3	3	4	4
22	2	2.5	3	3	3	2.5	3	2	3	3	3	2.5
23	3	3	3	4	2	2	2	3	3	3	2	2
24	3	4	3	3	3	3	3	3	2	3	2	2
25	4	3	3	2	4	3	3	4	3	3	3	2
26	4	4	4	4	4	3	3	4	3	2.5	2	2.5
27	3.5	3	2	2	3.5	2	3	2	1	3	1	1
28	4	4	4	3	2.5	3	4	3.5	2.5	3.5	3	2
29	1.5	4	2	3	2	3.5	4	3.5	2	2.5	2	1
30	3.5	3	3.5	3	4	3.5	4	3	3	3.5	3	2
31	3	3	3	2	2	2	3	2.5	2	2	1	3
32	3	2	3	3	2	2	2	1.5	2	2	1.5	2
33	3.5	3.5	3	3.5	3	2	3.5	3	3	3	3	4
34	3.5	3.5	3	3	3.5	3.5	3.5	3.5	4	3	3.5	3
35	2.5	3	2	3	2	2.5	3	2.5	2.5	2.5	2	3
36	4	4	4	4	3	3	4	4	3	3	3	4

TABLE A1.11 Analysis of Variance of Listening Test Results

***** ANALYSIS OF VARIANCE *****

SOURCE OF VARIATION	DF	SS	SS%	MS	VR	SL
SUBJEC STRATUM	35	85.2049	34.72	2.4344	8.256	0.1
SEX STRATUM	1	4.7712	1.94	4.7712	16.182	0.1
SUBJEC.SEX STRATUM	35	17.9305	7.31	0.5123	1.737	1
SEX.SENT STRATUM	10	8.7231	3.55	0.8723	2.958	1
SUBJEC.SEX.SENT STRATUM						
LOSS	5	27.0599	11.03	5.4120	18.355	0.1
RESIDUAL	345	101.7236	41.45	0.2949		
TOTAL	350	128.7835	52.48	0.3680		
GRAND TOTAL	431	245.4132	100.00			

GRAND MEAN 2.817
TOTAL NUMBER OF OBSERVATIONS 432

***** TABLES OF MEANS *****

GRAND MEAN 2.817

LOSS	9/8	8/256	8/8	7/768	7/8	6/2000
	3.137	3.110	2.787	2.825	2.599	2.444

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	LOSS
REP	72
SED	0.0905

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
SUBJEC	35	0.4504	16.0
SEX	1	0.1486	5.3
SUBJEC.SEX	35	0.2922	10.4
SEX.SENT	10	0.1557	5.5
SUBJEC.SEX.SENT	345	0.5430	19.3

TABLE A1.12 Design of Conversation Test

Subjects \ Cards	1	2	3	4	5	6	7	8
1 & 2	21L	9S	7L	11S	9L	7S	21S	11L
3 & 4	9S	9L	11L	7L	21S	11S	7S	21L
5 & 6	7L	21L	21S	9L	7S	9S	11L	11S
7 & 8	11S	11L	7S	21S	9S	21L	9L	7S
9 & 10	9L	21S	21L	7S	11S	11L	7L	9S
11 & 12	7S	7L	11S	11L	21L	9L	9S	21S
13 & 14	11L	11S	9L	9S	7L	21S	21L	7S
15 & 16	21S	7S	9S	21L	11L	7L	11S	9L

Subjects \ Cards	1	2	3	4	5	6	7	8
17 & 18	9S	9L	11L	21S	7L	11S	7S	21L
19 & 20	11S	7L	21S	11L	7S	21S	9L	9S
21 & 22	11L	9S	7S	7L	11S	9L	21L	21S
23 & 24	21S	21L	11S	7S	9L	9S	7L	11L
25 & 26	7L	7S	9S	11S	21S	21L	11L	9L
27 & 28	9L	11L	21S	21L	9S	7L	11S	7S
29 & 30	21L	21S	7L	9L	11L	7S	9S	11S
31 & 32	7S	11S	9L	9S	21L	11L	21S	7L

nL = 14 dB ORE (load)
nS = 34 dB ORE (soft)
} with channel capacity of n

TABLE A1.13 Details of subjects of Conversation Test

Subject no.	Sex	Profession	Age
1	M	Technician	35
2	M	Technician	28
3	M	Computer officer	30
4	M	Lecturer	30
5	M	Technician	50
6	M	Technician	50
7	M	Technician	50
8	M	Technician	50
9	M	Journalist	28
10	F	Housewife	25
11	F	Social worker	26
12	M	Ed. Psychologist	30
13	F	Secretary	50
14	F	Secretary	50
15	F	Schoolteacher	26
16	F	Student	21
17	M	Quantity surveyor	23
18	F	Clerical officer	20
19	M	Radio presenter	28
20	M	Unemployed	21
21	M	Student	26
22	F	Secretary	27
23	M	Electronics eng.	24
24	F	Cook	24
25*	M	Social worker	48
26	M	Various	35
27	F	Student	20
28	F	Nurse	26
29	F	Botanist	26
30	F	Unemployed	40
31	F	Student	21
32	F	Student	20

* Also took part in listening test

TABLE A1.14 Brief for Conversation Test

Imagine that your employers have offered to pay for 3 pictures for your staff common room, but there are only 6 suitable pictures in stock at the art shop, and even some of these may be sold by the time a decision has been made. You have a chance to look through postcard size replicas to form your own opinions, and then you have to speak to your colleague to try to arrive at a joint decision as to which 3 you want.

You are having 4 conversations, and four sets of picture cards are provided. Please make sure you have completed each opinion form before moving on to the next set of cards.

Please follow the instructions on the opinion form. If the green light comes on, call the other person - otherwise wait for them to call you.

TABLE A1.15 Opinion Form used for Conversation Test

OPINION FORM

(1) Before you commence your call, would you please insert below your order of preference for the picture cards. For identification , please use the serial numbers on the backs.

	1st	2nd	3rd	4th	5th	6th
Your Order of Preference						

(2) The green light indicates that it is your turn to originate the call. If the green light is on please call your companion as soon as you are ready by dialling 0 (N.B. there is no dialling tone).

(3) Please insert below the order of preference arrived at after discussion with your conmpanion:

	1st	2nd	3rd	4th	5th	5th
Agreed Order of Preference						

(4) When you have completed the conversation, please replace the handset.

(5) Please mark by a cross, your opinion of the telephone call you have just had.

N.B. Please do not discuss your opinion with your companion.

Excellent	Good	Fair	Poor	Bad

(6) Did you, or the person who spoke to you, have any difficulty in talking or hearing over the connection?

YES	
NO	

If YES, please explain in a few words the nature of your difficulty:

Now move on to the next set of picture cards.

TABLE A1.16 Results of Conversation Test

Cards	1		2		3		4		5		6		7		8		Total Difficulty
Link \ Room	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	
21L	4 3	4 5	* 4	5 4	5 4	4 5	4 5	4 5	5 4	4 5	3 5	5 4	5 4	4 4	5 4	4 4	0/31
11L	5 3D	4 4	3 5	4 4	5 5	* 2D	5 4	4 4	4 4	4 3	4 4	4 3	5 5	4 4D	5 3	5 3	3/31
9L	3D 4	4 3D	4 4	* 3D	5 4	4 3	* 3	* 5	5 3D	5 3	4 3	4 3	2 4	5 2D	3D 2D	3 2D	8/29
7L	* 3	3 2D	3D 4	2D 2D	5 1D	2 3D	4 1D	* 2D	3D 3D	3 3	2D 1D	3 2D	2D 3D	2D 3	2 2D	4 4D	18/30
21S	2D 4D	2D 2D	2D 2D	1D 4D	* 2D	* 4	2 2D	3D 1D	2 5	4 2D	5 2	3 3	2D 2D	2 4	2 2D	3D 1D	18/30
11S	2D 3	2D 3	4 3D	2 4	4D 3D	3 1D	2D 4	3 3D	3D 1D	2D 2D	2 1D	3 2D	2D 2D	4 3D	4 2D	3 3D	19/32
9S	2 2D	* 2D	2 1D	2 1D	2D 4	3 3D	4 3	3 4	2D 2D	3D 1D	4 2D	3 1D	2 2D	3D 2D	2D 3	2D 3	17/31
7S	3D 2D	2D 4	1D 2	2D 2D	2D 2D	1D 1D	1D 2D	1D 2D	2D 3	2D 2D	2D 1D	1D 2D	1 2D	1D 2D	3 2D	3D 1D	27/32

D = difficulty experienced
* = missing value

TABLE A1.17 Analysis of Variance of Conversation Test Results

***** ANALYSIS OF VARIANCE *****

VARIATE: SCORE

SOURCE OF VARIATION	DF(MV)	SS	SS%	MS	VR	SL
TEST STRATUM	7	3.4543	1.01	0.4935	0.899	NS
ROOM STRATUM	1	0.5930	0.17	0.5930	1.080	NS
TEST.ROOM STRATUM	7	5.7355	1.68	0.8194	1.492	NS
ROOM.CONV STRATUM	30	73.4975	21.49	2.4499	4.462	0.1
TEST.ROOM.CONV STRATUM						
VOLUME	1	104.0692	30.43	104.0692	189.521	0.1
LOSS	3	55.2422	16.15	18.4141	33.534	0.1
VOLUME.LOSS	3	7.7764	2.27	2.5921	4.721	1
RESIDUAL	193(10)	105.9795	30.99	0.5491		
TOTAL	200	273.0673	79.85	1.3653		
GRAND TOTAL	245	356.3476	104.20			

ESTIMATED GRAND MEAN	3.025
TOTAL NUMBER OF OBSERVATIONS	256
NUMBER OF MISSING VALUES	10

***** TABLES OF MEANS *****

VARIATE: SCORE

GRAND MEAN 3.025

VOLUME	LOUD	SOFT
	3.662	2.387

LOSS	NONE	LOW	MED	HIGH
	3.476	3.343	3.002	2.278

LOSS / VOLUME	NONE	LOW	MED	HIGH
LOUD	4.360	4.030	3.578	2.682
SOFT	2.593	2.656	2.425	1.875

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	VOLUME	LOSS	VOLUME LOSS
REP	128	64	32
SED	0.0926	0.1310	0.1853

(NOT ADJUSTED FOR MISSING VALUES)

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
TEST	7	0.1242	4.1
ROOM	1	0.0681	2.3
TEST.ROOM	7	0.2263	7.5
ROOM.CONV	30	0.5534	18.3
TEST.ROOM.CONV	193	0.7410	24.5

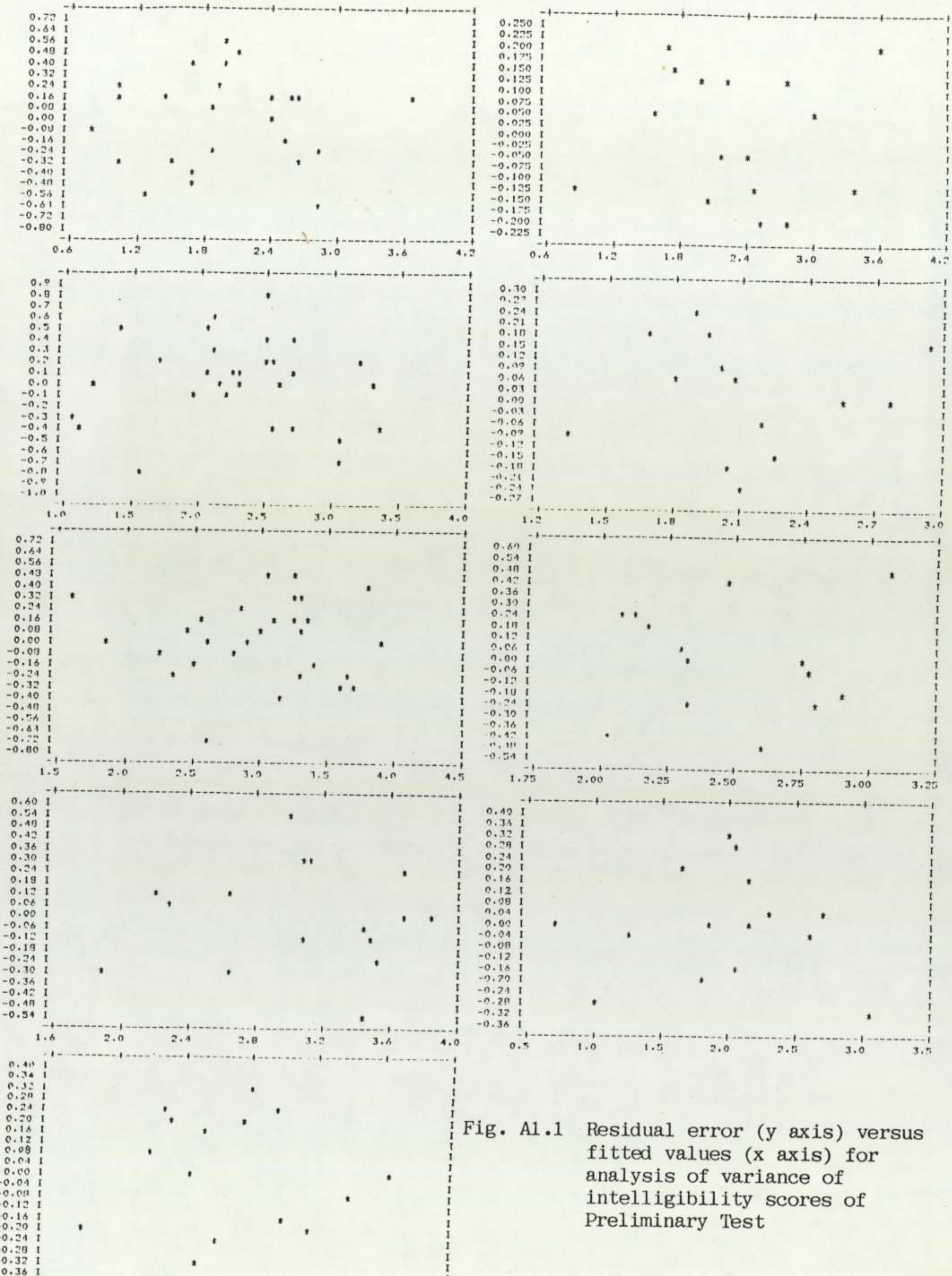
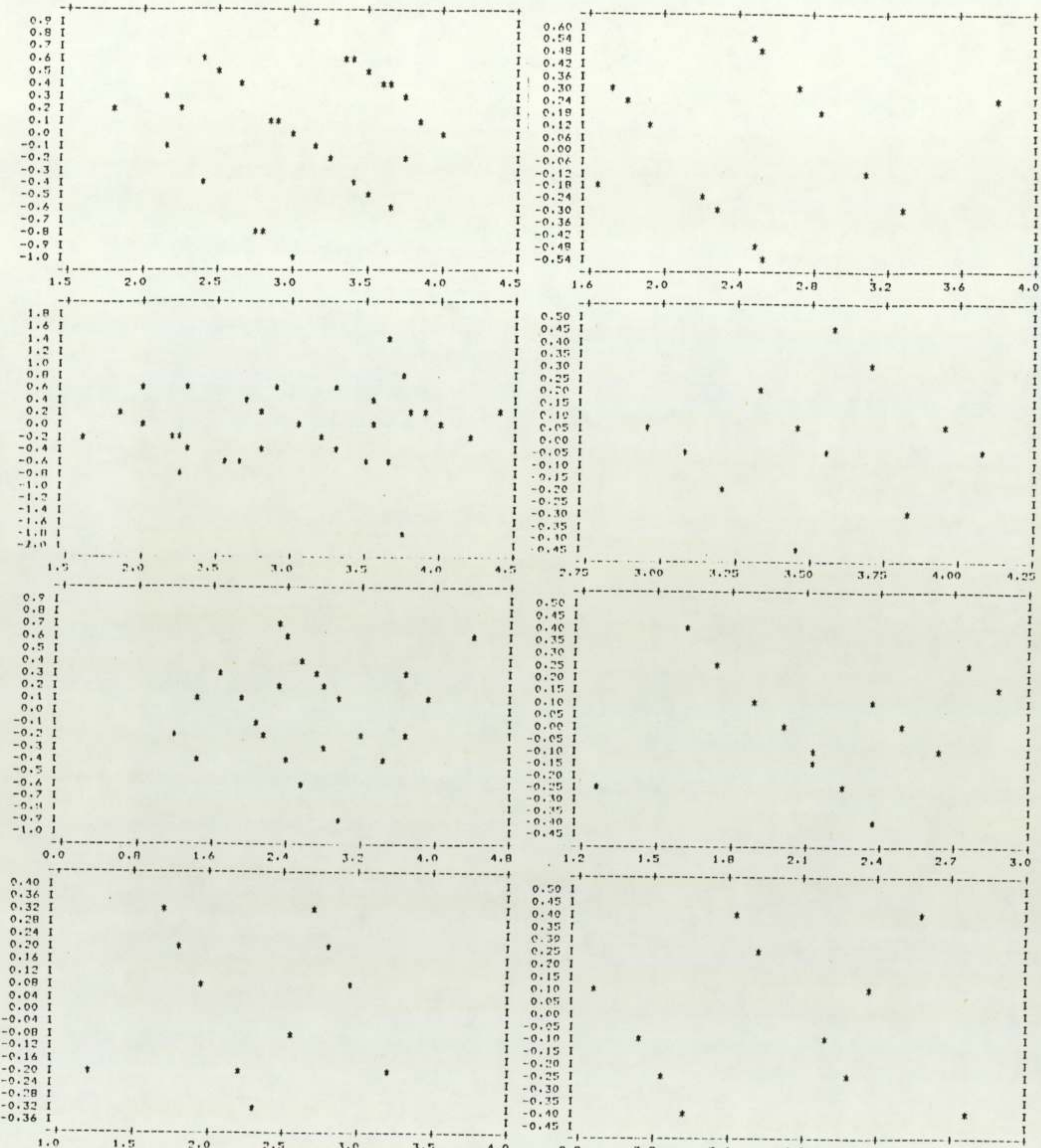


Fig. A1.1 Residual error (y axis) versus fitted values (x axis) for analysis of variance of intelligibility scores of Preliminary Test



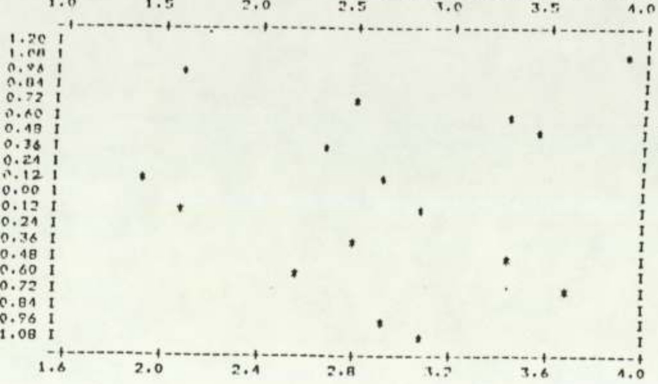


Fig. A1.2 Residual error (y axis) versus fitted values (x axis) for analysis of variance of opinion scores of Preliminary Test

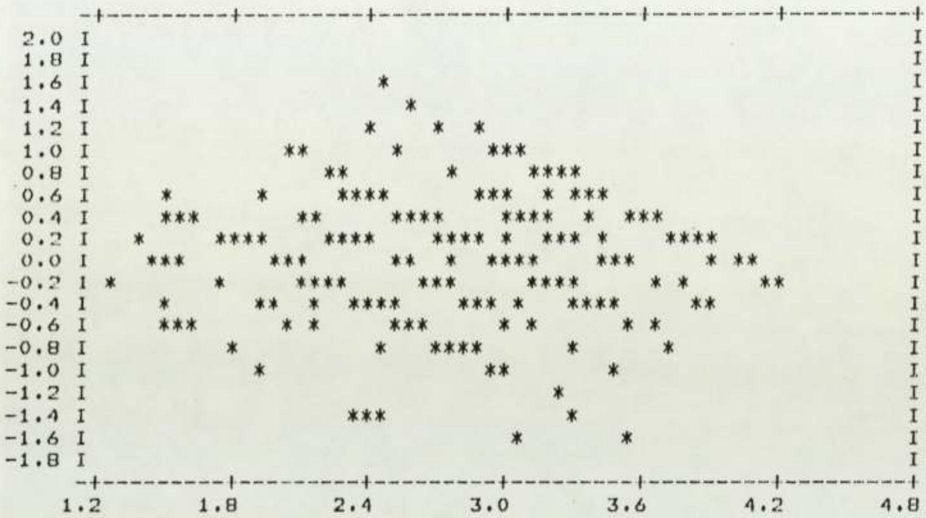


Fig. A1.3 Residual error (y axis) versus fitted values (x axis) for analysis of variance of Listening Test

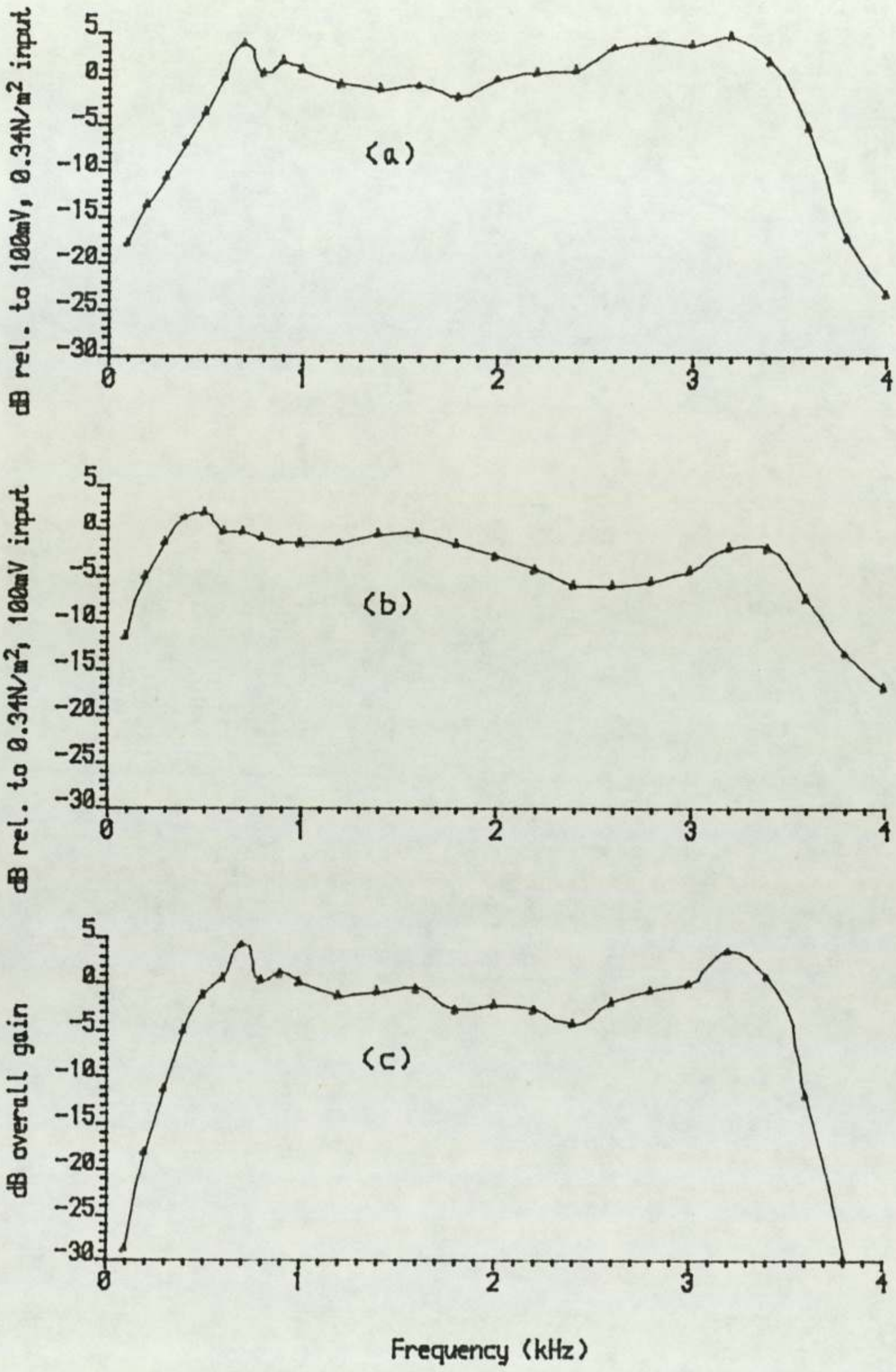


Fig. A1.4 Frequency responses of
a) telephone mouthpiece and speech interface
b) telephone earpiece
c) overall sound path

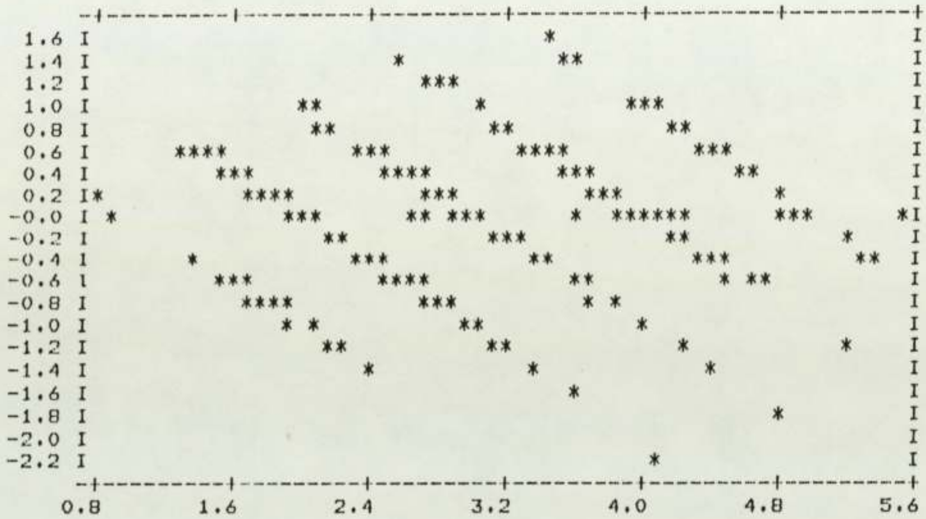


Fig. A1.5   Residual error (y axis) versus fitted values (x axis) for analysis of variance of Conversation Test.

APPENDIX 2 LISTING OF PROGRAM TO CALCULATE PACKET LOSS

```
C*****************************************************************************
C
C     This program computes the state matrix of a speech-packet queue.
C     A general solution to the state equations is obtained by
C     finding the eigenvalues and eigenvectors of the state matrix.
C     Boundary equations are formed and solved to give the percentage
C     of packets lost in the queue.
C
C*****************************************************************************
C
C              *****GLOSSARY OF VARIABLES*****
C
      DOUBLE PRECISION R(70,70),C(70,70),Z(70,70)
C     R contains the state matrix
C     C contains the eigenvectors
C     Z contains the LHS of the boundary equations
C     R and Z are also used as temporary workspace matrices
C
      DOUBLE PRECISION B(70)
C     This contains the overall probability of each state
      DOUBLE PRECISION EIG(70)
C     This contains the eigenvalues
      DOUBLE PRECISION EXINT(70)
C     This contains a function of the eigenvalues and the queue limit
      DOUBLE PRECISION COL(70)
C     This contains the RHS of the boundary equations
      DOUBLE PRECISION A(70)
C     This contains the solutions to the boundary equations
C
      INTEGER NOUT
C     NOUT is the output device
      INTEGER V
      REAL H,RO,RMU
C     V is the number of callers, H is the number of channels, RO is the
C     ratio of talkspurt to silence length, and RMU the average
C     talkspurt length
      INTEGER NVAL
      REAL DSTEP,MSTEP
C     NVAL is the number of delay points wanted, DSTEP the delay
C     increment between each point, and MSTEP the equivalent queue
C     length increment
      INTEGER IH
C     IH is the integer part of H
      INTEGER JU,JL
C     JU and JL are the lower and upper matrix limits
      REAL JAVE
C     Jave is the average value of J
      INTEGER NV,NH
C     NV is the size of the state matrix, NH indicates the row
C     which corresponds to the nearest state below (but not
C     including) H. These are different for integer and non-integer
C     values of H
      INTEGER IPOS,INEG
C     IPOS is the number of positive eigenvalues, INEG the number
C     of non-zero eigenvalues
      REAL M,FL
C     M is the queue limit, FL is the percentage packet loss
C
      DOUBLE PRECISION ER(70),EI(70),WKS1(70),WKS2(70),X02AAF
      INTEGER IWKS(70),IFAIL,IT
C     These are used in the NAG subroutines. In particular, ER will
C     temporarily contain the eigenvalues
C
C              *****DEFINE PARAMETERS*****
C
C
C     Examples are given here
      DATA NOUT/3/
      DATA V/160/,H/60.0/,RO/0.723/
      DATA RMU/1.2/,NVAL/20/,DSTEP/0.05/
C
  100 FORMAT(' PACKET LOSS WITH DELAY LIMIT OF ',F5.3,' IS ',F5.2)
```

```
C             *****MAIN PROGRAM*****
C
      MSTEP=H*DSTEP/RMU
      IH=H
C
C     Form bounds to JU and JL and call routine to find JU and JL
C     accurately
      JU=V*(0.268+0.327*RO)+15
      JU=MIN(V,JU)
      JL=(V-40)*.45*RO
      JL=MAX(0,JL)
      CALL MATLIM(V,RO,JU,JL,B,JAVE)
C
C     Form R matrix
      IF((IH-H).EQ.0) CALL INTMAT(R,JU,JL,NV,V,H,NH,IH,RO,B)
      IF((IH-H).NE.0) CALL NORMAT(R,JU,JL,NV,V,H,NH,IH,RO)
C
C     Put the matrix in Upper Hessenburg form
      CALL F01AKF(NV,1,NV,R,70,IWKS)
      CALL F01APF(NV,1,NV,IWKS,R,70,Z,70)
C
C     Find the eigenvalues and eigenvectors
      IFAIL=1
      CALL F02AQF(NV,1,NV,X02AAF(IT),R,70,Z,70,ER,EI,IWKS,IFAIL)
      IF(IFAIL.NE.0) GOTO 999
C
C     The eigenvalues with their corresponding eigenvectors are
C     ordered into first positive, then negative, then zero values
      CALL REORD(C,Z,EIG,ER,NV,IPOS,INEG)
C
C     Packet loss is calculated for a number of delays
      DO 10 I=0,NVAL
      M=I*MSTEP
C
C     Form the boundary equations
      CALL BEQ(C,Z,IPOS,INEG,NV,NH,EIG,M,V,H,IH,RO,EXINT,COL,B)
C
C     Solve the boundary equations
      IFAIL=1
      CALL F04ATF(Z,70,COL,NV,A,R,70,WKS1,WKS2,IFAIL)
      IF(IFAIL.NE.0) GOTO 10
C
C     Calculate the packet loss
      CALL PACLOS(C,PL,NH,NV,A,EXINT,B,H,IH,JAVE)
      WRITE(NOUT,100) I*DSTEP,PL
   10 CONTINUE
C
  999 STOP
      END
```

```
C::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
      SUBROUTINE MATLIM(V,RO,JU,JL,B,JAVE)
C
C     This subroutine finds the limits to the state matrix
C     and forms the B vector
C
C::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
C              *****PARAMETERS OF SUBROUTINE*****
C
C     (See calling program for glossary of these variables)
C
      DOUBLE PRECISION B(70)
      INTEGER V,JU,JL
      REAL RO,JAVE
C
C              *****LOCAL VARIABLES*****
C
      REAL PR,P
C     PR is the reciprocal of the probabilty of JL, P is the
C     probability of JU
      REAL S
      INTEGER I,J
C     These are temporary variables
C
C              *****MAIN PROGRAM OF SUBROUTINE*****
C
C     Assuming JU to be correct, the probability of JL is evaluated
C     for increasing values of JL, until PR<10**5
      PR=100001.
      JL=JL-1
   10 IF(PR.LT.100000.) GOTO 20
       S=1.
       PR=1.
       JL=JL+1
       DO 30 J=JL+1,JU
        S=S*(V-J+1)*RO/J
        PR=PR+S
   30   CONTINUE
      GOTO 10
   20 CONTINUE
C
C     JL has now been found. JU is calculated by finding the
C     probability of every state from JL upwards until P<10**-5
C     The previous state is then the desired value of JU
      P=1/PR
      J=JL
   40 IF(P.LT.0.00001) GOTO 50
       J=J+1
       P=P*(V-J+1)*RO/J
       GOTO 40
   50 CONTINUE
C
C     Both JL and JU have now been found. The B vector is formed
C     using these values; first the exact probability of JL is found
      JU=J-1
      S=1.
      PR=1.
      DO 60 J=JL+1,JU
       I=J-1
       S=S*(V-I)*RO/J
       PR=PR+S
   60 CONTINUE
C
      P=1/PR
      B(1)=P
      JAVE=B(1)*JL
      DO 70 J=JL+1,JU
       B(J-JL+1)=B(J-JL)*(V-J+1)*RO/J
       JAVE=JAVE+B(J-JL+1)*J
   70 CONTINUE
C
      RETURN
      END
```

```
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
      SUBROUTINE INTMAT(R,JU,JL,NV,V,H,NH,IH,RO,B)
C
C     This subroutine forms the state matrix for an integer
C     value of H
C
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
C              *****PARAMETERS OF SUBROUTINE*****
C     (See calling program for glossary of these variables)
C
      DOUBLE PRECISION R(70,70),B(70)
      REAL H,RO
      INTEGER JU,JL,NV,V,NH,IH
C
C              *****LOCAL VARIABLES*****
C
      INTEGER I,J
C     These are temporary variables
C
C              *****MAIN PROGRAM OF SUBROUTINE*****
C
      NV=JU-JL
      NH=IH-JL
C
C     The state matrix is cleared
      DO 10 I=1,NV
       DO 20 J=1,NV
        R(J,I)=0
   20   CONTINUE
   10  CONTINUE
C
C     First, the top of the matrix is formed
      DO 30 J=JL,IH-2
       I=J-JL+1
       R(I,I)=((V-J)*RO+J)/(J-H)
       R(I,I+1)=(J+1)/(H-J)
       R(I+1,I)=(V-J)*RO/(H-J-1)
   30  CONTINUE
C
C     The adjusted lines are formed
      I=IH-JL
      R(I,I)=H*RO*(V-H+1)/((V-H)*RO+H)-((V-H+1)*RO+H-1)
      R(I,I+1)=H*(H+1)/((V-H)*RO+H)
      R(I+1,I)=-(V-H)*RO**2*(V-H+1)/((V-H)*RO+H)
C
      I=I+1
      R(I,I)=(V-H-1)*RO+H+1-(V-H)*(H+1)*RO/((V-H)*RO+H)
      R(I,I+1)=-(H+2)
      R(I+1,I)=-(V-H-1)*RO/2
C
C     The bottom half is formed
      DO 40 J=IH+2,JU
       I=J-JL
       R(I,I)=((V-J)*RO+J)/(J-H)
       R(I,I+1)=(J+1)/(H-J)
       R(I+1,I)=(V-J)*RO/(H-J-1)
   40  CONTINUE
C
C     Finally, the top left and bottom right values are reset
      R(1,1)=(V-JL)*RO/(JL-H)
      IF(V.NE.(IH+1)) R(NV,NV)=JU/(JU-H)
C
C     The vector B is also adjusted by removing Bh and pushing the
C     values above Bh down by one element
      DO 50 J=IH-JL+1,NV
       B(J)=B(J+1)
   50  CONTINUE
C
      RETURN
      END
```

```
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
      SUBROUTINE NORMAT(R,JU,JL,NV,V,H,NH,IH,RO)
C
C     This subroutine forms the state matrix for a non-integer
C     value of H
C
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
C             *****PARAMETERS OF SUBROUTINE*****
C     (See calling program for glossary of these variables)
C
      DOUBLE PRECISION R(70,70)
      REAL H,RO
      INTEGER JU,JL,NV,V,NH,IH
C
C             *****LOCAL VARIABLES*****
C
      INTEGER I,J
C     These are temporary variables
C
C             *****MAIN PROGRAM OF SUBROUTINE*****
C
      NH=IH-JL+1
      NV=JU-JL+1
C
C     The state matrix is cleared
      DO 10 I=1,NV
       DO 20 J=1,NV
        R(J,I)=0
   20   CONTINUE
   10  CONTINUE
C
C     The matrix is formed
      DO 30 J=JL,JU
       I=J-JL+1
       R(I,I)=((V-J)*RO+J)/(J-H)
       R(I,I+1)=(J+1)/(H-J)
       R(I+1,I)=(V-J)*RO/(H-J-1)
   30  CONTINUE
C
C     Finally, the top left and bottom right values are reset
      R(1,1)=(V-JL)*RO/(JL-H)
      R(NV,NV)=JU/(JU-H)
C
      RETURN
      END
```

```
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
      SUBROUTINE REORD(C,Z,EIG,ER,NV,IPOS,INEG)
C
C     This subroutine re-orders the eigenvalues and eigenvectors
C     according to the sign of each eigenvalue. The positive
C     ones come first, then the negative ones and finally the
C     zero ones
C
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
C              *****PARAMETERS OF SUBROUTINE*****
C     (See calling program for glossary of these variables)
C
      DOUBLE PRECISION C(70,70),Z(70,70),EIG(70),ER(70)
      INTEGER NV,IPOS,INEG
C
C              *****LOCAL VARIABLES*****
C
      INTEGER K,I,J
C     These are temporary variables
C
C              *****MAIN PROGRAM OF SUBROUTINE*****
C
      K=0
C
C     First the positive eigenvalues are dealt with
      DO 10 I=1,NV
       IF(ER(I).LT.0.00001) GOTO 10
       K=K+1
       EIG(K)=ER(I)
       DO 20 J=1,NV
        C(J,K)=Z(J,I)
   20   CONTINUE
   10  CONTINUE
      IPOS=K
C
C     Next the negative eigenvalues are dealt with
      DO 30 I=1,NV
       IF(ER(I).GT.-0.00001) GOTO 30
       K=K+1
       EIG(K)=ER(I)
       DO 40 J=1,NV
        C(J,K)=Z(J,I)
   40   CONTINUE
   30  CONTINUE
      INEG=K
C
C    Finally the zero eigenvalues are dealt with
      DO 50 I=1,NV
       IF((ER(I).LT.-0.00001).OR.(ER(I).GT.0.00001)) GOTO 50
       K=K+1
       EIG(K)=0
       DO 60 J=1,NV
        C(J,K)=Z(J,I)
   60   CONTINUE
   50  CONTINUE
C
      RETURN
      END
```

```
C::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
       SUBROUTINE BEQ(C,Z,IPOS,INEG,NV,NH,EIG,M,V,H,IH,RO,EXINT,COL,B)
C
C     This subroutine forms the boundary equations
C
C::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
C              *****PARAMETERS OF SUBROUTINE*****
C     (See calling program for glossary of these variables)
C
      DOUBLE PRECISION C(70,70),Z(70,70)
      DOUBLE PRECISION EIG(70),EXINT(70),COL(70),B(70)
      REAL M,H,RO
      INTEGER IPOS,INEG,NV,NH,V,IH
C
C              *****LOCAL VARIABLES*****
C
      DOUBLE PRECISION EX(70)
C     EX contains a function of EIG and M
      REAL K1,K2
C     These are used in the calculation of the boundary equations
C     for the states immedeately above and below H
      INTEGER I,J
C     These are temporary variables
C
C              *****MAIN PROGRAM OF SUBROUTINE*****
C
C     First, EX and EXINT are calculated
      DO 10 I=1,INEG
       EX(I)=EXP(-ABS(EIG(I))*M)
       EXINT(I)=ABS((1-EX(I))/EIG(I))
   10  CONTINUE
      DO 20 I=INEG+1,NV
       EX(I)=1.
       EXINT(I)=M
   20 CONTINUE
C
C     Form boundary equations for J<H-1
      DO 30 J=1,NH-1
       DO 40 I=1,IPOS
        Z(J,I)=C(J,I)*EX(I)
   40   CONTINUE
       DO 50 I=IPOS+1,NV
        Z(J,I)=C(J,I)
   50   CONTINUE
   30  CONTINUE
C
C     Form boundary equations for J>H+1
      DO 60 J=NH+2,NV
       DO 70 I=1,IPOS
        Z(J,I)=C(J,I)
   70   CONTINUE
       DO 80 I=IPOS+1,NV
        Z(J,I)=C(J,I)*EX(I)
   80   CONTINUE
   60  CONTINUE
C
C     Form the boundary equations for H-1<J<H+1
      K1=(IH+1-H)/((V-IH)*RO)
      IF((IH-H).EQ.0) K1=K1*((V-H)*RO+H)/((V-H+1)*RO)
      K2=(H-IH)/(IH+1)
      IF((IH-H).EQ.0) K2=((V-H)*RO+H)/(H*(H+1))
C
      DO 90 I=1,IPOS
       Z(NH,I)=C(NH,I)*EXINT(I)+K1*C(NH+1,I)
       Z(NH+1,I)=C(NH+1,I)*EXINT(I)+K2*C(NH,I)*EX(I)
   90  CONTINUE
      DO 100 I=IPOS+1,NV
       Z(NH,I)=C(NH,I)*EXINT(I)+K1*C(NH+1,I)*EX(I)
       Z(NH+1,I)=C(NH+1,I)*EXINT(I)+K2*C(NH,I)
  100  CONTINUE
C
C     Form the RHS column vector
      DO 110 I=1,NV
       COL(I)=0
  110  CONTINUE
      COL(NH)=B(NH)
      COL(NH+1)=B(NH+1)
C
      RETURN
      END
```

```
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
      SUBROUTINE PACLOS(C,PL,NH,NV,A,EXINT,B,H,IH,JAVE)
C
C     This subroutine calculates the percentage packet loss
C
C:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
C
C              *****PARAMETERS OF SUBROUTINE*****
C     (See calling program for glossary of variables)
C
      DOUBLE PRECISION C(70,70),A(70),EXINT(70),B(70)
      REAL PL,H,JAVE
      INTEGER NH,NV,IH
C
C              *****LOCAL VARIABLES*****
C
      REAL LS
C     This is the end probabilty of each state
      INTEGER K,I,J
C     These are temporary variables
C
C              *****MAIN PROGRAM OF SUBROUTINE*****
C
      PL=0
      K=0
      DO 10 J=NH+1,NV
       K=K+1
       LS=0
       DO 20 I=1,NV
        LS=LS-A(I)*C(J,I)*EXINT(I)
   20   CONTINUE
       LS=LS+B(J)
       PL=PL+(IH+K-H)*LS
   10  CONTINUE
      PL=PL*100/JAVE
C
      RETURN
      END
```

REFERENCES

1. R. Cole, "Packet Voice: When it makes sense", Speech Technology, Sept./Oct. 1982, pp. 52-61

2. J. Gruber, "Performance considerations for integrated voice and data networks", Computer Communications, June 1981, Vol 4 No. 3, pp. 106-126

3. J. W. Forgie and A.G. Nemeth, "An efficient packetized voice/data network using statistical flow control", Conf. Rec. Int. Conf. Commun., 1977, Vol 3, pp. 38.2.44-48

4. K. Bullington and J.M. Fraser, "Engineering aspects of TASI", Bell Syst. Tech. Journ., March 1959, pp. 353-364

5. N.S.Jayant and S.W. Christensen, "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure", IEEE Trans. Commun., 1981, COM-29, pp. 101-109

6. T. Bially, B.Gold and S.Seneff, "A technique for adaptive voice flow control in integrated packet networks", IEEE Trans. Commun., 1980, COM-28, pp. 325-333

7. D.J. Goodman, "Embedded DPCM for variable bit rate transmission", Conf. Rec. Int. Conf. Commun., 1980, Pt III, pp. 42.2. 1-5

8. D. Seitzer, H. Gerhauser, G. Langenbucher, "A comparative study of high quality digital speech interprolation methods", Conf. Rec. Int. Conf. Commun., 1978, pp. 50.2.1-5

9. C.J. Weinstein and E.M. Hofstetter, "The tradeoff between delay and TASI advantage in a packetized speech multiplexer", IEEE Trans. Commun., 1979, COM-27, pp. 1716-1720

10. R.V. Cox and R.E. Crochiere, "Multiple user variable rate coding for TASI and packet transmission systems", IEEE Trans. Commun., 1980, COM-28, pp. 334-344

11. D.T. Magill, "Adaptive speech compression for packet communication systems", Proc. Nat. Telecommun. Conf., 1973, Vol 2, pp. 29 D1-5

12. C.K. Un and D.H. Cho, "Hybrid companding delta modulation with variable-rate sampling", IEEE Trans. Commun., 1982, COM-30, pp. 593-599

13. B.H. Atal, "Predictive coding of speech at low bit rates", IEEE Trans. Commun., 1982, COM-30, pp. 600-614

14. V.R. Viswanathan, J. Makhoul, R.M. Schwartz and A.W.F. Huggins, "Variable frame rate transmission: a review of methodology and application to narrow-band LPC speech coding", IEEE Trans. Commun., 1982, COM-30, pp. 674-686

15. M. Listanti and F. Villani, "An X.25-compatible protocol for packet voice communications", Computer Communications, Feb. 1983, Vol. 6 No. 1, pp. 23-31

16. D. Cohen, "A protocol for packet-switching voice communication", Computer Networks (Netherlands), Sept./Oct. 1978, Vol. 2 No. 4-5, pp. 320-331

17. J. W. Forgie, "Speech transmission in packet-switched store-and-forward networks", Proc. Nat. Comp. Conf., 1975, pp. 137-142

18. B. Gold, "Digital speech networks", Proc. IEEE, 1977, Vol. 65, pp. 1636-1658

19. C.J. Weinstein and J.W. Forgie, "Experience with speech communication in packet networks", IEEE Journ. Selected Areas in Commun., 1983, SAC-1, pp. 963-980

20. G.J. Coviello, O.L. Lake and G.R. Redinbo, "System design implications of packetized voice", Conf. Rec. Int. Comp. Conf. 1977, pp. 38.3.49-53

21. G.J. Coviello, "Comparative discussion of circuit - vs. packet-switched voice", IEEE Trans. Commun., 1979, COM-27, pp. 1153-1160

22. H. Ueda, I. Tokizawa and T. Aoyama, "Evaluation of an experimental packetized speech and data transmission system", IEEE Journ. Selected Areas in Commun., 1983, SAC-1, pp. 1039-1045

23. D. Conrads and P. Kermani, "Some issues of transmission of packetized voice through store-and-forwards data communication networks", Proc. 6th Int. Conf. Comp. Commun., London, England, Sept. 1982, pp. 725-9

24. T. Suda, H. Miyahara and T. Hasegawa, "Performance evaluation of a packetized voice system - simulation study", IEEE Trans. Commun., 1984, COM-32, pp. 97-102

25. M.J. Ross, A.C. Tabbot and J. A. Waite, "Design approaches and performance criterea for integrated voice/data switching", Proc. IEEE, 1977, Vol 65, pp. 1283-1295

26. D. Minoli, "Optimal packet length for packet voice communication", IEEE Trans. Commun., 1979, COM-27, pp. 607-611

27. D. Minoli, "Issues in packet voice communication", Proc. IEE, 1979, Vol. 126, pp. 729-740

28. W. A. Montgomery, "Techniques for packet voice synchronization", IEEE Journ. Selected Areas in Commun., 1983, SAC-1, pp. 1022-1028

29. D. L. Duttweiler, "Bell's echo-killer chip", IEEE Spectrum, Oct. 1980, pp. 34-37

30. M. M. Sondhi and D. A. Berkley, "Silencing echoes on the telephone network", Proc. IEEE, 1980, Vol. 68, pp. 948-63

31. R. G. Gould and G. K. Helder, "Transmission delay and echo suppression", IEEE Spectrum, Apr. 1970, pp. 47-54

32. D. L. Richards and J. Hutter, "Echo suppressors for telephone connections having long propogation times", Proc. IEE, 1967, Vol. 116, pp. 955-963

33. D. L. Richards, "Conversation performance of speech links subject to long propagation times", Proc. Int. Conf. on Satellite Commun., 1962, pp. 247-251

34. D. L. Richards, "Telecommunication by speech", Butterworth, 1973

35. P. T. Brady, "Effects of transmission on delay on conversational behaviour on echo-free telephone circuits", Bell Syst. Tech. Journ., Jan. 1971, pp. 115-134

36. R. M. Krauss and P. D. Bricker, "Effects of transmission delay and access delay on the efficiency of verbal communication", Journ. Acoust. Soc. Am., 1967, Vol. 41 No. 2, pp. 286-292

37. E. T. Klemmer, "Subjective evaluation of transmission delay in telephone conversations", Bell. Syst. Tech. Journ., July-Aug. 1967, pp. 1141-1147

38. CCITT White Book, Vol. V, Recommendation P. 14, 1968

39. "Handbook of Data Communications", National Computing Centre Ltd., 1982

40. J. Seguel, Y. Tanaka and M. Akiyama, "Simulation analysis of the waiting time distribution of a packetized voice concentrator", Trans. IEEE Japan, 1982, Vo. E65, pp. 115-122

41. D. Cohen, "Issues in transnet packetized communication", Proc. 5th Data Commun. Symp., 1977, pp. 6.10-13

42. G. Barberis, "Buffer sizing of a packet-voice receiver", IEEE

Trans. Commun., 1981, COM-29, pp. 152-156

43. G. Barberis and D. Pazzaglia, "Analysis and optimal design of a packet-voice receiver", IEEE Trans. Commun., 1980, COM-28, pp. 217-227

44. A. W. F. Huggins, "Effect of lost packets on speech intelligibility", NSC Note 78, Feb. 24 1976, and BBN Report No. 3263 Appendix D.

45. W. E. Naylor and L. Kleinrock, "Stream traffic communication in packet switched networks: destination buffering considerations", IEEE Trans. Commun., 1982, COM-30, pp. 2527-2534

46. N. Janakiraman, B. Pagurek and J. E. Neilson, "Multiplexing low-speed buffered data terminals", IEEE Trans. Commun., 1980, COM-28, pp. 1838-1843

47. N. Janakiraman, B. Pagurek and J. E. Neilson, "Delay versus TASI advantage in a packet voice multiplexer", IEEE Trans. Commun., 1984, COM-32, pp. 319-320

48. G. Barberis and G. Tamburelli, "Topics in analysis and design of generalised packet communication networks", Proc. GLOBECOM '82, 1982, Vol. 2, pp. 553-558

49. M. S. Nakhla and D. H. A. Black, "Analysis of a TASI System employing speech storage", IEEE Trans. Commun., 1982, COM-30, pp. 780-785

50. M. J. Fischer, "Delay analysis of TASI with random fluctuations in the number of voice calls", IEEE Trans. Commun., 1980, COM-28, pp. 1883-1889

51. S. A. Webber, C. J. Harris and J. L. Flanagan, "Use of variable-quality coding and time-interval modification in packet transmission of speech", Bell Syst. Tech. Journ., 1977, Vol. 56, pp. 1569-1573

52. D. Minoli, "General geometric arrival-constant server queue with applications to packetized voice", Conf. Rec. Int. Conf. Commun., 1978, Vol. 3, pp. 36.6.1-5

53. B. G. Kim, "Characterization of arrival statistics of multiplexed voice packets", IEEE Journ. of Selected Areas in Commun., 1983, SAC-1, pp. 1133-1139

54. Y. Tanaka, J. Seguel and M. Akiyama, "Performance analysis of a voice packet multiplexer", Trans. IEEE Japan, 1982, Vol. J65-B, pp. 539-546

55. P. T. Brady, "A statistical analysis of on-off patterns in 16

conversations", Bell Syst. Tech. Journ., 1968, Vol. 47, pp. 73-91

56. J. G. Gruber, "A comparison of measured and calculated speech temporal parameters relevant to speech activity detection", IEEE Trans. Commun., 1982, COM-30, pp. 728-738

57. C. J. Weinstein, "Fractional speech loss and talker activity model for TASI and for packet-switched speech", IEEE Trans. Commun., 1978, COM-26, pp. 1253-1257

58. J. W. Forgie, "Network spech: system implication of packetized speech", M.I.T. Lincoln Lab., Lexington, MA, Annu. Rep. ESD-TR-77-178 to Defence Commun. Agency, Sept. 1977

59. J. Gruber and L. Strawczynski, "Subjective effects of variable delay and speech loss in dynamically managed voice systems", Conf. Rec. GLOBECOM, Miami, FL, Nov-Dec. 1982, Vol. 3, pp. F7.3.1-5

60. J. M. Musser, T. T. Liu, L. Li and G. J. Boggs, "A local area network as a telephone local subscriber loop", IEEE Journ. on Selected Areas in Commun., 1983, SAC-1, pp. 1046-1053

61. J. M. Musser, T. T. Liu and F. P. Tredeau, "Packet-voice performance on a CSMA-CD local area network, "Int. Symp. Subscriber Loops and Service, Sept. 1982, pp. 40-44

62. G. A. Miller and J. C. R. Licklider, "The intelligibility of interrupted speech", Journ. Acoust. Soc. Am., 1950, Vol. 22 No. 2, pp. 167-173

63. D. J. Sharf, "Intelligibility of reiterated speech", Journ. Acoust. Soc. Am., 1959, Vol. 31 No. 4, pp. 423-427

64. R. R. Riesz and E. T. Klemmer, "Subjective evaluation of delay and echo suppressors in telephone conversations", Bell Syst. Tech. Journ., Nov. 1963, pp. 2919-2941

65. T. Bially, A. J. McLaughlin and C. J. Weinstein, "Voice communication in integrated digital voice and data networks", IEEE Trans. Commun., 1980, COM-28, pp. 1478-1490

66. Y. Yatsuzuka, "Highly sensitive speech detector and high-speed voice and data discriminator in DSI-ADPCM systems", IEEE Trans. Commun., 1982, COM-30, pp. 739-750

67. P. Di Carlo and M. R. L. Hodges, "Digital speech detection", Colloq. on "Digital Processing of Speech", London, England, Apr. 1983, pp. 9/1-7

68. J. W. Emling and D. Mitchell, "The effects of time delay and echoes on telephone conversations", Bell Syst. Tech. Journ.,

Nov. 1963, pp. 2869-2891

69. M. A. Marsan and G. Albertengo, "Integrated voice and data network", Computer Communications, June 1982, Vol. 5 No. 3, pp. 119-127

70. P. C. Ravasio, R. Marcogliese and R. Novarese, "Voice transmission over an Ethernet backbone", Local Computer Networks, North-Holland Publishing Co., 1982, pp. 39-65

71. T. T. Liu and L. Li, "A virtual token scheme, GBRAM, for local data and voice distribution", Proc. 6th Int. Conf. Comp. Commun., London, England, Sept. 1982, pp. 225-30

72. T. Tsuda, S. Hattori, R. Yatsuboshi and K. Yamauchi, "An approach to multi-service subscriber loop system using packetized voice/data terminals", IEEE Trans. Commun., 1979, COM-27, pp. 1112-1117

73. D. H. Johnson and G. C. O'Leary, "A local access network for packetized digital voice communication", IEEE Trans. Commun., 1981, COM-29, pp. 679-688

74. A. J. Mackie, S. E. Aidarous, S. A. Mahmoud and J. S. Riordon, "Design and performance evaluation of a packet voice system", IEEE Trans. Veh. Tech., 1983, VT-32, pp. 158-168

75. N. Shacham, E. J. Craighill and A. A. Poggio, "Speech transport in packet-radio networks with mobile nodes", IEEE Journ. on Selected Areas in Commun., 1983, SAC-1, 1084-1097

76. S. L. Casner, E. R. Mader and E. R. Cole, "Some initial measurements of ARPANET packet voice transmission", Proc. Nat. Telecommun. Conf. 1978, Vol. 1, pp. 12.2.1-5

77. J. W. Forgie, "Voice conferencing in packet networks", Conf. Rec. Int. Conf. Commun., 1980, pp. 21.3.1-4

78. G. Barberis, M. Calabrese, L. Lambarelli and D. Roffinella, "Coded speech in packet-switched networks: models and experiments", IEEE Journ. on Selected Areas in Commun., 1983, SAC-1, pp. 1028-1038

79. Numerical Algorithms Group, Fortran Library Manual Mark 9, Volumes 4 and 5, 1982

80. V. R. Dhadesugoor, C. Ziegler and D. L. Schilling, "Delta modulators in packet voice networks", IEEE Trans. Commun., 1980, COM-28, pp. 33-51

81. D. Bear, "Principles of telecommunication-traffic engineering", Peter Peregrinus, Stevenage, 1976

82. J. Raulin, G. Bonnerot, J. Jeandot and R. Lacroix, "A 60 channel PCM-ADPCM converter", IEEE Trans. Commun., 1982, COM-30, pp. 567-573

83. N. Alvey, N. Galwey and P. Lane, "An introduction to GENSTAT", Academic Press Inc. (London) Ltd., 1982

84. J. M. Griffiths (Editor) "Local Telecommunications", Peter Peregrinus, Stevenage, 1983