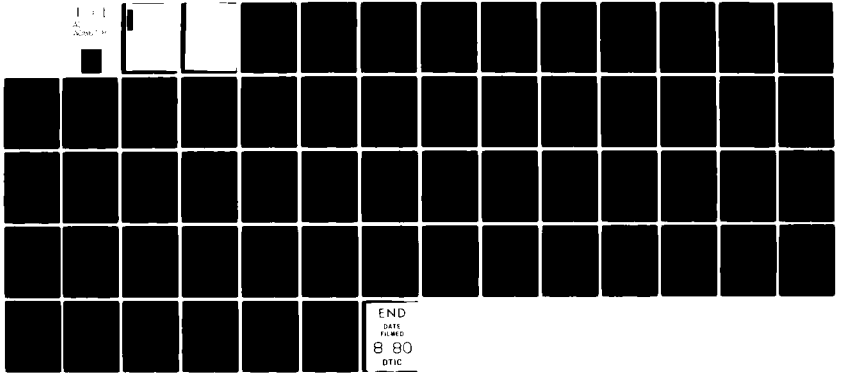
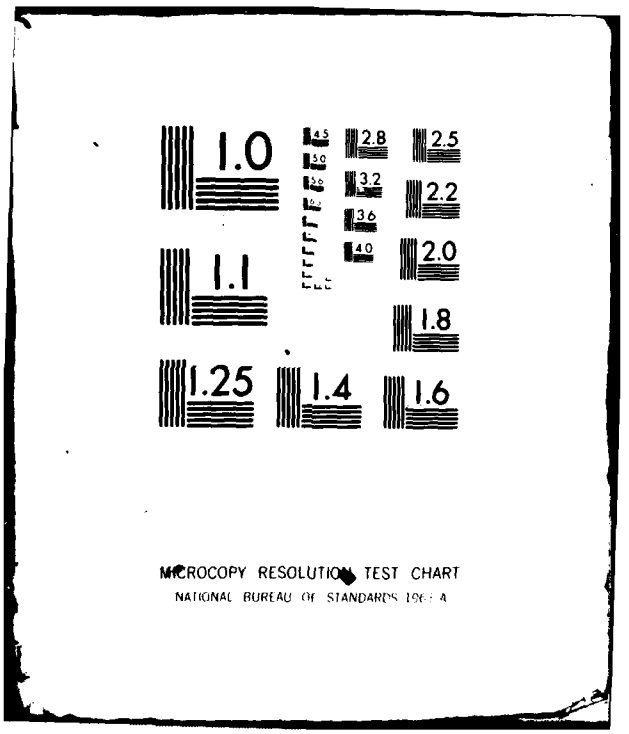
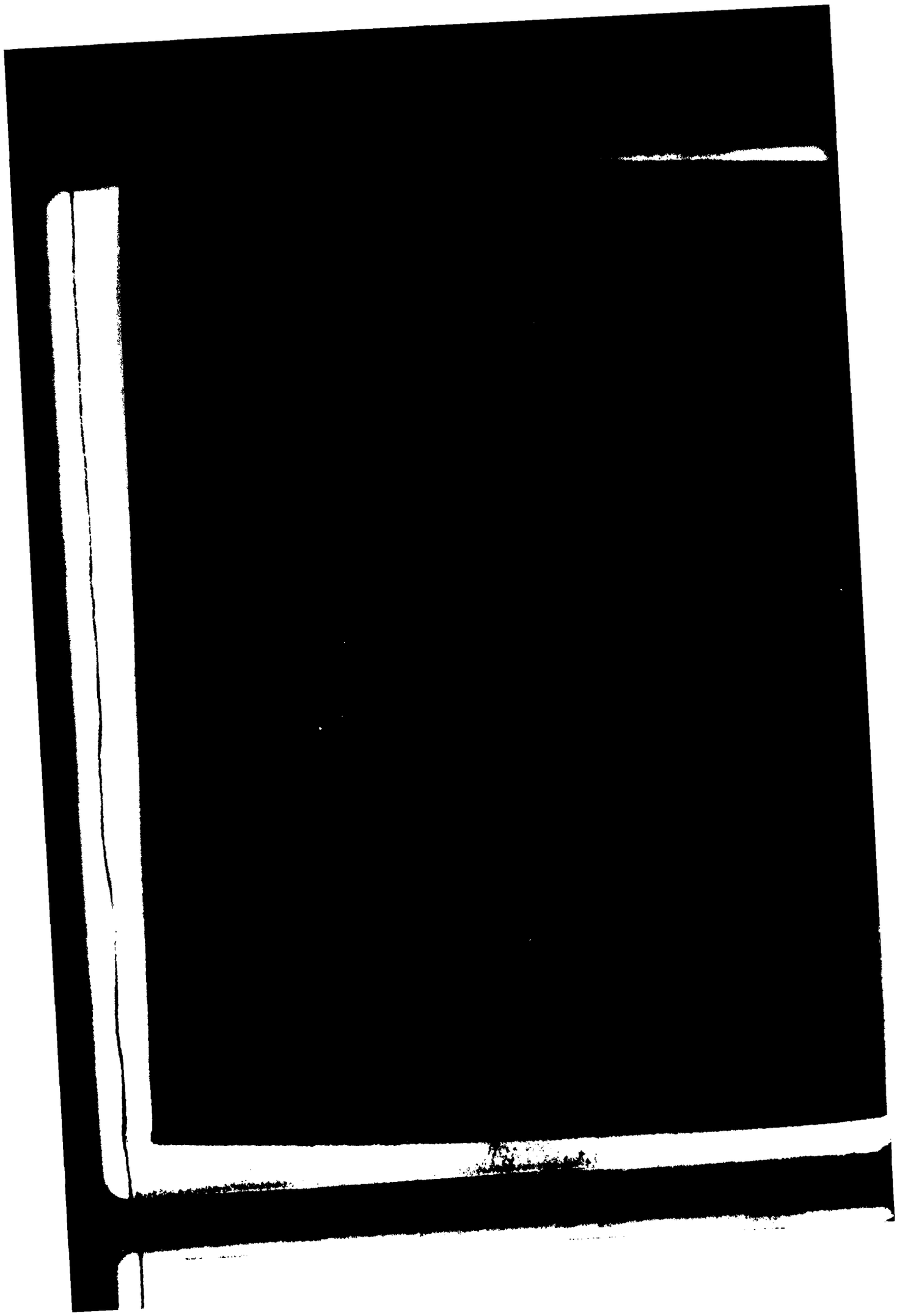
AD-A086 768 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/6 17/2
NETWORK SPEECH SYSTEMS TECHNOLOGY PROGRAM.(U)
SEP 79 C J WEINSTEIN F19628-80-C-0002
UNCLASSIFIED ESD-TR-79-313 NL



END
DATE FILMED
8 80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A



(12)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

NETWORK SPEECH SYSTEMS TECHNOLOGY PROGRAM

ANNUAL REPORT
TO THE
DEFENSE COMMUNICATIONS AGENCY

1 OCTOBER 1978 – 30 SEPTEMBER 1979

ISSUED 22 MAY 1980

LEXINGTON MASSACHUSETTS

ABSTRACT

This report documents work performed during FY 1979 on the DCA-sponsored *Network Speech Systems Technology Program.* The areas of work reported here are: (1) a switching and multiplexing study dealing with the analysis of buffered voice and data multiplexers and with a proposed technique for exploitation of speech activity detection to increase channel efficiency in a multi-link hybrid network; (2) a Demand-Assignment Multiple Access (DAMA) study focusing on prediction and buffering of digital speech streams for improved speech multiplexing performance on a broadcast satellite; (3) efforts in secure voice conferencing including protocol test and evaluation, support of the Secure Voice and Graphics Conferencing (SVGC) Test and Evaluation Program, and study of potential future-generation conferencing system strategies. Progress during FY 79 in experiment definition and planning for the Experimental Integrated Switched Network (EISN) test bed being developed under joint DCA/DARPA sponsorship is reported separately in a Project Report.

CONTENTS

NETWORK SPEECH SYSTEMS TECHNOLOGY

I. INTRODUCTION AND SUMMARY

This report documents work performed during FY 1979 on the DCA-sponsored Network Speech Systems Technology Program. The areas of work reported are: (1) a switching and multiplexing study dealing with the analysis of buffered voice and data multiplexers and with a proposed technique for exploitation of speech activity detection to increase channel efficiency in a multi-link hybrid network; (2) a Demand-Assignment Multiple Access (DAMA) study focusing on prediction and buffering of digital speech streams for improved speech multiplexing performance on a broadcast satellite; (3) efforts in secure voice conferencing including protocol test and evaluation, support of the Secure Voice and Graphics Conferencing (SVGC) Test and Evaluation Program, and study of potential future-generation conferencing system strategies. Progress during FY 79 in experiment definition and planning for the Experimental Integrated Switched Network (EISN) test bed being developed under joint DCA/DARPA sponsorship is reported separately in a Project Report.* The switching/multiplexing, DAMA, and experiment planning efforts represent continuations of work reported in the previous Annual Report.[1] A comprehensive summary of previous efforts in voice conferencing through mid-FY 79 has been reported separately.[2]

Section II deals with the switching/multiplexing study, which is a follow-on to the previous year's voice/data integration study.[2] A new analytic solution to the delay-vs-throughput behavior of a buffered speech multiplexer is presented. This result allows one to avoid time-consuming simulations and offers extendability to multi-link networks and to data-delay analysis in hybrid (combined circuit and packet) networks. A potential technique for achieving efficient Time-Assigned-Speech-Interpolation (or TASI-like) operation in multi-link hybrid nets is described, and issues to be resolved in the implementation of this technique are outlined.

The DAMA study is the subject of Sec. III. Focus is on the problem of taking full advantage of Speech Activity Detection (SAD) to achieve efficient channel utilization in a situation where a large number of ground stations, with a relatively small number of voice users at each ground station, share a broadcast satellite channel. This configuration is particularly relevant to a cost-effective, satellite-based architecture proposed for the next-generation AUTOVON.[3] The previous report[1] derived a talker activity prediction algorithm and a trade-off between buffering and TASI advantage, and proposed that prediction and buffering be used in conjunction with a dynamic DAMA algorithm to achieve efficient multiplexing. A multi-node satellite system simulation which combines these techniques and provides performance results has been developed and is described here. The results indicate that prediction and buffering can provide substantial improvement in system performance by significantly reducing the speech cutout fraction for a given system load.

The voice conferencing work, reported in Sec. IV, consists of three parts. First, conferencing protocol tests, conducted in the Lincoln Laboratory test facility with Air Force Subjects,

*"Experiment Plan for the Wideband Integrated Network – Supplement 1," to be published as Lincoln Laboratory Project Report EWN-1, Supplement 1 (originally issued as Wideband Working Note No. 5, 10 December 1979).

are reported briefly; comparisons to previous results[2] are discussed. Second, efforts in support of the Secure Voice and Graphics Conferencing (SVGC) Test and Evaluation Program, for which Naval Oceans Systems Center (NOSC) is the lead organization, are reported. Finally, an advanced secure conferencing systems study, which deals with classification of conferencing systems, interoperability issues, and future system alternatives, is reported.

Experiment definition and planning for EISN are covered in a separate Project Report that represents a supplement and update to last year's[1] preliminary experiment plan and includes more detailed planning information with respect to test-bed development, system validation, and advanced systems experiments. In the advanced systems experiment planning area, specific attention has been paid to the relevance of DAMA and integrated satellite/terrestrial experiments to proposed future DCS[4] and AUTOVON[3] architectures. The experimental network and associated planning activities are jointly sponsored by DCA and DARPA, and the experiment plan supplement reflects this joint sponsorship.

II. SWITCHING AND MULTIPLEXING STUDY

This section reports on efforts during FY 79 in the study of switching and multiplexing techniques for networks which can accommodate voice and data. It represents a follow-on to the FY 78 voice/data integration study[1,5] which focused on data traffic performance in hybrid multiplexers. Here, an analytical solution to the queueing behavior of a buffered speech multiplexer is developed. Comparison with simulation results and a discussion of extensions to multi-link and hybrid systems are included. In addition, a proposed systems approach for achieving efficient TASI operation in a multi-link hybrid network is discussed.

A. SWITCHING AND MULTIPLEXING

As we extend our study of communications networks, we must deal with traffic of widely differing statistical properties, sometimes mixed on the same link. In particular, the correlation times of the traffic sources vary greatly: a few milliseconds for packets from some interactive data terminals; a second or two for talkspurts in a TASI scheme; hundreds of seconds for call duration in a line-switched-voice scheme; even longer for large file transfers. In many cases, one of which is pointed out later in this section, it is misleading to model the traffic as a Poisson arrival process. Since anything but a Poisson model usually leads to analytical difficulties, our approach has been one of computer simulation of schemes for packet-switched speech communications.

Although their results are useful, the simulations require hours of computer time for the queueing behavior at a single node (making them hard to extend to the multi-node case). Also, they do not provide the insight into the congestion process that might be afforded by an analytical treatment. This section will describe an analytical approach to the queueing process in a buffered speech multiplexer and will compare the results to those of simulations and of a simplified Poisson-arrival model. It ends with some suggestions for extending the analysis to traffic combining speech and data. The analysis is sketched rather briefly here; more detail will be available in a forthcoming document.[6]

The problem differs in one important respect from classical queueing problems, namely in the statistical nature of the arrival process. Classical queueing analysis considers customers (packets) whose inter-arrival times are independent of each other, and the entire packet is assumed to arrive simultaneously. That is a reasonable model for data traffic such as that in Fig. II-1(a), where the tolerable delay D_d corresponds to a queue many packets long. It is not such a good model for traffic from a small number of speakers producing packets in the manner of Fig. II-1(b). If packets are considered customers, their inter-arrival times are far from independent. If talkspurts are considered customers, then the arrival of a single customer can change the queue length by an amount on the order of the tolerable delay D_s. Such a model is too coarse.

The model chosen treats speech as a continuous quantity, as in Fig. II-1(c). It preserves the correlation of speech over a talkspurt, but it does not model delays caused by the stochastic phasing of speakers. As an example of the latter, consider a link with capacity for 5 speakers. If exactly 5 speakers are in talkspurt, the continuous model says no queue would develop. But if it happened that all 5 produced packets simultaneously, the last-processed of these packets would be delayed 4 packet-processing times. In practical design problems, the delays caused by the stochastic nature of the lengths of talkspurts and silences will dominate the delays caused by the asynchronous phasing of speakers.

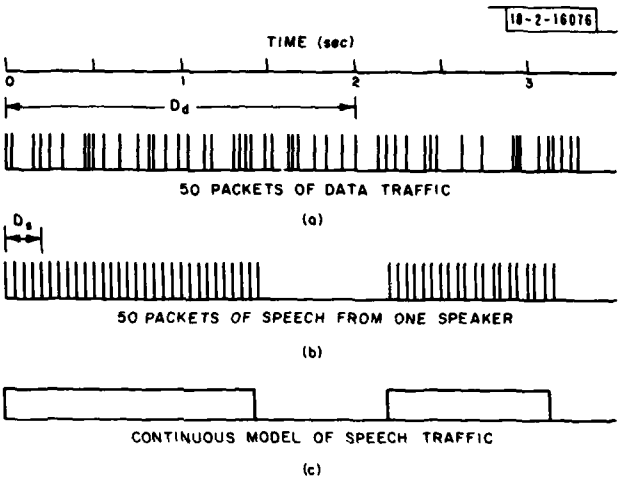


Fig. II-1(a-c). Traffic-arrival models.

The unit of speech used in the model is the speaker-second. A single speaker in talkspurt produces speech at a rate of 1 speaker-sec/sec. The number of off-hook callers is fixed at M. The link capacity is c speaker-sec/sec. The limit on queue length is Q speaker-sec, with overflowing speech being discarded. The average talkspurt and silence lengths are μ^{-1} and λ^{-1}, respectively.

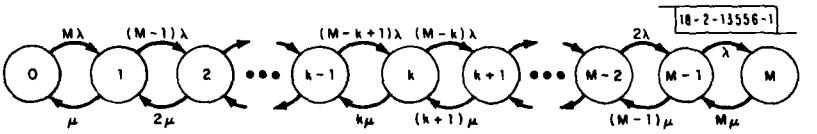


Fig. II-2. Speaker activity model.

The analysis gives a procedure for computing $q(x, a)$, the probability that a randomly timed snapshot will show a speakers in talkspurt and a queue of x speaker-sec or less. It uses the model of speaker activity shown in Fig. II-2, which appeared in last year's report[1] in connection with the prediction of speaker activity. The function $q(x, a)$ must obey the differential-difference equation

$$(M - a + 1)\lambda\, q(x, a-1) + (a+1)\mu\, q(x, a+1) - [(M-a)\lambda + a\mu]\, q(x, a)$$

$$= (a - c)\,\frac{\partial}{\partial x}\, q(x,a) \qquad \begin{matrix} 0 \leqslant a \leqslant M \\ 0 < x < Q \end{matrix} \qquad \text{(II-1)}$$

and the boundary conditions

$$q(Q-, a) = P(a) \qquad 0 \leqslant a < c \qquad \text{(II-2)}$$

$$q(0, a) = 0 \qquad c < a \leqslant M \quad . \qquad \text{(II-3)}$$

In Eq. (II-2), $P(a)$ is the probability of a randomly timed snapshot showing a active speakers, and $q(Q-, a)$ is the probability that there are a active speakers and the queue length is less than Q, i.e., the queue is not overflowing. $P(a)$ is simply a binomial distribution with mean $M\lambda/(\lambda + \mu)$.

Equation (II-1) was solved by a separation-of-variables technique as a linear sum of eigenfunctions

$$q(x, a) = \sum_{k=0}^{M} w_k A_k(a)\, e^{s_k x} \quad . \tag{II-4}$$

The analysis provides a quadratic equation for each of the eigenvalues s_k, and a finite procedure for calculating values of the associated function $A_k(a)$. The boundary conditions, Eqs. (II-2) and (II-3), provide a set of linear equations from which to compute the w_k. Once $q(x, a)$ is known, it is straightforward to compute such quantities as average delay.

Unfortunately, the algorithm for computing $q(x, a)$ is sufficiently complex that a computer program is needed in most cases of interest. Although the running time is much less than that of the equivalent simulation, a simple approximate expression for average delay is still needed. In the case of an unlimited queue ($Q \to \infty$), such an expression was found from the lumped-speaker model, which uses a reduced number N of speakers, each producing speech at a rate exceeding 1 speaker-sec/sec. The average rate of speech activity for all speakers combined and the auto-correlation function for that activity are kept the same as in the original, or unit-speaker, model. The formula for average delay is

$$d = \begin{cases} \frac{q}{\mu c}\left[\frac{q}{1-\rho} - \frac{c}{N}\right]\rho^{N-1} & 1 - q \leqslant \rho < 1 \\ 0 & 0 \leqslant \rho < 1 - q \end{cases} \tag{II-5}$$

where $q = \mu/(\lambda + \mu)$ and $\rho = M\lambda/c(\lambda + \mu)$. The symbol ρ has the usual meaning in queueing theory, namely the average arrival rate expressed as a fraction of the transmission capacity. N is an integer depending on M and ρ. It may be found from the chart in Fig. II-3. For example, let $\lambda = \mu$, so that $q = 1/2$. Let $M = 8$ and $c = 5$. Then, $M(1 - q)/q = 8$ and $c/q = 10$. Using these two coordinates in Fig. II-3, we can read a value of $N' = 4.6$. We would choose N to be the next smallest integer, 4. Alternatively, we could have found the same value of N using $c/q = 10$ and $\rho = 0.8$ as coordinates in Fig. II-3.

Figure II-4 shows the average delay experienced by incoming speech as a function of the channel utilization ρ. The capacity c is 5 speakers, and the number of off-hook callers varies from 6 through 10. For all five curves, the average silence and talkspurt are 1.34 and 1.23 sec, respectively. There was no limit to the queue length. In addition to values generated by the unit-speaker model and the lumped-speaker approximation, there are two (dashed) lines representing computer simulations of the same multiplexer. One used an exponential distribution of talkspurt and silence lengths, such as assumed in the models. The other used empirical distributions based on those of Brady.[7] The close agreement among the three lower curves gives confidence that the analysis was done without serious errors and that the simple formulas of the lumped-speaker approximation, such as Eq. (II-5), have merit. Their separation from the "Brady" curve shows that delay is sensitive to the shape of a talkspurt/silence distribution, not just to its mean. The fifth curve, labeled "Poisson Arrival of Packets," will be discussed below.

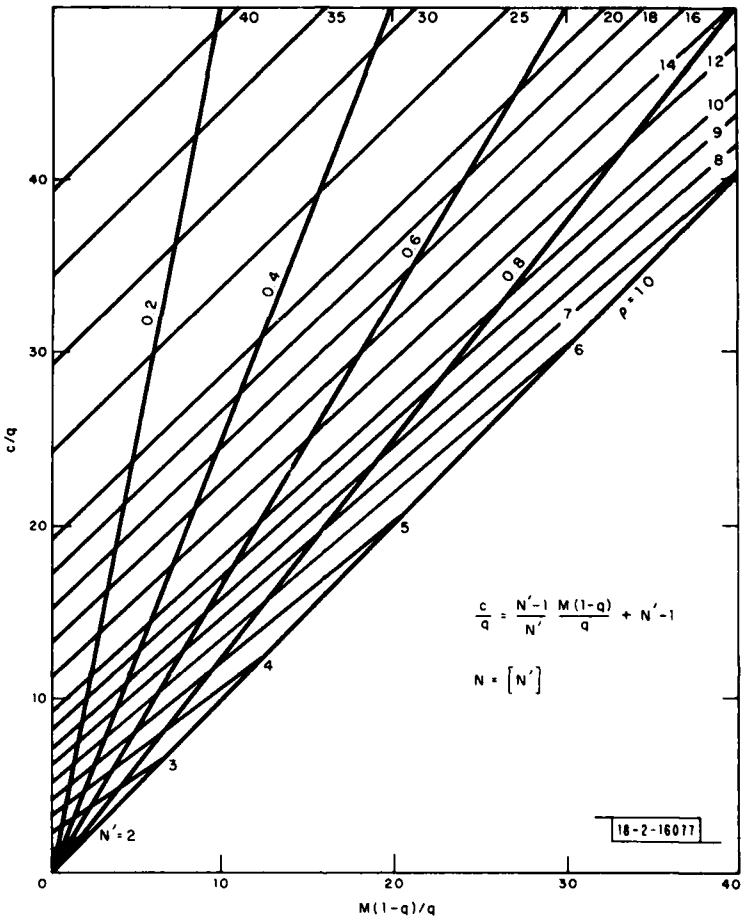


Fig. II-3. Chart for finding number of lumped speakers.

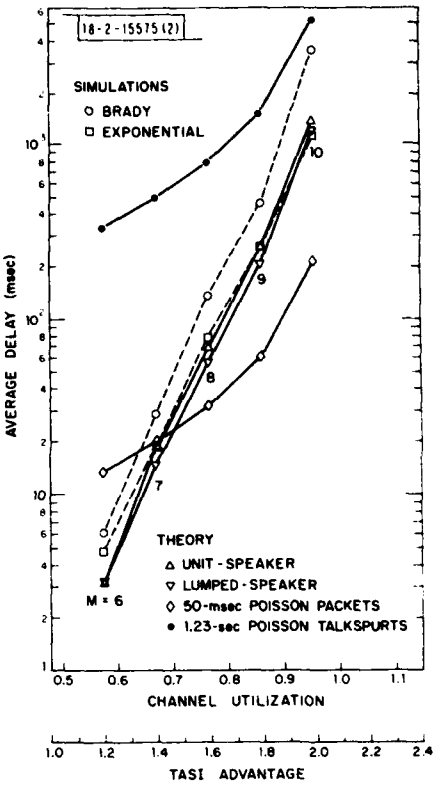


Fig. II-4. Delay vs TASI advantage.

It is instructive to compare the delay predicted by a model that incorporates the correlation of speech activity over a talkspurt, as in Eq. (II-5), with one that treats the arrival of speech packets as a Poisson process.[8] The latter gives the following asymptotic expression for average delay as the utilization ρ approaches unity.

$$\bar{d} \sim \frac{T_p}{c} \frac{1}{1-\rho} \tag{II-6}$$

where T_p is the length of a packet in seconds of coded speech. The asymptotic form of Eq. (II-5) is

$$\bar{d} \sim \frac{T_t}{c} \frac{q^2}{1-\rho} \tag{II-7}$$

where $T_t = \mu^{-1}$ = average length of a talkspurt in seconds.

Except for the factor q^2 which is roughly 1/4, Eq. (II-7) says that, for extremely heavy traffic (i.e., traffic producing average queues many talkspurts long), the speech arrivals can be treated as a Poisson process, but one in which the customers are talkspurts – not packets. But such a long queue corresponds to a delay that is usually intolerable for speech. Therefore, neither Eq. (II-6) nor (II-7) should be used as an approximation to Eq. (II-5). As an example of the error in Eq. (II-6), it is plotted in Fig. II-4 for T_p equal to 50 msec of speech per packet.

Although the above analysis is for voice traffic only, the process of deriving Eq. (II-1) gave some insight into how to treat the case of combined voice and data traffic feeding a buffered

multiplexer. Equation (II-1) was modified to include the effects of data traffic consisting of the Poisson arrival of packets whose sizes are independent and identically distributed. Although the modified equation was not solved, it looked obvious that the packet-arrival process could, in the right circumstances, be replaced by something even simpler than the Poisson-arrival process, namely a steady flow of traffic at the average rate. The "right circumstances" are that the average packet length be small compared with the average queue length. To analyze the buffered multiplexing of talkspurts and (short-enough) data packets, one would simply reduce the capacity c in (for instance) Eq. (II-5) by the average data rate (measured in equivalent speaker-sec/sec). The solution algorithm will still give the average queue length seen by a randomly timed observation. Whether the associated delay is suffered by speech or data depends on the queueing discipline. If, for instance, speech is allowed to get on-line ahead of data, and if the link can handle the maximum rate of speech, then all the delay would be borne by the data.

Although the explicit expressions (not given here) for eigenfunctions and eigenvalues in the above algorithm are valid only for binomially distributed speaker activity (such as that produced by a fixed number of independent speakers going into and out of talkspurt), the basic analysis in terms of eigenfunctions is valid for other activity distributions. In particular, it could be used (with standard numerical methods for finding the eigenfunctions) to analyze the combined multiplexing of line-switched voice and packet-switched data. Such multiplexing has been the subject of extensive simulations.[1,5] The details of the analysis described in this section will appear in a forthcoming paper.[6]

B. TASI IN MULTI-LINK HYBRID NETS

The Time-Assigned Speech Interpolation (TASI) system achieves an approximate 2:1 bandwidth improvement over conventional telephony by exploiting the silence intervals in normal speech for the transmission of other voice signals. In its familiar form, TASI is used over single links (e.g., submarine cables) in circuit switched networks that otherwise employ no statistical multiplexing. Since talkspurt and silence detection can only be performed on non-multiplexed speech streams, it is difficult to extend the concept to multiple-link systems without introducing ancillary control information for use by downstream TASI switches. Speech packetization formats afford a convenient means for conveying this information implicitly (i.e., packets are generated only during talkspurt) and, as a result, the concept of distributed statistical speech multiplexing has been promulgated mainly in the context of packet switching networks. Statistical multiplexing gains in hybrid nets have centered on the use of speech silence intervals for the transmission of packetized data traffic (Time-Assigned Data Interpolation, or TADI) since this has a much less complicated control implication.

This section outlines a technique for achieving TASI operation in hybrid nets. In brief, the packet handling capability is used for high-priority internode communication regarding the onset and termination of talkspurts, which, in turn, are transmitted in a circuit-switched format. The result is that end-to-end voice circuits are assigned separately to individual talkspurts, but along routes that are fixed for the duration of a conversation. There is potential in this system for improved bandwidth efficiency relative to packet nets, since the talkspurt/silence control overhead is amortized over entire talkspurts. On the other hand, voice cutout effects can be more severe than in a single link system and in this approach they will be concentrated at the beginnings of talkspurts instead of being distributed over the entire utterance, as they might be in a packet system.

1. System Description

Consider a multi-node net in which the wideband transmission links are organized into frames of N bits each, in a fixed TDMA fashion. A given voice connection C requires n_c of those bits in every link through which it is routed. The particular positions (not necessarily contiguous) of the bits are fixed for successive frame transmission on a given link, but they may differ from link-to-link along the route as shown in Fig. II-5. Small fixed delays (one frame) are introduced at the switching nodes in order to accommodate these bit position assignments.

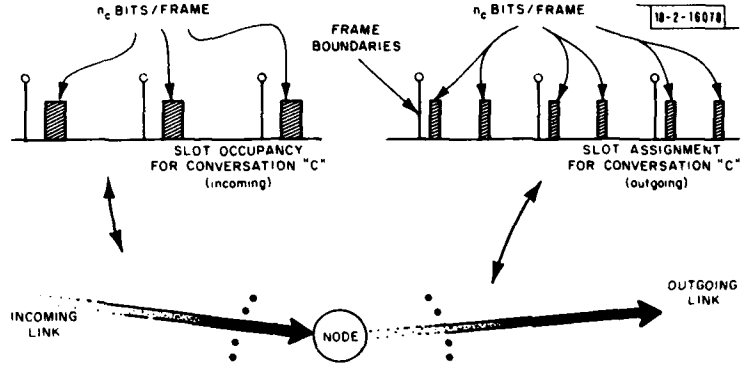


Fig. II-5. TDMA frame organization.

Voice routes through the net are determined at dial-up time in accordance with appropriate circuit routing procedures. However, time-slot assignments on the links along a given route are made separately for each talkspurt, and the slots that are occupied are released when the talkspurt terminates. They then become available for use by the talkspurts of other connections or for data packet transmissions. We assume for the time being that in each link all slots in a given frame that are not currently supporting talkspurt transmission can be aggregated and used for data packet transfers. The rate at which packet queues can "play out" into the network links will thus vary both with time and location in the net, as a function of instantaneous link voice load. Major issues in the design of a system of this type include:

(a) A control structure for assigning link bandwidth to talkspurts and releasing it when the talkspurt is over.

(b) Trade-offs between delay, TASI advantage, and cutout fraction that might be built into the nodal switching strategies.

(c) The bandwidth efficiency that can be achieved, given that talkspurts are typically much longer than speech packets, and can therefore amortize control overhead more effectively.

(d) Mechanisms for aggregating non-voice slots and using them for packet data traffic, on a dynamic, frame-by-frame basis.

(e) Robustness properties with respect to link transmission errors or node failures, and associated recovery problems.

(f) Implications for Voice and Data Security.

These are elaborated upon below.

2. Talkspurt Control

Talkspurt/silence detection is performed at the node of origin or at the transmitting voice terminal. Talkspurts are modeled as constant bit-rate entities that require synchronous, clocked digital connectivity from source to destination. No connection is needed for silence. At the onset of a talkspurt, the originating node or terminal creates a Talkspurt Control Packet (TCP) whose job it is to announce the appearance of the talkspurt to successive nodes along the selected route. Since timing will be critical for voice, chances are that TCPs may be treated preferentially with respect to other data traffic. A TCP might contain

(a) The ID of the virtual voice circuit to which it refers, and

(b) The TDMA slots or frame bit positions within which the talkspurt will be arriving on the incoming trunk.

The job of a switching node upon receiving such a TCP is to refer to its routing tables and select the appropriate outgoing link for this talkspurt, identify a set of unused slots or bit positions in the outgoing TDMA frame into which the incoming bits will be transferred, modify item (b) of the TCP accordingly, and send the TCP packet ahead to the next node in the route. It also establishes a semi-permanent circuit connection between the appropriate incoming and outgoing slots. A similar TCP is generated at the originating node at the completion of a talkspurt, for the purpose of releasing the slots that had been previously assigned. We thus add to the TCP contents a new item, i.e.,

(c) Start or end of talkspurt.

3. Performance Trade-Offs

The simple scenario described above will work as long as intermediate nodes can find the required number of free outgoing slots to accommodate newly arriving talkspurts. This will clearly not be the case at all times, and some performance compromises will result. Three possibilities suggest themselves, i.e., a node can introduce cutout, add delay, or effect some combination of both. For the cutout case, the node simply ignores the incoming voice stream until it can forward it properly. The TCP is held until a suitable outgoing circuit is identified, and when that occurs normal operation is resumed. The fact that a portion of the beginning of the talkspurt has been lost due to cutout need not be communicated to succeeding nodes.

The severity of the cutout phenomenon can be significant, especially since it can happen more than once to the same talkspurt at different nodes along the route. One way to mitigate the effect is for each node to buffer (instead of discard) its incoming talkspurts until outgoing circuits become available. Buffer delays introduced at successive nodes are additive and remain in effect for the entire talkspurt duration. Since excessive speech delays can be as undesirable as too much cutout, a balance between these two approaches might be in order. For example, one could add delay without introducing cutout until a predetermined maximum is

reached, and then impose cutout without additional delay. Since both delay and cutout are additive as a talkspurt carves its way from one node to the next, it might be reasonable to include the following control field in the TCP:

(d) Accumulated delay.

This field would be modified by intermediate nodes that introduce delay, and would be used by succeeding nodes as a guide in deciding whether to apply delay or cutout if forwarding circuits are not immediately available.

4. Bandwidth Efficiency

A very appealing aspect of this system is the enormous potential that exists in the average talkspurt for amortizing the bandwidth needed for TCP transmissions. If one thinks of a talkspurt as being a packet with a header (TCP) and a trailer (another TCP), and compares this with more conventional voice packets, the following emerges:

	16-kbps Voice		2.4-kbps Voice	
	Packet	Talkspurt	Packet	Talkspurt
Time Interval	20 msec	1.5 sec	40 msec	1.5 sec
Data Portion (bits)	320	24,000	96	3,600
Header Portion (bits)	64	2 × 64	32	2 × 64
Efficiency (D/D + H) (percent)	83	99	75	97

In the above, abbreviated (32-bit) headers are assumed for the narrowband packet speech example, along with a 40-msec (two parcels) packetization interval. TCPs are assumed to require the same number of bits as a nonabbreviated voice packet header, with two of these required for each talkspurt. The potential for high efficiency in the proposed system is obvious. In addition, if forward error correction is needed for control robustness purposes, it can be applied to the TCPs with very little loss in overall efficiency.

5. Data Transmission

As in any integrated voice/data scheme, the objective here is to use all the bandwidth that is not committed to talkspurt transmissions, for data packet transfers. Since packets will generally be of variable length, it would be nice to separate packet boundary and/or header issues from TDMA frame considerations. Referring to Fig. II-6, we note that if we ignore

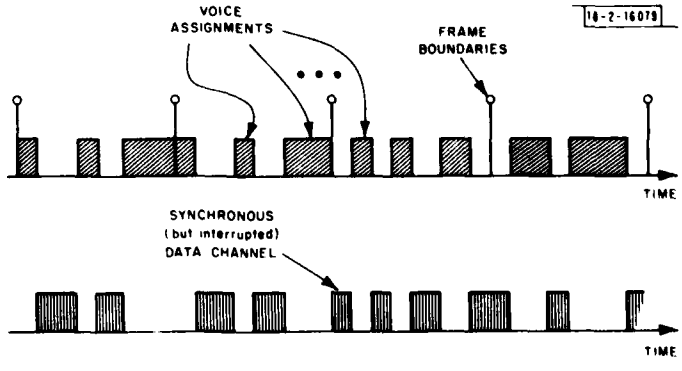


Fig. II-6. Bandwidth allocation scheme.

frame boundaries and simply "blank-out" those slots that are currently being used for voice, the result is the logical equivalent of a single synchronous clocked channel for data. This is basically the same notion that is applied in SENET[9] or SENET Virtual Circuit (SVC).

The "masking function" (i.e., the blanking of voice bits) will, in general, be different for every link in the net, and will change every time a TCP is sent along that link. TCPs, in turn, are simply packets that flow on this dynamically changing data channel. A reasonable requirement might be that TCPs bypass non-TCPs in packet queues in order to speed up the link bandwidth reallocation process.

6. Robustness Issues

Robustness is a major issue here, and no simple solution is obvious. The basic problem is to assure that the two nodes at either end of a given link are always in agreement as to what bits belong to which user and for what purpose, and to provide simple and effective recovery mechanisms when they're not. A serious concern is whether a small set of transmission errors or nodal failures can wreak havoc with all the users on a link, or if the damage is restricted only to those users whose bits were directly affected.

Suppose a start-of-talkspurt TCP is received in error. That talkspurt will clearly be aborted and the listener will hear prolonged silence. However, when the next talkspurt for that conversation appears and assuming its TCP is not destroyed, the circuit will be re-established. This example illustrates an interesting point; namely, that in this system the talkspurts are somewhat like packets, and the loss of one need not be felt by any others. A similar result follows for an error in a trailing TCP, which can in fact be viewed as the header for a silence "packet."

The robustness problem for data is less benign than for voice. With reference to Fig. II-6, if a single voice TCP for any user on the link is received in error, a block of data bits corresponding to the associated voice circuit will be erroneously added or deleted from the data stream in every succeeding frame. Since TCPs are transmitted as data, there is the added danger that an error in one TCP can spawn errors in succeeding ones, and cause even more catastrophic global sync problems. No simple answer is offered here. We simply observe that if all the voice slots are "compacted" to the beginning of each frame, and if a count of the total voice allocation is sent with the frame, the data channel robustness problem is eased considerably. On the other hand, a single TCP error can now affect many voice users in many successive TDMA frames, since slot allocations for talkspurts are no longer fixed from frame-to-frame. A fairly clean answer might be to avoid compacting the speech, but to quantize the slot sizes so that they all have M bits each. The leading bit in each slot could then flag whether it was carrying voice or data. Locations of the data bits are then unambiguously identified in each frame, and isolated mistakes cannot cause problems in future frames. The cost here is an a priori efficiency limit of $(M-1)/M$, plus whatever costs are associated with voice bit rates that are not exactly commensurate with an integer multiple of $(M-1)$ bits per TDMA frame.

7. Encryption Properties

One problem with secure packet speech is the potential for added overhead that comes with packet-oriented encryption methods such as the BCR Technique.[10] On the other hand, circuit-switched voice requires sync only at the start of an utterance. In our talkspurt-oriented TASI method, stream-oriented techniques could be used with appropriate sync provided at the start of each talkspurt interval. Methods exist in which crypto-sync is implicit in any contiguous

record of given length. In other words, one passes the received sequence through an unsynchronized (but properly keyed) crypto decoder, and after producing, say, 64 bits of unintelligible output, the remainder of the record is decoded properly. This technique is appealing in the context of a TASI system in which cutout can occur only at the beginning of a talkspurt. In fact, it appears that with this approach the fact that a voice stream may be encrypted could be transparent to the switching nodes.

8. Summary

The concept described here is very similar to the SENET and SVC notions in that a framed TDMA organization is used over wideband network links. A difference, however, is in the accommodation of TASI-like operation in the context of a multi-link network, and the attendant increase in flexibility and bandwidth efficiency that this affords. We note that the talkspurt-oriented distributed TASI technique can very simply degenerate to a pure SENET system by viewing entire conversations as single talkspurts. If speech activity detection is used, but the same link-slot assignments are kept for all talkspurts from a given speaker, the system behaves like SVC (TADI).

An important point worth emphasizing for this system is that it appears to be compatible with advanced voice-flow-control notions. For example, if embedded coding were of interest several parallel circuits could be established from source to destination, each carrying an embedded component of the voice stream. TCPs for these component circuits could contain their relative priority numbers, and switching nodes could refuse to connect low-priority streams when overload conditions exist. Unlike a packet-oriented embedded coding scheme, this one would alter bit rates on a talkspurt-by-talkspurt basis, or impose greater cutout on the less important bits. Perceptual problems due to instantaneous rate changes in mid-utterance would probably be avoided due to the built-in synchronization of these events with the talkspurt/silence boundaries.

The ideas outlined in this section are rather preliminary, and further analysis and simulation, as well as implementation exercises, will be needed before we can determine whether the concept will work in a practical sense.

III. DEMAND-ASSIGNMENT MULTIPLE ACCESS (DAMA) STUDY

A. INTRODUCTION

Recent trends in integrated voice/data communications network design have begun to favor configurations with a large number of small nodal switches serving small local user groups and with heavy reliance on broadcast satellites for transmission capacity.[3,4,11,12] Satellite channel capacity is an expensive commodity in such a system, and flexible DAMA schemes[13] must be relied upon to allocate this commodity efficiently according to fluctuating demands from the various earth terminals. Channel utilization can be significantly enhanced through the application of TASI (Time-Assigned Speech Interpolation)[14] or DSI (Digital Speech Interpolation)[15] wherein off-hook voice callers occupy channel capacity only during talkspurts and not during silence periods. Since talkers in conversation are typically silent more than 50 percent of the time, the potential capacity saving (generally referred to as the "TASI advantage") is greater than a factor of two. However, with standard TASI or DSI systems, achievement of the full potential TASI advantage requires that a large number of talkers (typically 50 or more) be statistically multiplexed at a particular node. The configuration of concern here is a number of small earth stations or nodes, where the number of off-hook callers at each node is too small to achieve efficient TASI multiplexing by standard techniques, but where the aggregate number of callers sharing the satellite would be large enough for efficient multiplexing if all the users were located at one node. This section describes and evaluates a proposed approach for achieving efficient TASI-like multiplexing in this configuration. The approach presupposes a DAMA scheme such as Priority-Oriented Demand Assignment (PODA)[13] which allows stations to request and rapidly obtain changes in their share of a Time-Division Multiple Access (TDMA) channel. The components of the approach are:

(1) Prediction[1] of the number of callers in talkspurt at each station ahead by the time (minimum of one satellite round-trip propagation delay) required to change channel capacity allocation, combined with requests for channel capacity on the basis of this prediction; and

(2) Variable-length buffering of speech at each station and trading of delay[16] for TASI advantage.[17]

The prediction algorithm and the basic trade-off between delay and TASI advantage were described in the previous report.[1] These ideas are refined and extended here, and a multi-node satellite simulation which combines the techniques and provides performance results is described. A separate Technical Note[18] describes the results in more detail and includes the prediction algorithm derivation.[1] The results, which are obtained primarily through simulation, show that this dual approach provides substantial potential improvement in TASI advantage over a system with channel allocations which cannot be changed rapidly enough to respond to talkspurt/silence variations and without variable buffering at the nodes. Note that the term "TASI advantage" is used generically here to refer to the ratio of the number of off-hook callers to the system channel capacity, where one unit of capacity is taken to be just sufficient to support one caller during talkspurt. The use of this term should not confuse the fact that the system under consideration is quite different from the classical TASI system.

B. SYSTEM MODEL AND STRATEGIES FOR IMPROVED TASI PERFORMANCE

The multi-node satellite-based communication system model of interest here is depicted in Fig. III-1. There are N ground stations, and the nodal processor at the i^{th} station supports M_i off-hook callers. Functions of the nodal processors include multiplexing and demultiplexing of local traffic, as well as the processing necessary to support the satellite demand-assignment algorithm. Application of Speech Activity Detection (SAD) and transmission only during talkspurt is assumed for each caller so that the transmission rate which must be supported at a node varies with the number of active talkspurts.

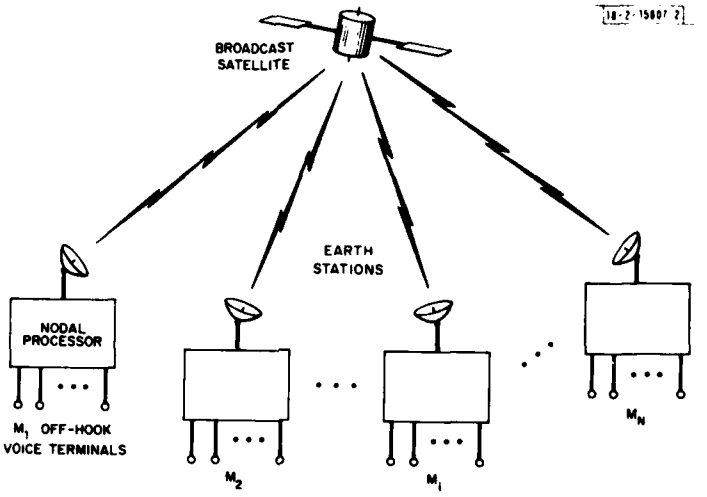


Fig. III-1. Configuration of multi-node satellite communications system.

The satellite channel capacity is assumed to be shared among the N nodes on a dynamically demand-assigned burst-TDMA basis. The capacity allocated to each individual station is assumed to be in the form of a variable-size "stream."[13] Once every T_S sec, the station has the opportunity to transmit a burst segment, where the maximum number of bits in this segment is the stream size. The DAMA algorithm is assumed to schedule these burst segments to be transmitted from the individual stations in a noninterfering and efficient manner. To minimize end-to-end delays it is desirable that T_S be kept as short as possible, on the order of 20 to 40 msec. It is not necessary that T_S match the frame interval which is associated with the TDMA pattern of the DAMA algorithm. Each segment, as shown in Fig. III-2, is assumed to contain a short reservation-request slot used to request changes in the size of the stream plus a set of speech slots each capable to carrying the amount of digital speech produced by one active voice terminal during one frame interval. For simplicity of simulation and analysis it is assumed here that all speech slots are of equal size, although the strategies and general nature of the results are not limited to this case. The nodal DAMA processor inserts into each reservation slot a request for a number of speech slots. This request number may vary slowly on the basis of variations in the number of off-hook callers M_i, or more rapidly on the basis of variations in the number of callers in talkspurt. In either case the request cannot be granted until it has been received by all stations, at least one satellite round-trip time (≈270 msec) after it is issued. A distributed

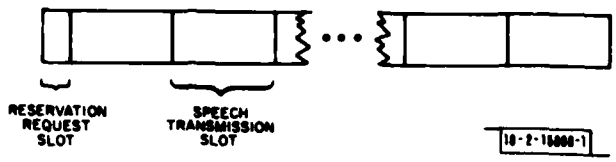


Fig. III-2. Format of burst segments transmitted in a single multiplexed speech stream. Reservation of stream capacity allows a node to transmit one of these segments every T_S sec. Stream size (number of speech slots) may be varied dynamically by changes in reservation request.

DAMA algorithm is assumed wherein the nodal processors at each station collect all requests and allocate speech slots based on identical, fair round-robin algorithms.[19] Generally, this channel allocation might occur synchronously with the frame structure of the DAMA algorithm. For convenience, in most of the simulation work here it has been assumed that allocations of channel capacity are updated every T_S sec, upon receipt of new reservation requests from all stations. Since T_S is typically much shorter than a satellite round-trip time, a "reservation pipeline" is formed wherein a number of reservation requests are propagating across the transmission link at any time.

The reservation response delay is not a significant limiting factor in responding to call initiations or terminations, where response times on the order of seconds are acceptable. However, achievement of efficient TASI multiplexing without nodal buffering in the case of a small number of callers per node requires that each node's slot allocation closely match the number of active (i.e., currently in talkspurt) speakers at that node. Because of the reservation response delay, the best each node can do to achieve this match is to issue slot requests based on a prediction of the number of talkers likely to be active one reservation response delay in the future. If, due to inaccurate prediction or limited overall satellite capacity, a node's slot allocation at a particular time becomes temporarily insufficient to support the instantaneous number of active talkspurts, then the overflow speech must either be discarded immediately or buffered (adding delay) at the node until transmission capacity becomes available or the buffer overflows. Both cases are considered here.

The strategies considered here can apply whether a packet-[20,21] or circuit-oriented[15] transmission format is utilized for the digital speech. As discussed in Ref. 8, the required control overhead for packet transmission can be reduced to a level comparable with that required to accommodate talker activity information in a circuit-switched system, if fixed virtual-circuit routing[22] is used for the packets. The primary remaining difference then becomes the flexible buffering allowed by the asynchronous nature of the packet system. However, a digital circuit-switched DSI system can also be augmented to include flexible buffering.[23] For convenience, the term "packet" will be used here to denote the speech information which is accommodated in a speech slot (see Fig. III-2), and speech buffers (when applied) are assumed to accommodate packet-sized units. However, it should be understood that the strategies and results are not limited to packet systems.

A block diagram of the functions to be carried out at each node is shown in Fig. III-3. The off-hook voice terminals transmit digital speech packets (during talkspurts only) through the multiplexer which feeds a multiplexed speech stream into the buffer. Once every stream interval T_S, the speech stream transmitter discharges from the buffer the number of packets that

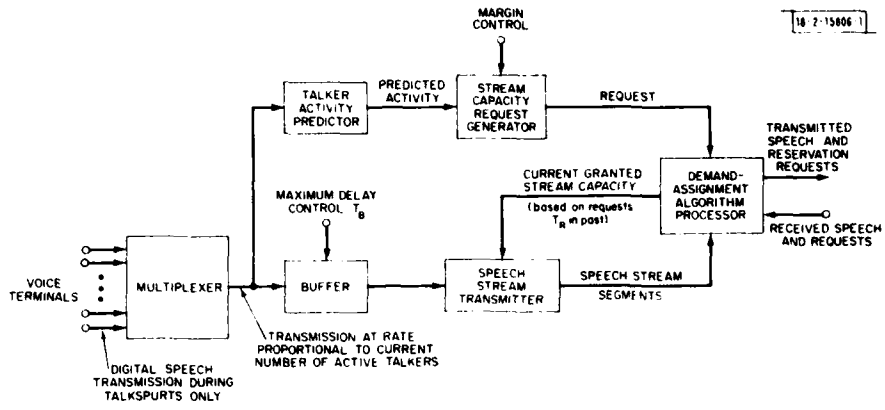


Fig. III-3. Block diagram of functions to be carried out at each node. Functions include multiplexing, buffering, transmission, prediction, reservation request generation, and execution of distributed demand-assignment algorithm.

can fit in its current stream segment. The maximum time T_B that a packet is allowed to remain in the buffer is set by a delay control parameter. Packets not discharged within this time are discarded. The cutout fraction, defined as percentage of packets discarded, is a key performance parameter in the system. Generally, cutout fractions less than 0.5 percent will be essentially unnoticeable to users, and cutout fractions on the order of 1 percent can be tolerated without significant degradation in user acceptability. This holds both for standard TASI systems[14] where cutouts occur only at talkspurt onsets, and for the system under consideration here where speech loss can be dispersed[24] through any part of a talkspurt. Minimal buffering delay (corresponding to a standard synchronous TASI or DSI system) results when the delay control parameter is set such that no packet remains in the buffer longer than one stream interval. Stream capacity is granted by the DAMA algorithm on the basis of the reservation requests most recently received from all stations and processed by the DAMA algorithm. The speech activity predictor observes the current number of active talkers in the multiplexed speech stream at the buffer input, and estimates the number of talkers likely to be active at a predict-ahead time T_P into the future. The request algorithm adds a margin M_A to this prediction to produce a reservation request for transmission along with the current speech frame. Margin is chosen (as discussed in more detail below) in order to balance optimally for a given overall satellite load, packet losses due to (1) insufficient reservation requests by the individual node, and (2) denial of reservation requests by the DAMA algorithm when the sum of all nodal requests exceeds channel capacity.

There are fundamental interrelationships in this system among the maximum buffer delay T_B, the required predict-ahead time T_P the reservation response time T_R, and the margin M_A. The growing uncertainty of predicting further into the future implies that M_A should increase with T_P. If T_B is set to zero, then T_P must equal T_R, which is lower-bounded by the satellite round-trip time. On the other hand, an increase in T_B has the effect of producing a corresponding decrease in T_P. In particular, if $T_B = T_R$ then no prediction is necessary because the speech can be buffered locally just long enough to make the desired change in channel allocation. The TASI performance of the overall satellite system for this special case will be as effective as if

all callers were multiplexed at a single node. The cost for obtaining this multiplexing performance is an added delay of T_R. The potential benefit of speaker activity prediction is to reduce this delay while still achieving efficient channel utilization.

The results of speaker activity prediction analysis[1,18] are summarized here. Consider M independent off-hook callers each alternating between talkspurt (active mode) and silence (inactive), and let n(x) denote the number of active talkers at time x. Assume a model of talkspurt and silence durations as exponentially distributed random variables with means μ^{-1} and λ^{-1}, respectively. This implies that n(x) is a Markov process[25-27] so that the optimum, least-squares predictor of $n(t+\tau)$ given the past history of n(x) prior to t is the conditional expectation $E[n(t+\tau)|n(t)]$. An explicit expression for this optimum predictor can be obtained as

$$\mu_j(\tau) \equiv E[n(t+\tau)|n(t)=j]$$

$$= \frac{M}{1+(\mu/\lambda)}\left\{1-\left[1-\left(1+\frac{\mu}{\lambda}\right)\frac{j}{M}\right]e^{-(\lambda+\mu)\tau}\right\} \quad . \tag{III-1}$$

Similarly, the mean-squared error of the optimum predictor is

$$\sigma_j^2(\tau) \equiv E\left\{[n(t+\tau)-\mu_j(\tau)]^2|n(t)=j\right\}$$

$$= \frac{M(\mu/\lambda)}{[1+(\mu/\lambda)]^2}\left[1-e^{-(\lambda+\mu)\tau}\right]\left\{1+\left[\frac{j}{M}\left(\frac{\mu}{\lambda}-\frac{\lambda}{\mu}\right)+\frac{\lambda}{\mu}\right]e^{-(\lambda+\mu)\tau}\right\} \quad . \tag{III-2}$$

Plots of Eqs. (III-1) and (III-2) for the case M = 10, $\mu^{-1} = \lambda^{-1} = 1.5$ sec are shown in Figs. III-4 and III-5. Inspection of these curves indicates that reasonably good prediction (±1 speaker rms error) can be realized for prediction times on the order of a round-trip satellite

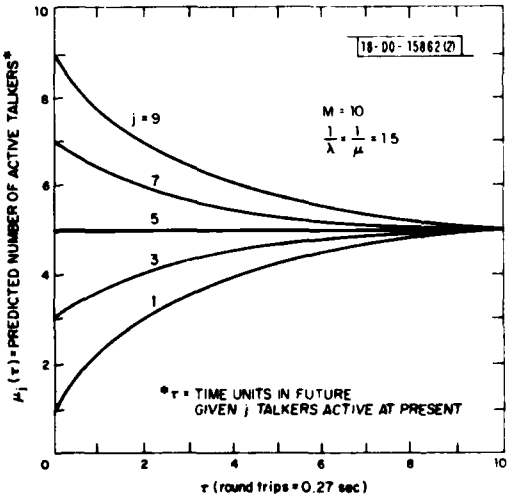


Fig. III-4. Optimum talker activity predictor.

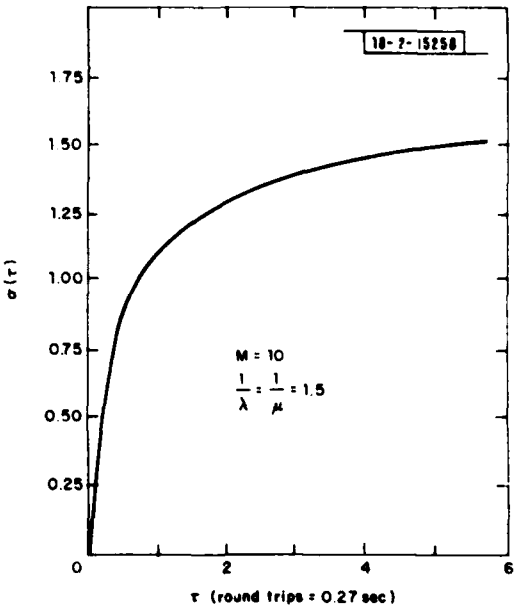


Fig. III-5. Prediction error of optimum talker activity predictor.

delay. Thus, the predictability of the speaker activity process, which results from time correlation due to typical talkspurt and silence durations, seems to offer potential for TASI advantage improvements along the lines discussed above. Simulation tests[1] have indicated that the above results are relatively insensitive to the Markov assumptions.

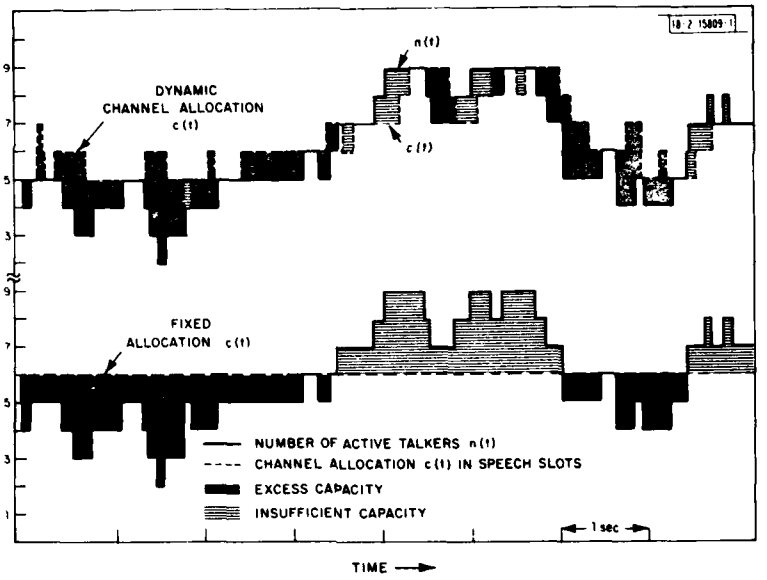


Fig. III-6. Sample functions of active talker process n(t) and channel allocation c(t) illustrating comparison of fixed allocation with prediction-driven dynamic channel allocation.

A graphical illustration of the potential benefits of speaker activity prediction is shown in Fig. III-6. The identical solid curves in the top and bottom parts of the figure represent an 8-sec segment of a talker activity time function n(t) obtained from simulation with M = 10 and exponential talkspurt/silence distributions. The average talkspurt duration was 1.23 sec, the average silence duration was 1.34 sec, the corresponding fractional talker activity p = 0.48, and the average talker activity $\overline{n(t)}$ = Mp = 4.8. The dashed curves represent channel allocation c(t) in slots/frame. The bottom part of the figure corresponds to a fixed allocation c(t) = 6. Dark gray areas indicate periods where n(t) $<$ 6 so that capacity is wasted. Light gray areas indicate periods where n(t) $>$ 6 and where, assuming no buffering, speech packets will be discarded. In the top curve, c(t) was obtained by predicting n(t) 280 msec into the future and adding sufficient margin so that the average $\overline{c(t)}$ = 6. It is apparent that the predictor, while far from perfect, does tend to track the changing talker activity. Both packet loss and wasted capacity are substantially reduced for this example with predictor-based allocations as compared with a fixed allocation with the same long-term average.

C. TASI PERFORMANCE IMPROVEMENTS WITH PREDICTION-DRIVEN *STREAM RESERVATIONS*

In this section, simulation results on system performance with prediction but without additional buffering delay at the individual nodes are presented. Referring to Fig. III-3, the constraint

applied is that speech packets which are not transmitted within the inter-packet interval T_S are discarded. The primary performance measure of the system is cutout fraction. A key issue was the selection of the correct margin level to minimize this loss fraction.

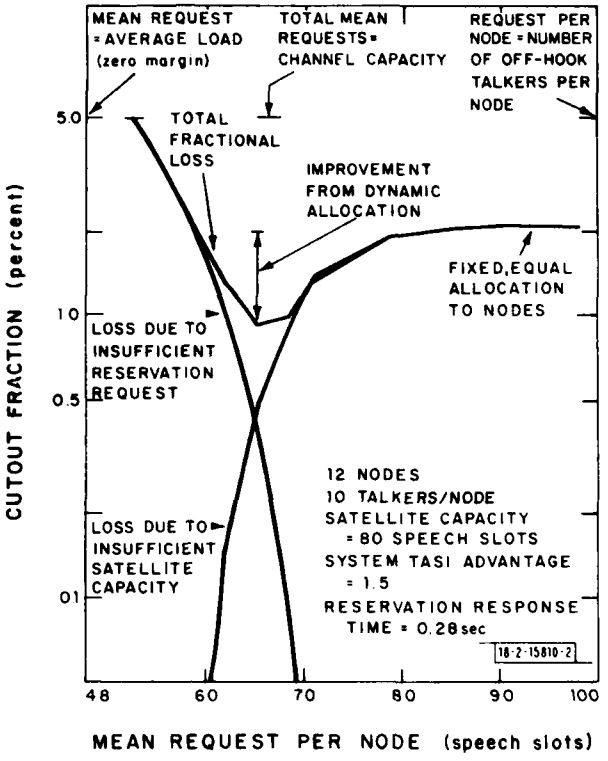


Fig. III-7. Example of effects of margin and of potential performance improvement with dynamic allocations.

An illustration of the nature of the simulation results as well as a discussion of the key system variables can be carried out in the context of the example shown in Fig. III-7. Here the variation of packet loss with margin is presented for the case of N = 12 nodes, M = 10 off-hook talkers per node, and an overall satellite capacity assumed to be sufficient to accommodate 80 voices in talkspurt. The system TASI advantage, or ratio of number of off-hook callers to channel capacity, is 1.5. In this, as in most of the runs, callers were assumed to generate one packet every T_S = 20 msec during talkspurt. During the runs, all pertinent system variables and statistics are generally updated every T_S sec. Talkspurt and silence durations were generated randomly from exponential distributions with means of 1.23 and 1.34 sec, respectively, for a talker activity fraction of 0.48. Each plotted point represents an average cutout fraction over 1200 sec of simulated real-time activity; this duration was found to be more than sufficient to obtain statistically stable results. Each station updates its prediction on the basis of the current number of local active talkers and issues a new reservation request every T_S sec. The system reservation response time T_R, which for this case is equal to the required predict-ahead

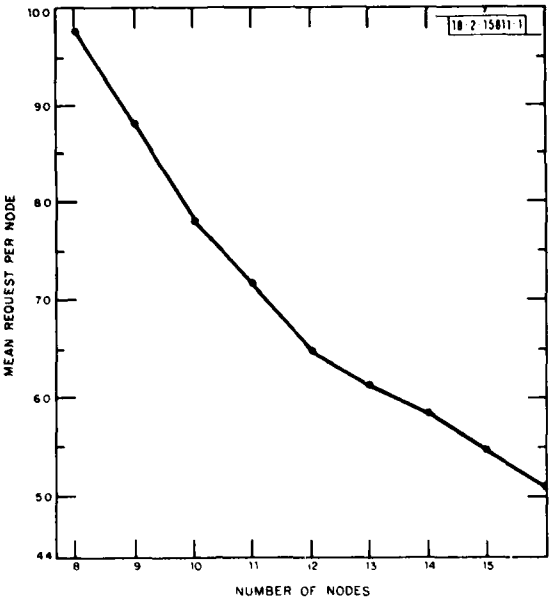


Fig. III-8. Mean request per node yielding smallest fractional speech loss, as a function of number of nodes. Results indicate that margin should be chosen such that total mean reservation request is approximately equal to satellite channel capacity.

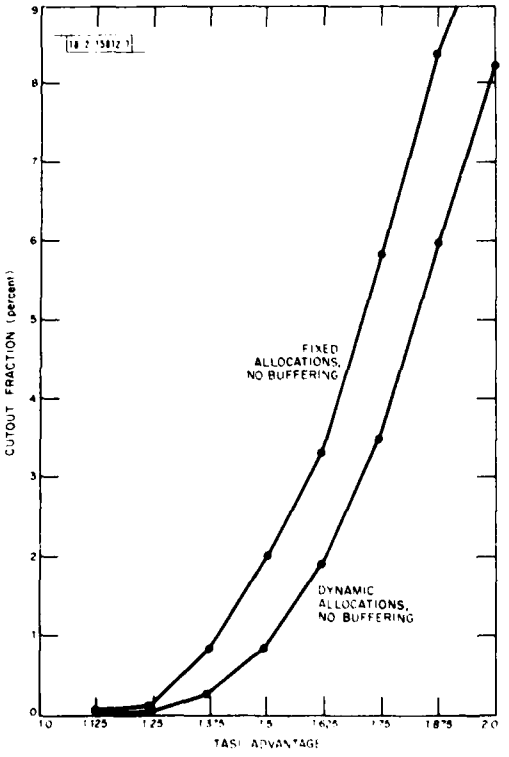


Fig. III-9. Comparison of cutout fraction with fixed and dynamic allocations, as a function of system TASI advantage. Referring to Fig. III-8, note that system TASI advantage varies from 1.0 to 2.0 as number of nodes varies from 8 to 16.

time T_P, is taken as 0.28 sec – just slightly longer than the satellite round-trip time. The manner in which reservation requests are generated from prediction and margin is

$$c_j = \min \{I[u_j(\tau) + M_A], M\} \qquad \text{(III-3)}$$

where c_j is the reservation request from the j^{th} node, and I denotes integer part.

In order to provide more insight into system behavior, we chose to plot fractional loss as a function of mean reservation request per node rather than directly as a function of margin M_A. Clearly, mean request increases with margin to a maximum of 10 slots/node.

As shown in Fig. III-7, cutout fraction can be divided into components arising from two causes: insufficient reservation request at the individual station, and denied reservation requests because satellite capacity was insufficient to accommodate all requests. For low mean request (and margin), almost all the loss is due to the first cause. As mean request per node increases, loss due to insufficient satellite capacity becomes dominant. Overall cutout fraction is minimized when these effects are balanced in an optimal way. For this example, the optimal mean request is about 6.6 slots/node. For the 12-node system, with an overall capacity of 80 slots, optimal performance is achieved when the overall average requested capacity is approximately equal to the total channel capacity. A mean request of 10 slots/node corresponds to the case where prediction is essentially ignored and each station always requests enough capacity to accommodate all 10 talkers. In this case, the round-robin DAMA algorithm will provide equal allocations to all nodes. The 2.0-percent packet loss for the case of equal allocations should be compared with the minimum loss of 0.9 percent. This graphically shows the potential improvement due to prediction-driven dynamic allocation with the correct choice of margin, as compared with equal allocation. An assumption which has been made in this work is that nodes are granted capacity only up to the amount they request, even if not all slots on the satellite channel are requested at a particular time. This excess capacity could be utilized by other traffic (e.g., data) on the channel. If no other traffic is present, then even for the case of optimal margin a small percentage of the available slots on the satellite channel is wasted because no node requests them. It has recently been found that a small degree of further performance improvement can be achieved by distributing unrequested slots among the nodes on a simple round-robin basis.

Performance curves similar to Fig. III-7 were obtained for numbers of nodes varying from 8 to 16, with all other system parameters kept the same. Figure III-8 plots the mean reservation request per node, minimizing cutout fraction as a function of the number of nodes. The observation that the margin should be chosen such that the total mean reservation request is roughly equal to the channel capacity is shown to hold for all cases. Figure III-9 compares percentage loss with variable allocation as determined by prediction with optimal margin against percentage loss with fixed allocations. The improved performance over the range of system TASI advantage is as illustrated.

The required predict-ahead interval (assumed to be 0.28 sec in Figs. III-7 through III-9) is a key parameter of this system. The further into the future one must predict, the less accurate prediction becomes and the less advantage can be obtained. Figure III-10 displays a family of curves, each for a different predict-ahead interval, showing the percentage packet loss at various TASI advantages. The case of equal allocations is included for reference; this can be considered as corresponding to an infinite predict-ahead interval since no improvement from prediction is possible. It should be noted that unless buffering delay is allowed at the nodes, the actual required predict-ahead interval must exceed the satellite round-trip time of 0.27 sec.

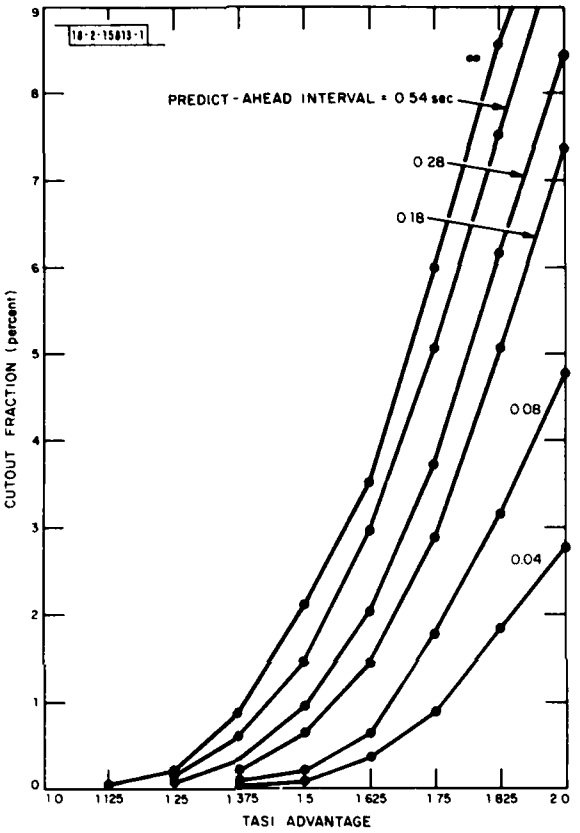


Fig. III-10. Cutout fraction as a function of system TASI advantage for various predict-ahead intervals.

The optimal margins in the results shown so far were determined by carrying out a number of runs with different but fixed values of margin and empirically determining an optimum. An investigation was carried out to determine if margin could be adapted automatically to system conditions (number of nodes, number of talkers, etc.). To zero in on optimal margin, the result was applied that the total mean reservation request should be close to channel capacity. Each node was allowed to observe the total number of reservation requests currently being made, and then to make incremental adjustments in its own margin (and mean request level) to bring the total request level closer to channel capacity. The results were quite encouraging. The nodes quickly approached the optimal margin and stayed at or near this value with small oscillation. Percentage packet loss was very close to the results obtained with optimal fixed margins.

D. TASI PERFORMANCE IMPROVEMENTS WITH COMBINED PREDICTION AND SPEECH STREAM BUFFERING

As shown in the previous section, dynamic allocations based on prediction can improve system performance, decreasing percentage packet loss for a given TASI advantage. Further improvements are possible if buffering is allowed at the nodes. Buffering can avoid packet loss during temporary overload conditions and can effectively reduce required predict-ahead interval. The advantages of buffering for the case of a single multiplexer with fixed-channel capacity are discussed in Ref. 17.

For the multi-node system considered here, the effects of both fixed- and variable-delay buffering have been examined. In fixed-delay buffering, each speech packet is held in a buffer at the transmitting node for a fixed period of time. When this time has elapsed the packet is transmitted if there is sufficient allocation, or discarded otherwise. Fixed delay results in a direct reduction of required predict-ahead interval by the length of the delay. As shown in Fig. III-10, smaller predict-ahead intervals result in more accurate prediction and lower percentage packet loss. As an example, refer to Fig. III-10 and consider a TASI advantage of 1.625. When prediction 0.28 sec into the future is required to match the system reservation response time, there is a 2-percent packet loss. However, a 0.2-sec fixed delay reduces predict-ahead time to 0.08 sec for the same reservation response time, and reduces packet loss to 0.61 percent. Of course, the users must tolerate the increase in speech delay.

For the case of variable delay, the buffer is also limited to a fixed maximum delay but packets stay in the buffer only as long as necessary. Buffer size and delay tend to grow when many talkers are active, and diminish when many talkers are silent. Variable delay also tends to decrease the required predict-ahead interval, but the relationship is not as direct as with fixed delay. However, the need for optimal prediction and margin is not as crucial for the case of variable delay since the buffer tends to smooth out momentary mismatches.

Figure III-11 summarizes simulations that have been run to measure the interrelationship and performance improvement gained from combinations of fixed and variable allocation in conjunction with fixed- and variable-delay buffering. System parameters not given explicitly are as in Figs. III-7 through III-9. For comparison purposes, the results with no buffering delay and fixed allocation are shown. Buffer limits of 100 and 200 msec were considered. For each buffer size, progressively improving performance resulted for the following three cases: (1) fixed channel allocation, variable delay; (2) fixed delay, variable channel allocation; (3) variable delay, variable channel allocation. For the case of a 200-msec variable delay with variable allocation, the packet loss performance is excellent in that the system could be run at a TASI advantage of approximately 1.9 with only 0.5-percent packet loss.

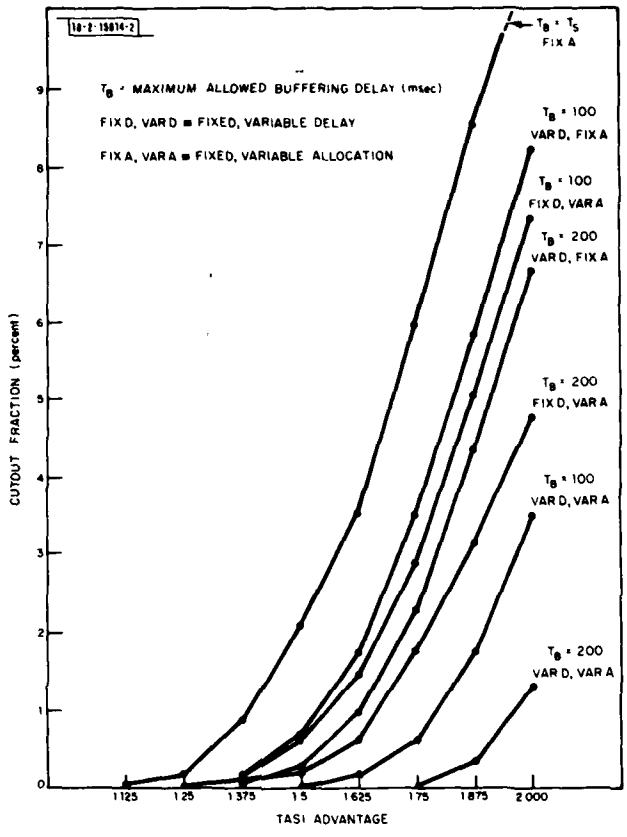


Fig. III-11. Cutout fraction as a function of TASI advantage for various combinations of buffering and allocation strategies.

E. SUMMARY AND CONCLUSIONS

A summary of potential TASI advantage improvements as determined by the simulations is presented in Fig. III-12. Here, system TASI advantage is plotted as a function of the number of off-hook callers at each node for various combinations of prediction and buffering. The results were obtained by requiring the cutout fraction not to exceed 0.5 percent and to determine at what TASI advantage this level of performance would be achieved in each case. For example, the results for 10 speakers/node are obtained from Fig. III-11 by determining at what TASI advantages the various curves cross a cutout fraction threshold of 0.5 percent. As mentioned earlier, this is a conservative threshold for cutout fraction. The satellite capacity is taken as 8M slots, where M is the number of speakers per node. The assumed reservation response time was 280 msec as in most previously presented cases.

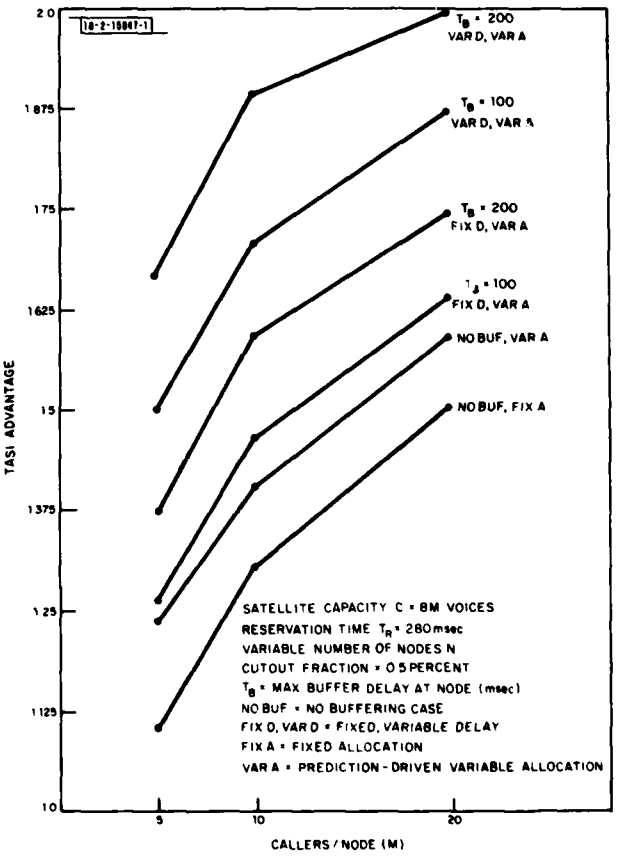


Fig. III-12. TASI advantage as a function of number of off-hook callers per node for various combinations of buffering and allocation strategies.

Referring to the no buffering, fixed allocation case as a baseline, the various levels of performance improvement are apparent. Even for the case of only 5 speakers/node, respectable values of TASI advantage can be achieved. The ordering of performance for various combinations of prediction and buffering follows the previous discussion regarding Fig. III-11.

Prediction and buffering of digital speech streams has been shown to provide potential performance improvement in the statistical multiplexing of speech on a demand-assigned satellite channel, in the case where only a small number of users are multiplexed at each node. One can take advantage of this improvement either by accommodating more callers at a given cutout fraction or by providing a lower cutout fraction to a fixed number of users. Taking maximum advantage of prediction requires a rapidly responsive demand-assignment algorithm capable of changing channel allocations within slightly more than a satellite round-trip time.

The simulations have shown that reservations for channel capacity should be based on prediction plus a correctly selected margin. The "optimal" margin was empirically determined to be the quantity which results in a system-wide reservation level that is approximately equal to the channel capacity. It is possible for the system to adaptively establish such a margin in a dynamic fashion by observing the system-wide reservation rate and making adjustments to the margin currently being used by a node.

Additional performance improvement can be achieved by buffering packets before their transmission. This improves prediction by reducing the required predict-ahead interval. In addition, variable-length buffering provides a smoothing action between temporary overloads and more quiescent time periods. Variable-delay buffering was shown to be more effective than fixed-delay buffering.

IV. SECURE VOICE CONFERENCING

Research on voice conferencing technology at Lincoln Laboratory started in FY 1977 with the construction of an experimental conferencing facility. This facility has been used to carry out a series of experiments designed to evaluate the relative acceptability of different conferencing techniques from a human factors point of view. The goal of the research has been to recommend and demonstrate the best secure conferencing techniques for future defense communication needs. Lincoln Laboratory has been supported in this work by human-factors specialists from Bolt Beranek and Newman, Inc. (BBN), who have carried out the human-factors aspects of the research under contract with Lincoln Laboratory.

Conferencing work in FY 79 has been directed toward three tasks. The first to be described is the continuation of the experimental work started in prior years. The second is effort in support of the Secure Voice and Graphics Conferencing (SVGC) Test and Evaluation Program. The third is a study of advanced secure conferencing systems which deals with system issues as opposed to the human-factors orientation of our other work in the area.

A. CONFERENCING PROTOCOL TESTING AND ANALYSIS

Altogether, six sets of human-factors experiments have been carried out using the Lincoln experimental conferencing facility and the test scenarios and procedures developed by BBN. The first four sets, called "Phases I through IV," were carried out using Laboratory volunteers as subjects. They compared a wide variety of conferencing configurations using both centralized and distributed control techniques. The results of those experiments are included in a comprehensive report[2] on our work in this area that was prepared earlier in FY 79. The last two sets of experiments, Phases V and VI, were carried out subsequent to that report. They constituted repetitions of experiments reported as parts of Phases II and IV in the comprehensive report. The repetitions were undertaken to explore the effects of using different subject populations in the experiments. In particular, we felt that military users of conferencing systems might differ from our group of civilian volunteers in their subjective judgments of system acceptability. To test this hypothesis, we arranged for a group of eight Air Force personnel to participate in a series of experiments. We first ran a comparison among eight versions of the Shared Channel with Distributed Control (SCDC)* conferencing technique and an analog bridge with similar communication delay. This comparison was called Phase V and made use of the Word Match scenario previously used in Phase IV and described in Ref. 2. We then ran a comparison among centrally controlled simplex broadcast and speaker/interrupter systems together with the analog bridge. This comparison, called Phase VI, used the discussion scenario called "Consensus" previously used in Phase II. The following two sections reproduce BBN summaries of the results of the new experiments.

B. SUMMARY OF PHASE V RESULTS

1. Introduction

Results obtained during Phase IV of the Lincoln/BBN teleconferencing study[2] suggested that for SCDC systems simplex broadcast and broadcast interrupter protocols employing short (24-msec) and moderately long (300-msec) preambles were more acceptable to conference

* A brief description of the SCDC technique can be found in Sec. E-4-b below. For a more detailed description, see Sec. 2.7 of Ref. 2.

TABLE IV-1
SUMMARY OF PHASE V EXPERIMENTAL SCHEDULE

Session No.	Run No.	Conditions
1	1	Analog Bridge (AB)
	2	Broadcast Interrupter-short (BIs)
	3	Simplex Broadcast-extra long (SBx)
	4	Broadcast Interrupter-long (BIl)
	5	Analog Bridge (AB)
2	6	Simplex Broadcast-long (SBl)
	7	Analog Bridge (AB)
	8	Speaker Interrupter-slow (SIsl)
	9	Broadcast Interrupter-extra long (BIx)
	10	Simplex Broadcast-long (SBl)
3	11	Simplex Broadcast-short (SBs)
	12	Speaker Interrupter-fast (SIf)
	13	Analog Bridge (AB)
	14	Speaker Interrupter-fast (SIf)
	15	Simplex Broadcast-short (SBs)
4	16	Broadcast Interrupter-extra long (BIx)
	17	Speaker Interrupter-slow (SIsl)
	18	Simplex Broadcast-extra long (SBx)
	19	Broadcast Interrupter-long (BIl)
	20	Broadcast Interrupter-short (BIs)
	21	Analog Bridge (AB)

participants than similar systems employing very long (1067-msec) preambles. Performance scores in the Word Match scenario were found to be very highly correlated ($r_s = 0.934$) with judged acceptability.

A limitation in Phase IV, as in prior phases, was that insufficient time and resources existed for collection of data on different populations of subjects. As a consequence, the generality of results is unknown.

The purpose of Phase V was to extend our analyses of system acceptability and performance to a sample of military personnel. The group of participants differed from the group of civilian volunteers that served in Phase IV in a number of respects, among which were the following:

(a) The military participants were much less experienced, both as subjects serving in laboratory experiments and as judges of alternative conferencing systems.

(b) The military group contained seven males and one female, whereas the civilian group was evenly divided between males and females (4, 4).

(c) More regional dialects were evident in the military group than in the civilian group.

(d) An explicit distribution of ranks, from Lt. Colonel to Airman, existed across the military group. No explicit distribution existed across the Phase IV group, although some implicit hierarchy related to job category (secretary/technician/technical staff) may have been present.

2. Procedure

The training techniques and experimental procedures employed earlier in the program were employed here. Subjects were given 2 h of training during which they were given practice on the Word Match task and on completion of the questionnaire items. In the course of practice sessions, the subjects were given experience with each of the nine teleconferencing systems of interest during this phase.

When training was complete, a series of five experimental sessions was begun, each of which lasted approximately 1 h. The schedule of conditions followed over the series appears in Table IV-1. As in the previous series, subjects were not told what system they were using at any given time. Procedures for acquiring and analyzing questionnaire responses and for accumulating performance data were identical to those used in Phase IV.

3. Results

a. Overall Rating Item

Results obtained with the overall rating item are presented in Fig. IV-1. Note that the rating associated with the Analog Bridge system is higher (i.e., is more favorable) than that associated with all other systems tested. This outcome stands in distinction to that of Phase IV, the results of which are shown in Fig. IV-2, where all but the simplex broadcast and broadcast interrupter systems with extra long preamble were judged to be more favorable than the Analog Bridge.

As earlier, the Wilcoxon test was applied to all possible pairs of ratings [$\binom{9}{2} = 36$] of overall system goodness. The results of the series of tests are presented in Table IV-2 where differences among pairs found to be significant during Phase IV are labeled "4," and those found to

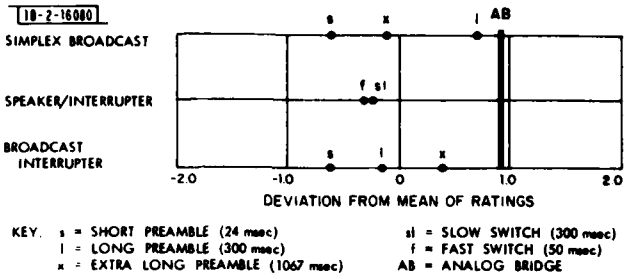


Fig. IV-1. Summary of results obtained with overall rating items during Phase V. Data have been adjusted as described in an earlier report.[2]

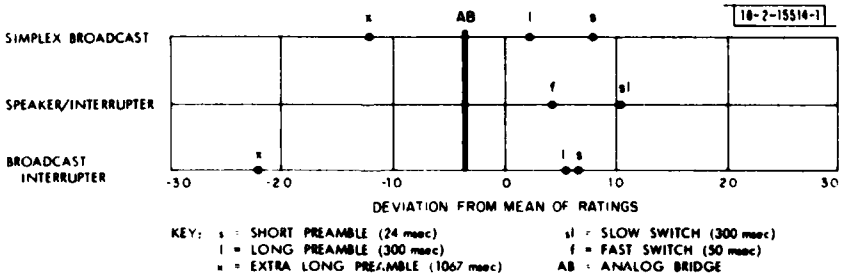


Fig. IV-2. Summary of results obtained with overall rating items during Phase IV. Data have been adjusted as described in an earlier report.[2]

TABLE IV-2

COMPARISON OF WILCOXON TEST OUTCOMES ON OVERALL RATINGS
FOR PHASES IV (4) AND V (5)
(Cell Entry Indicates $p \leq 0.05$, Two-Tailed)

Protocol	AB	SB			BI			SI	
Preamble/Switching Time		x	l	s	x	l	s	sl	f
Analog Bridge (AB)				4	4	5	5 4	5 4	
SB extra long (SBx)			4			4		4	
SB long (SBl)					4				
SB short (SBs)					4†				†
BI extra long (BIx)						4	5 4	4	4†
BI long (BIl)									
BI short (BIs)									*
SI slow (SIs)									*
SI fast (SIf)									

*Not testable in Phase IV due to ties.

† Not testable in Phase V due to ties.

be significant in Phase V are labeled "5." Empty cells in the table denote comparisons found not significantly different in either phase.

Brief examination of Table IV-2 indicates that only 4 of the 36 comparisons proved to be significantly different in Phase V and that 3 of these are associated with the Analog Bridge condition. Bearing in mind that the latter was judged to be superior to all others in Phase V and better than only 2 in Phase IV, it should be clear that the three instances in which significance was demonstrated in both phases (viz. AB vs BIs, AB vs SIf) represent complete reversals. That is to say, whereas both conditions were judged superior to the Analog Bridge in Phase IV, they were judged inferior to it in Phase V. Such a reversal also occurs with respect to the BIx vs BIs condition. The remaining comparison found to be significant, AB vs BIx, represents an addition to the list of comparisons found earlier to be significant.

b. Word Match Performance Scores

A summary of the total times taken to complete Word Match under each of the system conditions is presented in Table IV-3, along with the total times for corresponding conditions in Phase IV. Comparison of the rank order of these times from shortest to longest with the order of points from right to left in Fig. IV-1 suggests a very low correlation of performance with judged overall goodness. A test of the relationship between these parameters indicates a slight, nonsignificant negative correlation ($r_s = -0.03$).

TABLE IV-3
SUMMARY OF PHASES IV AND V WORD-MATCH PERFORMANCE TIMES

System	Mean Performance Time (sec)		
	Phase IV	Phase V	Difference (IV-V)
Analog Bridge (AB)	305	303.8	1.2
SB extra long (SBx)	403	356.5	46.5
SB long (SBl)	285	344.5	−59.5
SB short (SBs)	229	378	−149
BI extra long (BIx)	368	357	11
BI long (BIl)	272	369.5	−97.5
BI short (BIs)	248	291	−43
SI slow (SIs)	256	340	−84
SI fast (SIf)	261	322.5	−61.5

A comparison of the performance times obtained during Phase V with those obtained in Phase IV indicates that, in most cases, the military group took longer to complete Word Match tasks than did the civilian group. Actual differences are presented in column 4 of Table IV-3 where a positive entry indicates a Phase V time shorter than that observed in Phase IV, and a negative entry indicates a time longer.

4. Discussion

The results obtained in Phase V differ from those obtained in Phase IV in the following respects:

(a) Rank order of system conditions with respect to judged quality,

(b) Frequency and identity of systems that differ (statistically) significantly from each other,

(c) Absolute times taken to perform Word Match, and

(d) Degree of correlation between judged quality and Word Match performance time.

Reasons for these differences are impossible to establish with confidence on the basis of the small quantity of data available, but two characteristics of the performance of the military group vis-à-vis the civilian group provide grounds for speculation. One is the difference in conference "pace" alluded to in connection with the data of Table IV-3. The second, revealed by audits of the t[illegible]ngs, is a tendency for sequences of interactions among Phase V participants to proceed in accord with military rank, despite the fact that this affords no advantage in the Word Match task. The results of these two effects is to reduce the amount of competition for the communication channel and, over a set of systems, to minimize differences that might otherwise be associated with preamble duration and switching time in highly competitive contexts. We suspect that, at this reduced level of competition, the military group was unable to experience the effects of different preamble durations and switching times that had formed the basis for the distribution of earlier system ratings.

Although the differences in pace and interaction may help explain why fewer experimental conditions were found to be significantly different from each other in Phase V than in Phase IV, they do not, of course, provide a reason for the pattern of differences that was found. Why, for example, were BIs and BI1 less satisfactory than BIx? One might expect that, at any given level of performance pace and interaction, collisions with a preamble would be more likely if the preamble were long than if it were relatively shorter. Hence, to the extent that overall quality would be expected to depend upon such factors as ease of gaining the "floor" when desired or relative listening effort, one would expect that, in comparison with a system containing no preamble (Analog Bridge), BIx should be the worst condition encountered, followed by BI1 and, finally, BIs.*

Once again, it is extremely important to recognize that the amount of data upon which the outcomes of comparisons in both Phases IV and V are based is much less than that which would be required to make inferences with a high degree of confidence. For the most part, each of the systems in Phase IV has been evaluated on the basis of a single conference. In Phase V, all but AB have been evaluated on the basis of only two conferences each. The statistical tests that have been applied to these small samples make the assumption that each of the participants provides an independent judgment of the overall quality of each of the systems, and the set of eight (Phase IV) or sixteen (Phase V) judgments is generally sufficient to satisfy constraints on degrees of freedom for the test utilized. One must recognize, however, that if a conference conducted over an otherwise very good (or very bad) system happened to go particularly badly

* A similar expectation might arise with respect to differences among the SB systems. In this study, however, no differences were found among SB-AB combinations, and the point remains moot.

(or well), participant's judgments, even though arrived at independently, might tend to reflect the fact and lead to a spurious finding of significance (or nonsignificance).

The smallness of the data base presents at least one further problem in the analysis of differences among systems. As suggested above, a set of eight S's is generally sufficient for application of the nonparametric test used in these studies. However, ties occasionally occur among the judgments of given participants, reducing, in effect, the number of responses available for analysis. When the number of available responses is less than six (i.e., when three or more participants' judgments are tied), the statistical test cannot be run.*

From the point of view of a designer attempting to choose among system alternatives, the difference between a finding of nonsignificance and an inability to apply a statistical test because of ties may not be of practical importance. He may conclude that if a large percentage of participants are essentially neutral with respect to the alternatives, they might as well be viewed as equivalent and the choice made on some other basis. He should recognize, however, that there are at least two other courses of action: (1) He might attempt to identify a statistical test that can be applied to the reduced sample of judgments, or (2) he might acquire more data by replicating the test conditions. We believe that, during these early stages of conferencing system evaluation, increasing the size of the data base is much to be preferred over increasing the resolution of the statistical tests or concluding that differences which cannot be tested for the reason given above are unimportant.

C. SUMMARY OF PHASE VI RESULTS

1. Introduction

Results obtained during Phase II suggested that there is no significant difference in quality between voice-controlled simplex broadcast (VC/SB) and voice-controlled speaker-interrupter (VC/SI) systems with centralized conference control. The experimental comparison of these systems was embedded within a much larger set of comparisons that included, in addition to VC/SB and VC/SI, control-signal-switched (CSS), push-to-talk (PTT), and shared-channel distributed control (SCDC) systems employing procedural and automatic controls.

The research conducted on voice-control systems in Phase II left at least two questions unanswered: (1) How does relative quality (judged "goodness" or "badness") of these systems compare with that of a standard Analog Bridge (AB) system? (2) How might the judgments of quality rendered by a sample of military participants compare with those obtained from civilians?

The purpose of Phase VI was to address these questions. The group of eight participants for this experiment was the same as that for Phase V and differed from the earlier group in the same respects.

2. Procedure

Training techniques and experimental procedures employed earlier in the program were employed here. Subjects were given 1 h of training on the Consensus task and on completion of the questionnaire. During the hour, they had an opportunity to practice the task both on a face-to-face basis and over the AB teleconferencing system that would later be used during the experimental comparisons.

---

* An examination of Table IV-2 will indicate that this occurred two times in Phase IV and three times in Phase V.

When training was complete, a series of three experimental conditions (VC/SB, VC/SI, and AB) was administered. All duties relating to starting and stopping of discussion and to completion of the questionnaire were handled by the experimenter over a dedicated conference line, rather than by an arbitrarily chosen chairperson as in Phase II. This minor departure in procedure was viewed as being consistent with the lower experience level of the Phase VI group and did not represent a material change in the conduct of the task. Procedures for acquiring and analyzing questionnaire responses were identical to those used in Phase II.

3. Results

Results obtained with the overall rating item are presented in Fig. IV-3. Data points labeled "2" are associated with Phase II and are presented for comparison with those "6" from the current phase. The rating for the AB (plotted as a heavy line in the figure) is, of course, unique to Phase VI.

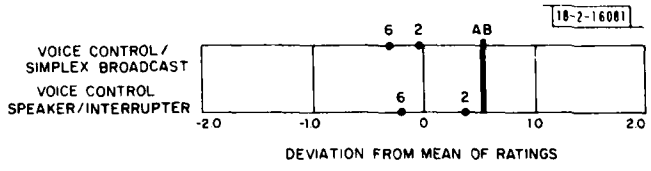


Fig. IV-3. Summary of results obtained with overall rating items during Phases VI (6) and II (2).

As noted in the Introduction (Sec. C-1 above), the difference between the simplex broadcast (SB) and speaker/interrupter (SI) in Phase II was not statistically significant. A similar analysis of the difference between these systems in Phase VI replicates that finding. A difference that is significant ($p < 0.05$) occurs between the SI and AB systems. Although one might anticipate, on the basis of the relative locations of the means plotted in the figure, that the SB would also differ significantly from the AB, that has been found not to be the case. This is due to the fact that the distribution of individual participant ratings for SB overlaps with that associated with AB to a relatively greater degree than does SI.

4. Discussion

The Phase II finding of no significant difference with respect to judged overall quality between VC/SB and SI systems has been replicated in this comparison. Within the context of free conversation provided by the Consensus task, there appears to be little reason to expect that military and civilian personnel will differ in their ratings of the relative quality of these systems.

In terms of mean conferee response, both systems appear to rate lower than the AB system, although only the SI system has been demonstrated to be significantly different from the AB in a statistical sense. This asymmetry may suggest that, at least for a small portion of the population of users, the quality of the SB system is more like that of the AB than is the quality of the SI system.

The VC/SI condition was replicated after the set of three had been accomplished because participants had succeeded in reaching a consensus unusually quickly in the first run.

D. SUPPORT OF SECURE VOICE AND GRAPHICS CONFERENCING (SVGC) TEST AND EVALUATION PROGRAM

The SVGC Test and Evaluation Program is concerned with testing a number of new conferencing systems under conditions approximating field operating environments. The Lincoln/BBN mission in the Program is to provide test scenarios and consultation with respect to procedures for carrying out and evaluating the human-factors testing of the systems. The actual testing will be conducted by the Naval Ocean Systems Center (NOSC) in San Diego, California. In addition, we are providing some help with the preparation of test plans, and we expect to carry out similar tests using our simulation facility as a means of comparing the field-test results with those obtained under laboratory conditions.

Since the equipment to be tested has not yet been installed, our efforts in FY 79 have been limited to help with the test plan and the reworking of some of our test scenarios. In the area of test planning, we attended a plan review meeting where we accepted the task of reworking the section of the draft plan that dealt with the human-factors testing. We submitted the revised section and provided an additional document that enumerated the combinations of system configurations that would have been required to be tested by the original specification. This document pointed out that there were far too many distinct configurations (3984) to permit exhaustive testing, and suggested that further work was needed to prune the test set to a manageable size (100 to 200). We have had further interaction on how to carry out the pruning so as to best serve the needs of the many parties interested in the tests. We are now working on an appendix to the test plan that will specify the tests in more detail and will take the pruning into account.

In the area of test scenario development, we have provided detailed information on the scenarios used in our laboratory experiments. It has been agreed that the "word-go-round" and "word match" scenarios were the types to use in the field-test environment. It was requested that we change the words used in the scenario to those to be found on a Navy word list that is used for other testing purposes. The changes have been carried out and programs are now available to produce the test materials when they are needed.

Some of the SVGC configurations to be tested can produce splits in a conference in which some participants hear a different speaker from that heard by others. Such a split occurs as a result of a collision (two or more participants starting to talk at the same time). To facilitate the testing of configurations subject to this effect, a new version of the "word-go-round" scenario has been generated that is intended to increase the likelihood that collisions of the type that could cause splitting will occur. The scenario is also intended to provide information to allow the experimenters to readily detect that a split has occurred and measure its effect on the conference.

Laboratory experiments involving simulations of SVGC configurations have not yet been carried out. Until the SVGC equipment becomes operational, we cannot determine whether or not our simulations are correct. We have already tested a number of systems that we expect will be very similar to those used in the SVGC tests; but, since our experience has shown that small details can be important in the subjective judgments given to a system by the test subjects, we will not make any laboratory tests of SVGC configurations until we are convinced that our simulations are correct in sufficient detail.

E. ADVANCED SECURE CONFERENCING SYSTEMS STUDY

1. Introduction and Summary

The goal of this study has been to define and analyze future-generation secure conferencing techniques with the objective of identifying viable alternatives that could be implemented in future-generation defense communication systems (DCSs). Security, survivability, efficiency of communications, ease of operation and control, and interoperability with existing and projected systems were properties on which attention was focused. At the outset of the study we expected that it would be possible to address the problems of interoperating with existing and projected systems in some detail, and perhaps to subject some issues to evaluation by simulation. Our experience with simulating conferencing systems had shown that conceptually small details could have a major effect on system acceptability, and we expected that some of the details of interoperation might well lead to unsatisfactory performance and that simulation would be useful in demonstrating the difficulties to be expected. However, it became clear soon after the start of the study that questions of interoperability with projected systems could not be addressed in an adequate fashion because information about the properties of projected systems could not be obtained in enough detail. Consequently, we have had to consider interoperability questions in general terms only and have not made use of simulation in the study.

In this study, we have examined a range of conferencing capabilities, some of which are sufficiently similar to current practice that they could be implemented in the next-generation DCS. Others, which we find to be more desirable, could probably be realized only in the more distant future.

The scope of the study has been limited to the qualitative aspects of conferencing. We have not addressed the question of estimating the number of users who might be expected to require conferencing capabilities in the future, or the quantities of equipment and communication bandwidth that would be needed to serve their requirements.

The plan of this study report is as follows. In Sec. 2 we define three classes of users of conferencing capabilities and show how their requirements place conflicting demands on any system that would attempt to meet all future requirements in a uniform fashion. In Sec. 3 we discuss the interoperability problems that we anticipate will have to be dealt with in future systems. In Sec. 4 we present three alternatives for providing conferencing capabilities in future systems and discuss their relative advantages and disadvantages.

Our conclusions are distributed throughout the report, but may be briefly summarized as follows:

(a) Interoperability problems due to differences in speech encoding and cryptographic equipments are likely to remain in future systems, since they result from fundamental differences in the communication requirements of different user communities. Problems due to a mix of half- and full-duplex communications and terminals are also likely to remain in future systems and to place half-duplex participants at some disadvantages in otherwise full-duplex conferences.

(b) Solutions exist for all the interoperability problems identified, but some are unattractive because they involve additional equipment that adds to system cost and pose problems with respect to siting.

(c) Conferencing using distributed control in a future integrated communication system offers the promise of good performance, low communication costs, and high survivability. These features are not offered together by any other alternative examined in the study.

2. Requirements for Secure Conferencing

Requirements for secure conferencing are expected to exist at many levels and locations in future DCSs. For the purposes of discussion, we define three classes of conferences that we believe encompass expected requirements:

Class I – Local Conferences. These take place within some closed communication environment such as a ship, a base complex, or a tactical radio net. The potential participant group is relatively small, and all members have compatible terminal equipment and communication capabilities. Users are likely to be familiar with the conference capability as a result of regular participation in conferences. In some cases, security requirements can be met by the closed nature of the environment, and crypto equipment may not be required.

Class II – Global Conferences. These are primarily concerned with high-level command and control. The participation is largely fixed, but there is a need to expand it on occasion to bring in arbitrary new participants. Conference durations can be very long, as in the case of crisis management. There is a need for record and graphics conferencing as well as voice. Participants and support personnel can be given special training in the use of any special conferencing equipment, and can be expected to use the equipment on a regular basis to maintain proficiency. Security is very important, and crypto equipment is required for the long-haul communications involved. The terminal equipment available to all participants may not be identical, but the importance of the conference and the fixed nature of most sites justify the use of extra equipment to minimize the effects of any incompatibilities. The use of dedicated communication channels is also justified by the importance of the conference.

Class III – Dial-Up Conferences. These are conferences carried out using the generally available dial-up capability of the switched DCS. The potential participant group is large, compatibility of terminal equipment is not assured, and the ease of setting up and carrying out a conference can be expected to vary depending upon the participant set. Users are likely not to be regular conference participants, and, consequently, the conferencing capability must be easy to learn and to use. Survivability is not usually crucial for this Class, since the conference can be redialed in the event communication is lost. However, in a heavily loaded system, it is likely that users would need to request a high precedence level to avoid losing parts of the conference connection due to preemption.

In principle, the requirements of Classes I and II could be met with a common system-wide conferencing capability that would handle them as special cases of Class III. Historically, it has not been possible to do so both because of technical difficulties and cost factors. The alternative approach of using specialized systems to meet specialized requirements has appeared to be both more tractable and less expensive, and has been used in realizing existing and currently planned future conferencing systems. Looking ahead, we see possibilities for a general conferencing capability that could handle Class II requirements as a special case of Class III. There can be substantial cost benefits if some of the special equipment and dedicated communications currently used to meet Class II requirements can be avoided by using capabilities available to the general user community. Section 4-c below discusses one approach by which such a capability might be achieved without undue increase in costs to the general user. In the case of Class I users, however, it does not appear likely that it will become economical to handle their requirements as special cases of Class III. For example, it is hard to imagine that there would be cost benefits in involving "outside" facilities in a conference that could be handled entirely within a PBX area. Rather, we anticipate that Class I requirements will continue to use specialized conferencing capabilities and that difficulties will continue to be experienced when interoperation is required. Section 3 below discusses the problems that arise when it becomes necessary to interconnect or extend existing and currently planned systems, and it also points out some possibilities for improving the situation in the future.

3. Interoperability Issues

The current (and probably also the future) interoperability problems that affect secure voice conferencing are primarily due to differences in the basic communication techniques that have been developed to meet the special communication needs of the various user communities. Differences in conferencing technique, per se, have been important in the past due to the use of analog bridges that were incompatible with the narrowband speech encoding required by some users, but these differences should become less significant in the future with the use of signal-selection conferencing instead of analog bridges. The differences in basic communications arise from the special needs of users for achievement of effective communication under difficult conditions such as those posed by limited bandwidth, noise, jamming, and size and weight requirements for mobile applications. We anticipate that these differences will remain in the future. Attempts to force uniformity are likely to result in compromised performance for users in difficult environments or significantly increased costs to users in benign environments. Since neither alternative is acceptable, we conclude that future conferencing systems will have to cope with a number of interoperability problems if broad coverage is to be achieved.

There are two types of interoperation relevant to voice conferencing. The first, which we call "extension," refers to the process of adding to an existing conference a participant who does not have the same equipment as the others or who must be connected through communications that have different protocols or cryptographic techniques than those used by the others. The second, called "interconnection," refers to a situation in which two or more independent conferences are to be merged in such a way that the controllers of the individual conferences continue to function (hopefully cooperatively) in the merged conference. The principal problems relative to extension are due to differences in speech encoding and cryptographic techniques, although terminal differences can also be troublesome. Interconnection can have all the problems of extension, as well as problems arising from differences in conferencing protocols and control procedures.

The following subsections discuss the causes of interoperability problems and suggest some possible solutions, first considering those relevant to extension and then those peculiar to interconnection. Although many problems are identified, none are the sort that give cause for alarm. Solutions exist for all, though some of the solutions are unattractive in one sense or another. In particular, many of the solutions involve gateways, translators, or tandeming points. These elements have the common disadvantages that they are costly, must generally be physically secure, and are difficult to distribute so as to have them in the right place at the right time.

a. Speech Encoding Techniques

The most troublesome interoperability problems are due to differences in the speech encoding equipment available to potential conferencing participants. If common equipment is not available, it is necessary to perform some kind of translation to allow any communication to take place. The conventional solution to this problem has been to decode the speech to the equivalent of analog representation and then to re-encode it. This process, called tandeming, introduces some degradation in the signal that will be more or less severe depending upon which particular techniques are being tandemed. The degradation can be rather severe in some cases. Another disadvantage of the tandem solution is that the signal must be decrypted prior to tandeming, thereby requiring that the tandem process takes place at a physically secure location.

There is another approach that can be used in situations such as high-level command-and-control conferences where most of the conference participants can have both wideband and narrowband equipment, but some can have only narrowband equipment. In this approach, the users with both types of equipment transmit using both techniques simultaneously. Other such users listen to the wideband signal when it is present; otherwise, they listen to the narrowband signal. This approach avoids any degradation due to tandeming, but it is expensive in requiring extra encoding equipment and additional communication bandwidth. Another disadvantage is that conference controllers must deal with two distinct communication links for the wideband and narrowband signals. These signals will have different transmission delays if conventional circuits and cryptographic techniques are used. Also, this approach works only when all participants have a common technique (narrowband in this case) and helps only to the extent that it provides better quality speech at times when narrowband users are not talking.

For future use, the embedded coding techniques currently being studied[28,29] hold promise for allowing a mix of wideband and narrowband communications. In this case, all users have compatible encoding/decoding equipment but are connected by communication links with different capacities. The embedded coding equipment is intended to be able to produce speech from all or from one or more subsets of the bits transmitted by a user on a wideband link. If only a narrowband subset were available, listeners would observe some loss of quality or robustness with respect to acoustical noise at the talker's site, but they would still be able to communicate satisfactorily, and there would not be a marked difference in sound quality when changing between wideband to narrowband talkers. Advantages of this approach are that the conference controller has to cope with only a single bit stream, and that no decryption at intermediate points is required. However, some new complexity in crypto equipment at the user's terminal is required to maintain proper synch when only a subset of the transmitted bits arrives at the receiver's terminal. Current research in embedded coding is aimed at developing encoding algorithms with appropriate properties. Some success has been achieved in this area, but questions remain about the cost of the processing required. On the basis of current knowledge, we cannot assume that this technique will be of low enough cost to postulate its widespread use in the future DCS.

If the promise of embedded coding should fail to be realized, it may be necessary to use tandeming or some other form of translation between encoding techniques in future systems. Any translation requires that the signal be decrypted at the translation point, but there are possibilities for translation without producing the analog speech equivalent required for tandeming. For example, within a family of similar encoding techniques such as transform coders at various bit rates, it is possible to compute the representation at one bit rate from that at another. There should be no loss of quality in going from a low rate to a higher rate, and minimal loss in the other direction.

In the case of true tandeming, there is hope for the future because some of the worst tandeming problems occur when going between different techniques operating at narrowband (say 2400 bps) and medium band (9600 bps or so), and there should be little need for such in the future. The primary need for medium-band communication grows out of current requirements for sending digital speech over analog lines. With the future availability of true digital communications, the need for encoders in this range should disappear. If the general user had PCM or some other waveform encoder offering almost PCM quality, tandeming with narrowband users would not introduce significant degradation and could be considered an acceptable solution to this interoperability problem.

b. Cryptographic Techniques

The second most troublesome interoperability problems are due to differences in crypto techniques used in different communication situations. If a would-be participant has different crypto gear than that used by others in a conference, he must be connected through a gateway that can translate between his crypto algorithm and that used by the conference. The need for such gateways would disappear if all users could be provided with identical equipment; but that situation is not likely to occur because, even if the same basic algorithm is used to encrypt the signal, there are special requirements for synchronization, etc. in some communication situations that force differences in the techniques used in those cases. For example, in a conference involving broadcast communications the participants must have a common means of decrypting each other's transmission. A simple encryption technique with all participants sharing a common key variable is not acceptable, because if two or more participants should transmit at the same time (a probable event in a conference situation) the security of the transmissions would be compromised. (Potentially, the keystream could be derived by adding together two encrypted streams using a common key.) To avoid this problem, it is necessary to use a different technique in which additional information is sent as a preamble to the data to be transmitted. The preamble has to have the same effect as would be achieved if each participant had his own key variable. Unfortunately, this preamble adds delay to the transmission. The length of the delay depends on the extent to which the preamble contents must be protected from transmission errors by the use of coding techniques. Protection is required, since an error in the preamble contents would prevent the following transmission from being decrypted correctly. In a worst-case situation of narrowband communication in the presence of high noise or jamming, the delay could be quite long, and users communicating under more benign conditions would consider such a delay to be unacceptable. Even within the narrow scope of broadcast conferences, it is unlikely that worst-case preambles would be acceptable for general use.

If packet techniques are used in future speech-transmission systems, the problem of delay caused by crypto preambles can be largely avoided. Secure packet transmission requires the

use of preambles, but because the actual transmission rate is much higher than the net speech bit rate, the real-time delay due to the preamble is substantially reduced. If a preamble is used on each packet, which is a desirable procedure when there is a significant probability of losing packets in transmission, it is not necessary to protect the preamble against noise because only one packet will be lost as a result of an error in a preamble. Of course, there is a price to pay for this advantage. There is a loss of transmission efficiency due to the preambles, and there is some delay due to the packetization procedure itself. Since we cannot assume that packet techniques will be used in all future communication systems, we conclude that there will be a continuing need for translation between crypto techniques to meet the diverse requirements of the user communities.

c. Terminal Capabilities

Current conferencing systems aimed at Classes II and III requirements assume full-duplex communications so that a participant can hear the conference when attempting to speak to it. The ability to hear allows him to quickly detect a failure to get the conference "floor" when he tries to do so. Experiments with audible signals to indicate detection of channel collisions have demonstrated the value of such signals in speeding the flow of the conference and increasing user confidence. The half-duplex terminals used in some communication situations deny these benefits to participants who would have to use such terminals in a conference in which other participants had full-duplex capability. Currently, half-duplex terminals are used in situations where the communication medium is inherently half-duplex, for example in a radio net, as well as in other narrowband environments where the motivation for half-duplex is largely one of cost saving since the same processor can be used alternately for the analysis and synthesis tasks in the narrowband encoding algorithm. We anticipate that processing costs will decrease sufficiently in the future that the latter motivation for half-duplex terminals should disappear. However, it is likely that half-duplex communications will continue to be used, and the problem of extending full-duplex conferences to include half-duplex participants will remain.

The half-duplex participant in a full-duplex conference is at a greater disadvantage than he would be if all users had half-duplex terminals, since in the latter case protocols would be used that would tend to minimize the difficulties associated with half-duplex operation. For example, half-duplex conferencing is the normal mode of operation in a radio net, and formal procedures for handing over the right to talk prevent collisions in most cases. If a collision should occur, no one will hear good speech, and all parties will be in the same state with respect to the conference scenario. However, consider the case of a user on a radio net who is connected into a full-duplex conference outside the radio net. If he should attempt to speak to the conference and fail due to some other participant starting to speak a little bit sooner or having higher priority, he would be unaware of his failure, and the other participants would have heard the other speaker and would not have been aware of the radio net user's attempt to speak. He would have to deduce from the ensuing conversation that he had not been heard and that he should try again to make his point. Fortunately, this situation is not always damaging to the conference as a whole, but it makes participation difficult for the disadvantaged half-duplex participant. In our experience with test scenarios, we have seen a few instances in which this kind of difficulty caused significant problems for the conference as a whole. Protocols at both the human and conference controller levels can help to minimize the problem by giving some compensating advantage to the disadvantaged user. At the human level, the full-duplex users can force themselves to delay

their attempts to get the floor to allow the half-duplex users to get there first. Also, the conference chairman can explicitly ask for responses from individual participants, thereby reducing the chance for contention to become a problem. Our experiments involving Air Force subjects suggest that ordinary military deference to rank will help in this case by reducing the probability of contention for the conference floor. Future conference controllers can help half-duplex users by giving them a priority advantage, allowing them to override full-duplex users. However, high-ranking full-duplex participants may find occasional interruption by lower-ranking half-duplex participants to be unacceptable, and we do not assume that this technique will be used.

Another possible aid for the half-duplex user in future systems is the use of out-of-band signaling to warn him when he attempts to speak and fails to get the floor. Packet techniques allow such signaling with a minimum requirement for additional communication capacity. It is likely that other means could be provided in the absence of packet capabilities.

Another area of possible problems due to terminal differences relates to the ability of a user to set up and control access to a conference from any terminal. To be able to do so requires all terminals to have the same ability to handle signaling with conference controllers. At setup time, there is not likely to be much of a problem, since ordinary dialing capability should be sufficient; but once a conference is under way, signaling for control purposes requires additional communication capacity and terminal flexibility. We expect that it will be possible to provide the required capabilities in future systems and terminals without significant cost penalties.

If a conference is to involve record or graphics communication as well as voice, it is necessary for the participants to have appropriate compatible equipment. Since this type of conferencing represents a new capability, it should be possible to provide compatible equipment for the relatively small number of potential users who will require it. In that case, interoperability problems would not arise. However, there is always the chance that it would be necessary to use some mix of existing equipment that would pose problems and require some intervening translation in order to communicate. It does not appear likely that any significant technical problems would be encountered in carrying out such translations, but, as with any translation equipment, there is always the problem of having the right equipment at the right place at the right time.

d. System Interconnection

Interoperability problems can occur when it is necessary to interconnect conferencing systems. Such interconnection is most likely to be desired between special systems built to suit the needs of Class II users. For example, if it was decided to interconnect command-and-control conferences using broadcast satellite systems in the Atlantic and Pacific Oceans, interoperability problems could occur if these systems used different protocols and control procedures or even if they were identical but lacked provision for interconnection. If protocols are identical or similar enough, such systems can be interconnected by taking the output (selected speaker) of one system and introducing it into the other system as an additional participant, and vice versa. We have experimented informally with this simple kind of interconnection, and have concluded that it could be expected to operate satisfactorily most of the time. However, if one participant in each system starts speaking at the same time, the conference will split into separate conferences with the participants in each system hearing that system's speaker without being aware of the other system's speaker. The split will continue until silent intervals coincide in both systems. A slow-moving, polite conference will not experience much splitting. A heated conference

with aggressive contention for the floor can be expected to have many problems from this source. The damage caused by splitting will depend on the state of the conference scenario at the time a split occurs and cannot be realistically evaluated in a laboratory situation. In our opinion, operational experience will be required to assess the acceptability of simple interconnections subject to splitting.

Splitting problems can be overcome quite simply by means of a global control policy. The simplest of such policies is to give each participant a global precedence value and, if a split is detected, to abort the lower precedence speaker. The difficulty of implementing such a policy depends upon the design of the individual conferencing systems. It should not be difficult to design future systems so that they could support such a global policy gracefully.

If more complex interconnections were attempted, such as those symbolized in Figs. IV-4 and IV-5, other difficulties could be expected. The circles in the figures represent areas of satellite coverage, as well as the extent of individual conference control regimes. The boxes labeled G represent gateways that are assumed to do whatever they can to effect the interconnection of the systems. In the case of Fig. IV-4, the cascade of delays between A and D would cause difficulty even if a global precedence control policy were in effect, because the delay would increase both the probability of splitting and the time required to recover from it. If an attempt were made to overcome the problems caused by delay in Fig. IV-4 by increasing the connectivity of the systems to produce a configuration like Fig. IV-5, other problems would be introduced that would require more complex algorithms to control. For example, without special routing control, speech from a talker in B would be fed simultaneously to A and D, each of which

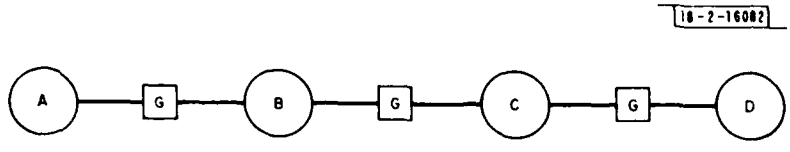


Fig. IV-4. Interconnection of controllers with problems due to cascaded delays.

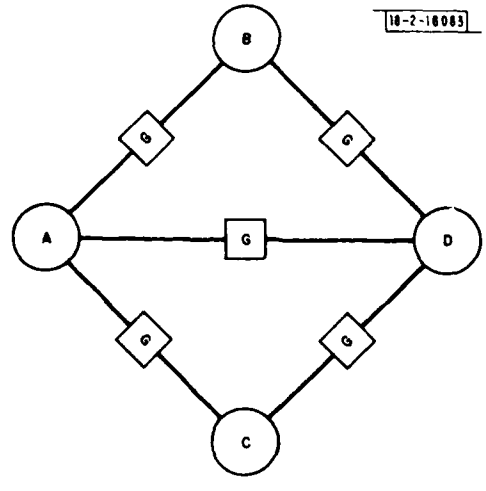


Fig. IV-5. Interconnection of controllers with potential routing problems.

would attempt to pass it along to C, where the two simultaneous inputs would collide and cancel each other. This kind of difficulty could be corrected by introducing a global routing policy that determined whether or not each gateway should pass a signal on to the next system. The decision would be based on the origin of the transmission. There is no conceptual difficulty in designing an appropriate routing algorithm for such a network, but its implementation may pose problems because of the inability of the individual systems to provide information about connectivity and the origin of the transmissions. In the example (Fig. IV-5) it is not only necessary to avoid the collision caused by A and D simultaneously transmitting to C, it is also necessary to make sure that the transmission stops at D and is not allowed to loop indefinitely.

If future conferencing systems are designed with interconnection in mind, it should not be difficult to avoid the problems discussed here almost entirely. If a global conferencing capability such as that discussed below in Sec. 4-c were to be implemented, there would be no need for system interconnection. In either case, we expect that interoperability problems due to system interconnection can be reduced to insignificance in the future.

4. System Alternatives

The intent of this section is to discuss three of the many possible conferencing system architectures that could meet some or all of the Class II and III user requirements. We believe that these three are representative of systems that could be built in the not-too-distant future. They are presented in the order in which they depart from conventional conferencing capabilities.

All three systems could support any of the conferencing protocols [simplex broadcast (SB), speaker-interrupter (SI), etc.] that we have examined in our human-factors experiments. Since those experiments showed that SB was the preferred protocol, we have assumed that it would be used in future systems and do not discuss it further. If other considerations should lead to the choice of some other protocol, we expect that the comparisons among the three alternatives made in this section would still apply because the system considerations that are the focus of the comparisons are essentially independent of the choice of protocol.

The alternatives chosen represent particular combinations of the design choices between centralized and distributed control on the one hand, and communication technique on the other. We have combined central control with conventional point-to-point circuit communications, and distributed control with advanced techniques that offer broadcast capabilities. Other combinations are possible, but in our opinion they offer no advantages over those chosen. Distributed control is prohibitively expensive with point-to-point communications, and central control adds delay and reduces survivability when broadcast communications are available.

a. Alternative I: Central Control with Point-to-Point Communications

Historically, voice conferencing has been handled by making point-to-point connections between each participant and a conference bridge located at some convenient place. The natural extension of this technique for future use would be to utilize digital circuits and to replace the bridge with a conference controller that would select a speaker rather than sum and signals from the participants. Selection allows conferencing to occur satisfactorily with the narrowband encoding required by some users. This combination, called Alternative I, is schematically represented in Fig. IV-6. The circuits that connect participants to the controller could be either terrestrial or satellite links, and could carry low-data-rate out-of-band signals for conference control and supervision as well as encoded speech signals. With additional multiplexing, these

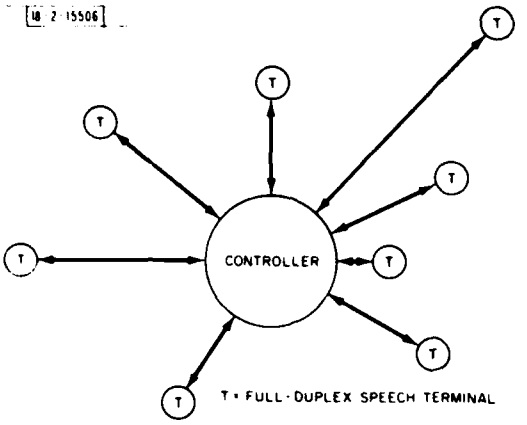


Fig. IV-6. Centrally controlled conferencing configuration.

circuits could also carry low-data-rate information for record or graphics augmentation of a voice conference, but effective augmentation is likely to require more bandwidth than could be achieved in this way. Any multiplexed use of these lines would require different terminal equipment and line formats than those now in use for secure speech transmission.

Alternative I has the following advantages as a candidate for future system use:

(1) Extension to bring in participants with nonidentical equipment is relatively easy. Any necessary tandeming or crypto translation can be carried out at the controller. There is no problem in deciding where to locate such conversion equipment, as there is if control is distributed.

(2) Collision handling (dealing with the situation where two or more participants start talking at the same time) can be optimized. All necessary information is available at the controller, and the confusion that causes problems for distributed controllers when delay is present is not a problem. Though there may be some difficulty caused by differences in the communication delays between participants and the controller, these differences can be equalized or compensated if desired.

(3) No additional delay is introduced by crypto preambles since they are not needed for the point-to-point circuits. Any extra communication required for crypto operation can be handled at the time the conference connection is set up.

(4) A minimum of special equipment is required at a subscriber's terminal, and no special equipment is needed in communication switches since all control functions are performed at the central controller.

On the other hand, Alternative I has two serious disadvantages:

(1) Survivability is poor due to the centralization of control and translation functions. Additionally, communications in the vicinity of a controller are vulnerable because of their increased density relative to the system as a whole.

(2) Communication costs are high due to the inefficient use of channel capacity (many channels carry the same information in a large centrally controlled conference) as well as the need to provide large-capacity nodes to support conference controllers.

The survivability problem can be solved to some degree by providing backup controllers that are kept up to date with respect to conference participation. Automatic switchover to a backup controller could be carried out in the event that the primary controller failed or was destroyed. With future digital communication systems it should be possible to effect such a switchover in a few seconds, and the resulting hiatus would not be very disturbing to the conference. Unfortunately, this solution further increases the cost of this alternative.

b. Alternative II: Shared Channel with Distributed Control (SCDC)

It is possible to share a broadcast satellite channel among a number of earth terminals in a fashion that minimizes the channel capacity needed to support a conference. By distributing the channel control among a set of cooperating controllers following the same algorithm, survivability can be increased relative to a centrally controlled approach. However, the satellite itself remains as a vulnerable common point. The technique is applicable to any broadcast medium such as radio, but its attractiveness for future conferencing systems lies in its use with satellites where low-cost long-haul communications are required to meet the needs of Class II users. This type of system is currently being explored in the SVGC Test and Evaluation Program, and we have simulated the control algorithms and found them to be acceptable in human-factors tests.

The control algorithm involves sensing the presence or absence of signal in the shared channel, and starting to transmit a participant's speech only when the channel is observed to be free. Because of the delay in the satellite transmission, there is a period of time between the start of transmission and the instant at which the other controllers detect that the channel has become busy. During this period other controllers may also start transmitting, thereby causing a collision in channel usage that will prevent useful communication until corrective action has been taken by the controllers. A variety of recovery algorithms has been explored, the best of which requires at least one satellite round-trip time plus one crypto preamble time to regain use of the channel. The preamble time depends upon the speech coding rate and the noise characteristics of the channel, and can be quite long for narrowband communication in the presence of noise or jamming.

Figure IV-7 shows a schematic representation for Alternative II. As indicated in the figure, it is expected that a single controller would serve more than one conference participant because the cost of an earth station and associated controller is too high to provide one for each possible participant. In this configuration, the controllers serve as central controllers for their local participants. If they are provided with equipment for tandeming and crypto translation, the ease of extending the system to bring in participants with nonidentical equipment would be comparable to that for Alternative I. However, it is more expensive to provide enough conversion equipment when it is distributed than when it is centralized, because more must be provided to have the same probability of being able to satisfy needs with locally available equipment.

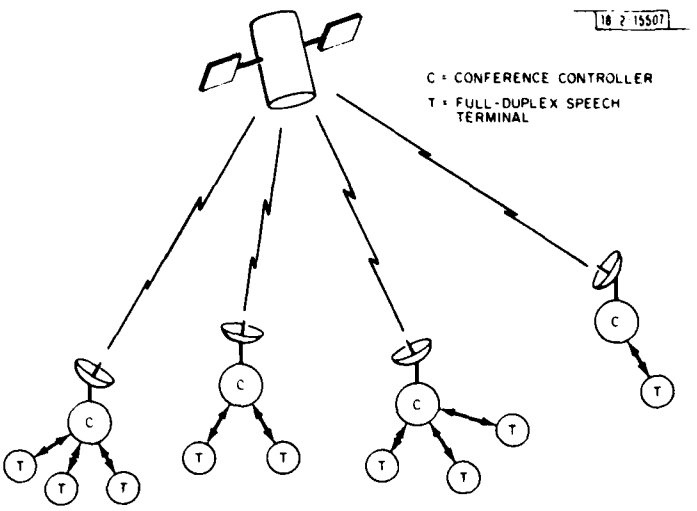


Fig. IV-7. Distributed conference controllers sharing a satellite channel.

The advantages of Alternative II are:

(1) Distributed control provides high survivability. There is no critical site whose loss would completely disable a conference. This feature is especially important for Class II users.

(2) Long-haul communication costs are minimized since only one satellite channel is required to support a conference with many participants. There also may be savings in terrestrial costs if the controllers are favorably located with respect to the "local" users. Although communication costs for Alternative II are expected to be much lower than for Alternative I, controller cost will be higher because more controllers are required and they are individually more complex to deal with the shared channel. We expect that overall cost comparisons would remain favorable.

(3) The probability of successfully setting up and maintaining a conference in a heavily loaded system is increased relative to Alternative I because less capacity is required to support it with Alternative II.

The disadvantages are:

(1) Subjective tests have shown that conferencing performance is less good with Alternative II than with Alternative I. The performance problems are caused by the less graceful handling of collisions. More speech is lost in a collision due to the delay between the distributed controllers as well as the delay introduced by the need to use crypto preambles.

(2) In order to extend the long-haul cost benefits beyond the coverage of a single satellite, it is necessary to interconnect systems. Interconnection significantly complicates the distributed control process and magnifies the undesirable effects of collisions. Section 3-d above contains a discussion of the problems to be expected when interconnecting systems of this type.

(3) Out-of-band signaling for conference control and supervision is difficult because only the current speaker's controller has transmit access to the channel. In the case of Alternative I, the circuits between the controller and the participants are always available for signaling.

(4) Conference setup and supervision are more complex with distributed than with central control. This complexity is of no concern for many Class II users whose conference participation is fixed, but it could be troublesome in using Alternative II for Class III applications.

(5) Augmentation to include record or graphics information is straightforward but requires the use of a second shared channel, since it is unlikely that the current speaker and the current sender of graphics information would be at the same site.

c. Alternative III: Integrated Communications with Distributed Control

This alternative is presented to show the potential value of new communication capabilities in supporting conferencing. An integrated communication system is assumed to handle both voice and data, and to provide an opportunity for terminals and controllers to exchange control information independent of the flow of voice signals. This control information can allow graceful recovery from collisions, provide for priority interrupts during a conference, and facilitate conference supervision and control. Record and graphics augmentation is also straightforward in an integrated system. Some integrated systems, such as those using packet technology for voice, could offer additional advantages by providing multi-address delivery of conference packets and by transmitting at a bit rate substantially higher than the speech encoding rate, thereby reducing the delay associated with crypto preambles. The higher transmission rate could also allow speech from more than one talker at a time to be received during a collision event, further improving collision handling by allowing the receiving controller to choose which talker's speech to accept.

The distributed controller for Alternative III would allow transmission whenever speech was not being received and some one of its participants was above threshold at his speech activity detector. Initiating transmission, a controller would send a control packet indicating that it was starting to transmit. It would then listen for control packets from other controllers while continuing to transmit its talker's speech. If it received a control packet from some other controller transmitting for a higher-precedence speaker, it would abort its transmission, signal its participant, and prepare to play out the speech from the higher-precedence speaker. During the collision event, the load offered to the communication system would be higher than the nominal single speaker load associated with the conference on the average. Depending upon the system design, current overall load, etc., the system might deliver all, some, or none of the momentary excess traffic load. If it delivered all the speech, the receiving controllers could play out the speech of

the highest-precedence participant and could therefore minimize any glitch in the conference due to the collision. In this case, the controllers would not need to make use of the control packet since the speech itself carries the same information. In the worst case, none of the speech would be received because of channel collisions en route, and the controllers would depend upon the control packet's arrival to bring the collision event to an end. If the speech were not packetized, there would be some further delay in this case to transmit a new crypto preamble. It is important for collision control that the control packets traveling in parallel with the speech not be subject to contention that could prevent their delivery.

In an integrated communication system offering multi-address delivery, much of the mechanism for setting up and maintaining conference connectivity would be provided by the system itself. The responsibility for routing the conference connection would be distributed through the system in such a way that broadcast capabilities would be utilized where possible to minimize the cost of serving the conference. Where no inherent broadcast capability existed – as for terrestrial links, for example – the system would replicate signals as required. Rerouting around failed switching nodes or links would be automatic. This distributed responsibility could offer both improved survivability and higher probability of being able to set up and maintain a conference under heavy load than could be obtained with either Alternative I or II.

Since transmission occurs only when speech is present, the communication cost of keeping conference lines open for long periods would be minimal for this alternative. Such usage is desirable in crisis management situations.

Alternative III offers a means of meeting the requirements of both Class II and III users with a common system. The Class II users merely require higher-precedence connections. On the other hand, Alternative III is at a disadvantage in requiring somewhat more complex translation equipment if interoperation with existing circuit-oriented equipment is required.

In summary, Alternative III offers the following advantages:

(1) Survivability is excellent. Both conference control and communications are distributed.

(2) Communication costs are low. There need be no parallel paths carrying the same information, and transmission capacity is used only when speech or other signals are present.

(3) Collision handling approaches the optimal performance of Alternative I.

(4) Out-of-band signaling for control and supervision is very easy.

(5) Augmentation with record and graphics is straightforward.

(6) If packet techniques are used for speech, delay due to crypto preambles can be substantially reduced.

(7) The relatively low communication requirements of this Alternative increase the probability that a conference can be set up and maintained in a heavily loaded system.

(8) The task of a distributed controller is relatively simple, and we estimate that it could be incorporated into an individual user's terminal at an acceptable cost.

The principal disadvantage of Alternative III is that it represents a significant departure from current capabilities. This departure means that interoperation with more conventional systems would pose some additional problems because Alternative III makes use of properties of the communication systems that are not present in current systems. The evolutionary process of getting from here to there would be more difficult for this alternative than for the others. Many of the claimed advantages have yet to be demonstrated, and costs have yet to be assessed in sufficient detail to satisfy all concerned. The whole question of how (or even whether) to provide integrated communications is a currently controversial topic. In our opinion there are substantial advantages to be gained from such systems, particularly for conferencing applications, and we favor continued exploratory work in this area. We expect that opportunities to demonstrate conferencing capabilities similar to those described here for Alternative III will occur in the next few years as part of the EISN Experiment. That experiment should also afford opportunities to explore some of the problems to be expected in interoperating with more conventional techniques.

REFERENCES

1. Network Speech Processing Program Annual Report, Lincoln Laboratory, M.I.T. (30 September 1978), DDC AD-A068318/5.
2. Voice Conferencing Technology Program Final Report, Lincoln Laboratory, M.I.T. (31 March 1979), DDC AD-A074498.
3. "Design Concepts for the Next Generation CONUS AUTOVON," Defense Communications Engineering Center Technical Report 18-78 (December 1978).
4. "The DCS III Architecture and Advanced Systems Concepts," Defense Communications Engineering Center Technical Report 21-78 (December 1978).
5. C. J. Weinstein, M. L. Malpass, and M. J. Fischer, "Data Traffic Performance of an Integrated Circuit- and Packet-Switched Multiplex Structure," Technical Note 1978-41, Lincoln Laboratory, M.I.T. (26 October 1978), DDC AD-A065449/1.
6. R. Berger, "The Queueing Behavior of a Buffered TASI Multiplexer," to be submitted to the IEEE Transactions on Communications.
7. P. J. Brady, "A Technique for Investigating On-Off Patterns of Speech," Bell Syst. Tech. J. 44, 1-22 (1965).
8. G. J. Coviello, "Comparative Discussion of Circuit- vs. Packet-Switched Voice," IEEE Trans. Commun. COM-27, 1153-1160 (1979).
9. G. J. Coviello and P. A. Vena, "Integration of Circuit/Packet Switching by a SENET (Slotted Envelope NETwork) Concept," National Telecommunications Conference, New Orleans, December 1975, pp. 42-12 to 42-17.
10. S. T. Walker, "ARPA Network Security Project," 1977 EASCON Proceedings, pp. 14.5A to 14.5C.
11. R. D. Rosner, "Optimization of the Number of Ground Stations in a Domestic Satellite System," 1975 EASCON Proceedings, pp. 64A to 64J.
12. J. D. Barnla and F. R. Zitzmann, "Digital Communications Satellite System of SBS," 1977 EASCON Proceedings, pp. 7-2A to 7-2I.
13. I. M. Jacobs, R. Binder, and E. V. Hoversten, "General Purpose Packet Satellite Networks," Proc. IEEE 66, 1448-1467 (1978).
14. K. Bullington and J. M. Fraser, "Engineering Aspects of TASI," Bell Syst. Tech. J. 38, 353-364 (1959).
15. S. J. Campanella, "Digital Speech Interpolation," COMSAT Technical Review 6, 127-159 (1976).
16. P. J. Brady, "Effects of Transmission Delay on Conversational Behavior on Echo-Free Telephone Circuits," Bell Syst. Tech. J. 50, 115-133 (1971).
17. C. J. Weinstein and E. M. Hofstetter, "The Tradeoff Between Delay and TASI Advantage in a Packetized Speech Multiplexer," IEEE Trans. Commun. COM-27, 1716-1720 (1979).
18. W. Kantrowitz, C. J. Weinstein, and E. M. Hofstetter, "Prediction and Buffering of Digital Speech Streams for Improved TASI Performance on a Demand-Assigned Satellite Channel," Technical Note 1979-75, Lincoln Laboratory, M.I.T. (15 November 1979), DDC AD-A081593.
19. R. Binder, "A Dynamic Packet Switched System for Satellite Broadcast Channels," Proc. ICC 1975, San Francisco, California, June 1975.
20. J. W. Forgie, "Speech Transmission in Packet-Switched Store-and Forward Networks," National Computer Conference Proceedings, Anaheim, California, 19-23 May 1975.
21. I. Gitman and H. Frank, "Economic Analysis of Integrated Voice and Data Networks: A Case Study," Proc. IEEE 66, 1549-1570 (1978).

22. J. W. Forgie and A. G. Nemeth, "An Efficient Packetized Voice/Data Network Using Statistical Flow Control," Proc. ICC 1977, Chicago, Illinois, 12-15 June 1977.

23. J. M. Elder and J. F. O'Neill, "A Speech Interpolation System for Private Networks," NTC Conference Record, Birmingham, Alabama, December 1978, pp. 14.6.1 to 14.6.5.

24. J. W. Forgie, "Network Speech System Implications of Packetized Speech," Annual Report to Defense Communications Agency, Lincoln Laboratory, M.I.T. (30 September 1976), DDC AD-A045455/3.

25. C. J. Weinstein, "Fractional Speech Loss and Talker Activity Model for TASI and for Packet-Switched Speech," IEEE Trans. Commun. COM-26, 1253-1257 (1978), DDC AD-A065035/8.

26. R. Syski, Introduction to Congestion Theory in Telephone Systems (Oliver and Boyd, London, 1960), pp. 243-245.

27. L. Kleinrock, Queueing Systems, Vol. I: Theory (Wiley, New York, 1975).

28. G. Kang, L. Fransen, and E. Kline, "Multirate Processor (MRP)," Naval Research Laboratory Report (September 1978).

29. B. Gold, "Multiple Rate Channel Vocoding," EASCON '78 Record, pp. 693-697.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ESD-TR-79-313	2. GOVT ACCESSION NO.: AD-A086768
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Network Speech Systems Technology Program	5. TYPE OF REPORT & PERIOD COVERED: Annual Report, 1 Oct 78 – 30 Sep 79
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Clifford J. Weinstein	8. CONTRACT OR GRANT NUMBER(s): F19628-80-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Program Element No. 33126K
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Communications Agency 8th Street & So. Courthouse Road Arlington, VA 22204	12. REPORT DATE: 30 Sep 1979 13. NUMBER OF PAGES: 60
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Electronic Systems Division Hanscom AFB Bedford, MA 01731	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

None

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

network speech processing
secure voice conferencing
speech algorithms
integrated networks
packetized speech
human factors
speech encoding
teleconferencing

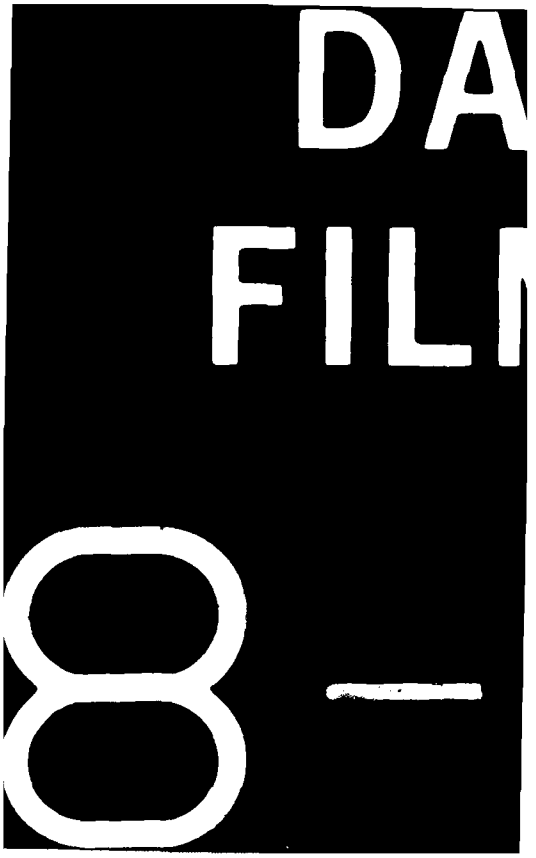
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report documents work performed ~~during FY~~ 1979 on the DCA-sponsored Network Speech Systems Technology Program. The areas of work ~~reported here~~ are: (1) a switching and multiplexing study dealing with the analysis of buffered voice and data multiplexers and with a proposed technique for exploitation of speech activity detection to increase channel efficiency in a multi-link hybrid network; (2) a Demand-Assignment Multiple Access (DAMA) study focusing on prediction and buffering of digital speech streams for improved speech multiplexing performance on a broadcast satellite; (3) efforts in secure voice conferencing including protocol test and evaluation, support of the Secure Voice and Graphics Conferencing (SVGC) Test and Evaluation Program, and study of potential future-generation conferencing system strategies. Progress during FY 79 in experiment definition and planning for the Experimental Integrated Switched Network (EISN) test bed being developed under joint DCA/DARPA sponsorship is reported separately in a Project Report.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


DA
FIL
8 —