

CHI

CULLER-HARRISON INC.

150-A Aero Camino, Goleta, California 93017

12

Telephone (805) 968-1064

FINAL TECHNICAL REPORT

May 1, 1977 to December 31, 1977

ARPA Order: 3402

Contract: MDA903-77-C-0220

Contractor: Culler/Harrison, Inc.
150-A Aero Camino
Goleta, CA 93017

Effective Date: May 1, 1977

Expiration Date: December 31, 1977

PROJECT SCIENTISTS: Drs. Judith Bruckner and Michael McCammon

TITLE: SPEECH COMPRESSION PROTOTYPE DEVELOPMENT

Research Sponsored By

Defense Advanced Research Projects Agency
under ARPA Order No. 3402, Contract No.
MDA903-77-C-0220, Monitored by DCASMA
Oxnard.

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

DDC
RECEIVED
MAR 22 1978
RECEIVED

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

AD A051607

AD NO. 1
DDC FILE COPY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING |
|--|-----------------------|--|
| 1. REPORT NUMBER 14 CHI-FIN-001 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) 6 SPEECH COMPRESSION PROTOTYPE DEVELOPMENT. Final Technical Report Research on speech compression prototype development | | 5. TYPE OF REPORT & PERIOD COVERED 9 Final Technical Report. May 1977 - December 1977 |
| 7. AUTHOR(s) 10 Judith Bruckner, Michael McCammon | | 6. CONTRACT OR GRANT NUMBER(s) 15 MDA903-77-C-0220, ARPA Order-3402 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Culler/Harrison, Inc. 150-A Aero Camino Goleta, California 93017 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Code: P7P10 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Defense Supply Service -- Washington Room 1D 245, The Pentagon Washington, D.C. 20310 | | 12. REPORT DATE 11 December 1977 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) DCASMA Oxnard 500 Esplanade Dr., Suite 990 Oxnard, California 93030 | | 13. NUMBER OF PAGES 14 42 15p. |
| | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 16a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce for sale to the general public. | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No. 3402. | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number) ARPANET Digital Voice Transmission Network Voice Conferencing Speech Compression | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report covers work performed by Culler/Harrison, Inc. (CHI) under contract no. MDA903-77-C-0220 for the period May 1, 1977 to December 31, 1977. During this period, Culler/Harrison, Inc. has analyzed trace recordings of network voice conferences to determine the nature of delays and range of variations in network transit time, and implemented an automatic delay adjustment algorithm to enhance reproductive quality. | | |

CONTENTS

| | Page |
|--|------|
| I. INTRODUCTION AND SUMMARY | 1 |
| II. EFFECT OF PACKET SIZE ON REPRODUCTIVE QUALITY. | 2 |
| III. AUTOMATIC DELAY ADJUSTMENT | 3 |
| REFERENCES | 11 |

| | |
|---------------------------------|---|
| ACCESSION FOR | |
| NTIS | White Section <input checked="" type="checkbox"/> |
| DDC | Soft Section <input type="checkbox"/> |
| UNANNOUNCED | <input type="checkbox"/> |
| JUSTIFICATION | |
| BY | |
| DISTRIBUTION/AVAILABILITY CODES | |
| Dist. | AVAIL. and/or SPECIAL |
| A | |

CHAPTER 1. INTRODUCTION AND SUMMARY

The reproductive quality of voice communication on a packet switched network is enhanced by reducing the total delay between generation of the original signal and its reproduction at the destination, and by maintaining as much as possible, the continuity or smoothness of the reproduction. In general, these two goals cannot be pursued simultaneously: smoothness can be ensured by accepting a sufficiently long delay, while minimizing delay can be accomplished only at the expense of losing part of the transmitted message or "sliding time", which affects smoothness adversely.

The delay factor depends upon packet size, network transit time and initial wait between receipt and playback of the first packet of a burst of communication. Smoothness depends upon the same factors. Since the sender and receiver have no control over network transit time, they must pursue an optimal strategy of choice of packet size and wait factor to maximize reproductive quality.

During this quarter, we analyzed trace recordings from previous conferences to determine the nature of delays and the range of variations in network transit time. We conjectured that a system of automatic adjustment of the wait factor to comply with current network conditions would be a good method for reducing delay to the minimum consistent with smooth reproduction. Such a method was implemented in our LPC conference programs and tested in conferences with ISI.

Details of the algorithm are discussed in Section 3 which also contains graphs of delays for a conference in which automatic delay adjustment was employed, as well as for a conference in which a single delay factor was used throughout the conference. The effect of packet size on reproductive quality is discussed in Section 2.

CHAPTER 2. EFFECT OF PACKET SIZE ON REPRODUCTIVE QUALITY

Let us confine our discussion to a single burst of speech containing no periods of silence. Assume that time is measured in units of frames, that a parcel contains the parameters for one frame and that a packet (or message) contains n parcels. If the data from the first parcel was generated at time t_0 , the packet cannot be sent until time $t_0 + n$. There is thus a delay of at least

$$[n + \text{network transit time}]$$

between generation of data and its playback. At each node of the transmission, the entire packet must be received before it can be resent. Thus the network transit time itself increases with n .

If the message length (in terms of the number of parcels) varies, the wait W (between receipt and playback of the first packet of the burst) should be at least as long as the maximum number of parcels in any message of the burst, minus the number of parcels in the first packet.

$$W \geq \max. \text{ PC} - \text{PC of first message} \quad (1)$$

Otherwise, assuming the network transit time was approximately constant for packets in the burst, the time for playback of the longest message would arrive before that message itself was received. This might also happen for other messages of the burst. See Figure 1 for a schematic drawing of this situation.

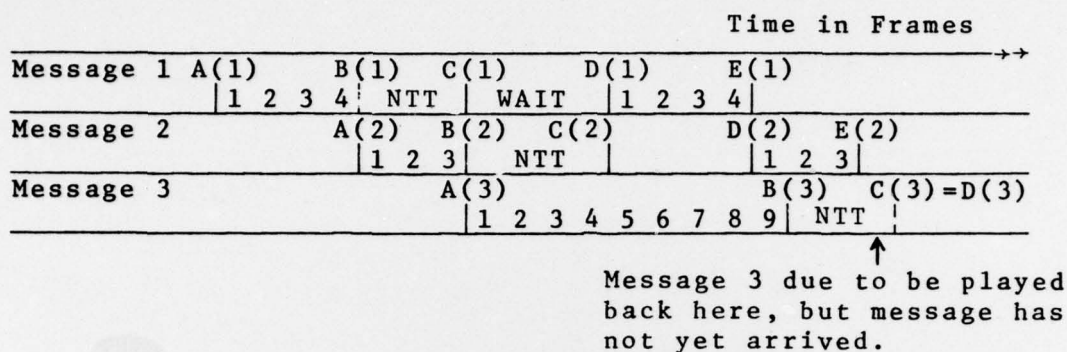


Figure 1

A(i) - represents the time the first parcel of message i was generated.
 B(i) - time the message was sent.
 C(i) - time the message was received.
 D(i) - beginning of playback for message i.
 E(i) - end of playback for message i.

- NOTE: 1. For each message, the data is represented by a count of parcels (e.g., message 1 has four parcels).
2. $C(i) - B(i)$ is network transit time, assumed here to be equal to 3.
3. $A(i+1) = B(i)$
4. $D(i+1) = E(i)$ for smooth playback.
5. The wait, W, in this example is 4.
6. For message 3, length = 9, length of message 1 = 4 and $W = 4$. Relation (1) does not hold. Thus message 3 is due to be played back before it is received.

In setting packet size, there are a number of considerations. If packet size varies, the longest message must be allowed for in setting W, otherwise smoothness will suffer; yet using the longest message increases the overall delay. On the other hand, sending packets of constant parcel length facilitates smoothness of reproduction. If that constant parcel count is small, overall delay is reduced. However, sending very short packets would be an inefficient use of the network. If all transmissions on the network were in

minimal packets, the network transit time might degrade.

Analysis of trace recordings disclosed two algorithms were used for packing messages for transmission.

1. Send a packet as soon as a preset minimum number of parcels m has been generated, as long as a maximal bit length has not been exceeded. Increase message size to maximal bit length whenever a backlog of messages to be sent builds up.

2. Send a packet as soon as a preset minimum number of bits has been generated, so long as a maximum parcel count M has not been exceeded. Increase message size to maximal bit length whenever a back log of messages to be sent builds up.

Method 1 results in messages which have relatively constant parcel count but may differ radically in bit count. Method 2 leads to messages with stable bit count but whose parcel count varies widely. If $m < M$, a shorter delay W would be needed for Method 1 than for Method 2. Thus, from the point of view of smoothness of reproduction and minimal delay, the first method is superior.

CHAPTER 3. AUTOMATIC DELAY ADJUSTMENT

Let us confine our attention to an interval of speech preceded and followed by silence and comprising a minimum of M messages. We use the following abbreviations for factors associated with a packet.

TG - time at which the first frame of the message was generated
TS - time the message was sent
TR - time the message was received
TD - time the message was due to be played back
TP - time at which the message was played back
OT - observed network transit time for the message ($=TR-TS$)
NT - expected transit time for the message (since OT varies drastically from packet to packet, NT is a smoothed version of OT)
var NT - variation in network transit time ($=OT-NT$)
PC - parcel count of the message
D - delay before playback of first packet

D is a fixed quantity for the interval; all other factors can vary. Since it is not known at the time the first message is received whether its parcel count is large or small, the time the first message should be played back is calculated, not in terms of the time sent or the time received, but in terms of the time the first parcel was generated as:

$$TP = TG + D + NT \quad (2)$$

D must accommodate variation in parcel count and variation in network transit time.

The time later messages are due to be played back is determined by the requirement of continuity or smoothness:

$$TP(i+1) = TP(i) + PC(i)$$

The time the $i+1$ st message is due to be played out is the time the i^{th} message is played out plus the parcel count of the i^{th} message. If a message has not been received when it is due to be played, its

playback is delayed until it arrives ("sliding time"). However, if a later message is due to be played, and has arrived before the packet in question is received, playback continues with the later message. The earlier message is considered to be lost. If it subsequently arrives, it is discarded.

The time at which a message is received is:

$$TR = TG + PC + OT \quad (3)$$

Now $TP-TR$, for messages in an interval, gives a good indication of whether the choice of the delay factor was optimal for quality reproduction of the interval. $TP-TR$ can never be negative, since a message cannot be played back before it was received. But, if it was often 0 for messages in the interval, that is an indication that time was forced to "slide" and smoothness was poor. Similarly, if $TP-TR$ stayed large for the interval, D could have been smaller and the continuity of playback would not have suffered. To see this, we calculate $TP-TR$ from (2) and (3)

$$TP-TR = D-(OT-NT)-PC = D-(varNT+PC) \quad (4)$$

$$\min (TP-TR) = D-\max(varNT+PC)$$

where the min and maximum are taken over all messages in the interval.

If there are a sufficient number of messages in the interval (number of messages greater than M), maximum var NT and maximum PC are approximately constant from interval to interval. Then a decrease in D (for the next interval) will result in an approximately equal decrease in $\min (TP-TR)$ for that interval while an increase in D would result in an increase in $TP-TR$ for the next interval. Of course, such a relationship is overridden for large decreases by the fact that $TP-TR \geq 0$.

In automatic delay adjustment, $TP-TR$ is calculated for each message received during a time interval. The minimum of those values, m , over the time interval is used to adjust the factor D used for the next interval.

Let ϵ be the desired minimum delay over an interval and m the minimum $TP-TR$ for that interval. D is calculated as follows:

$$\text{new } D = \begin{cases} (\text{old } D)+2\epsilon & , \text{ if } m=0 \\ (\text{old } D)+(\epsilon-m), & \text{ if } 0 < m \leq 3\epsilon \\ (\text{old } D)-2\epsilon & , \text{ if } 3\epsilon < m \end{cases}$$

These equations mean that if 'new D' had been used during the last interval (rather than 'old D') and $0 \leq m \leq 3\epsilon$, then the minimum TP-TR would have been exactly ϵ . If m was greater than 3ϵ in the last interval, minimum TP-TR would have been decreased by 2ϵ had 'new D' been used. The last equation limits the amount D can change in any one adjustment.

The calculation ("adjustment") of D is performed whenever a message arrives after a period of silence and a preset number of messages has been generated since the last calculation. If an insufficient number of packets has been received, statistics on TP-TR continue to accumulate, and the old D is used to determine TP for the first packet of the burst.

D should "home in" on the minimum delay which preserves continuity of playback.

Figures 2 and 3 show a schematic picture of the relationships.

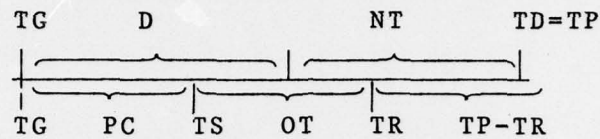


Figure 2

Schematic representation of TP-TR for a message which was received before it was due to be played back.

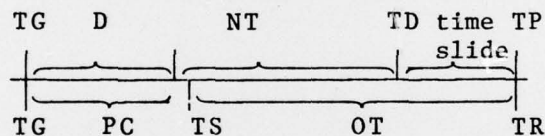


Figure 3

Schematic representation for a message which was not received before it was due to be played back. $TP-TR=0$

For convenience, the abbreviations used here are repeated:

- TG - time first parcel generated
- TS - time sent
- TR - time received
- TD - time due to be played back
- TP - time played back
- NT - expected network transit time
- OT - observed network transit time
- PC - parcel count
- D - delay

Shown below are graphs of TP-TR for parts of two conferences. Figure 3 illustrates the situation for a conference which did not have automatic delay adjustment. D was set at 60 for the entire conference. When one participant spoke, D was insufficient and time was frequently forced to "slide". When the other participant spoke, D was overgenerous and could have been reduced without affecting continuity.

The conference depicted in Figure 4 included provision for automatic delay adjustment. D was initialized at 60 and ϵ set to 5. While the first participant spoke, there was no silence, and D was not adjusted. During the interval the second participant spoke, D was adjusted until minimum TP-TR approached 5.

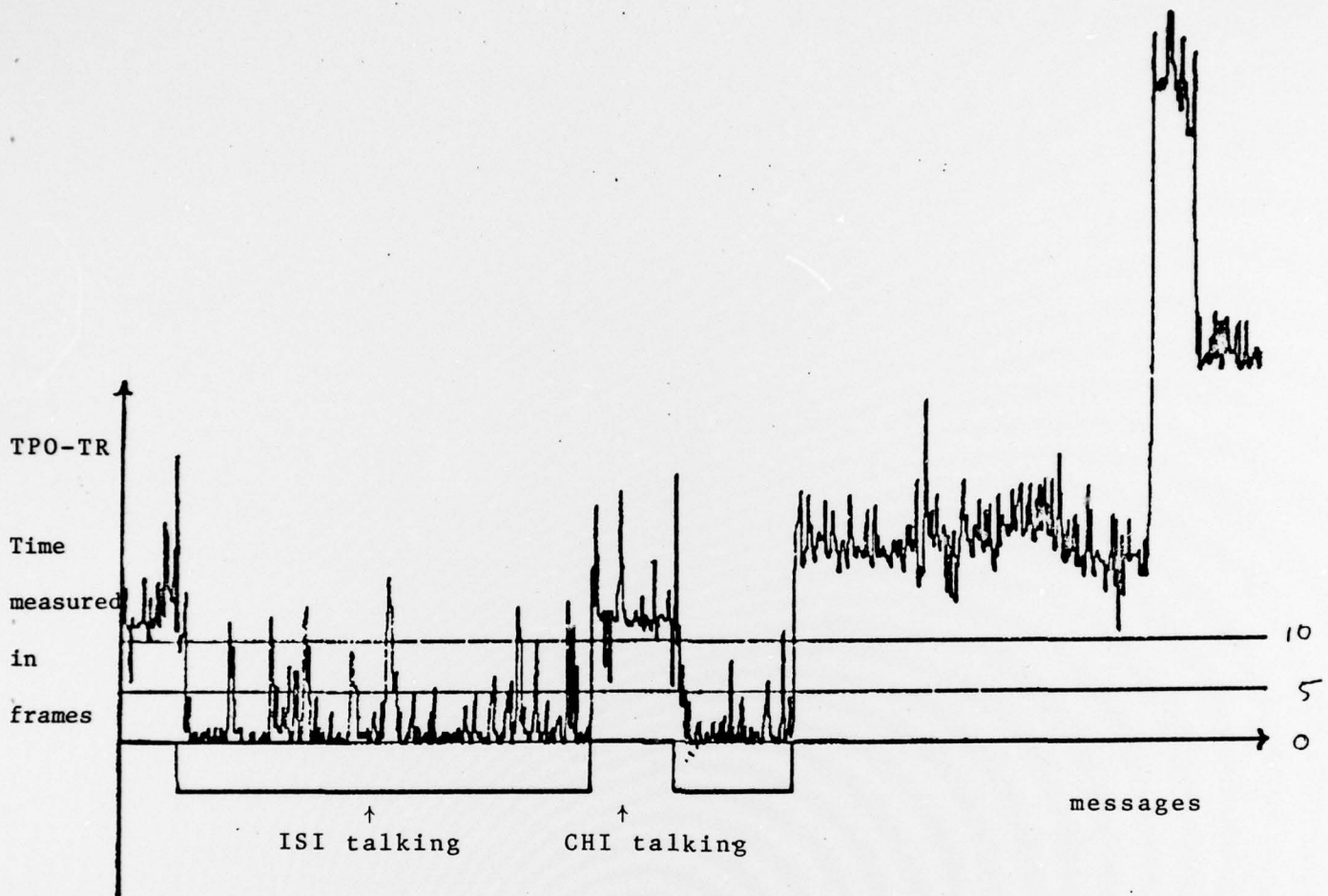


Figure 3
TP-TR for 750 messages. No automatic delay.

D=60 throughout conference. The 0,5 and 10 level lines are shown. Intervals when ISI was talking are underlined; CHI talking in remaining intervals. When ISI was talking, TP-TR was frequently 0, indicating time slide. During last interval of CHI speech, D could have been smaller and continuity would have been preserved.

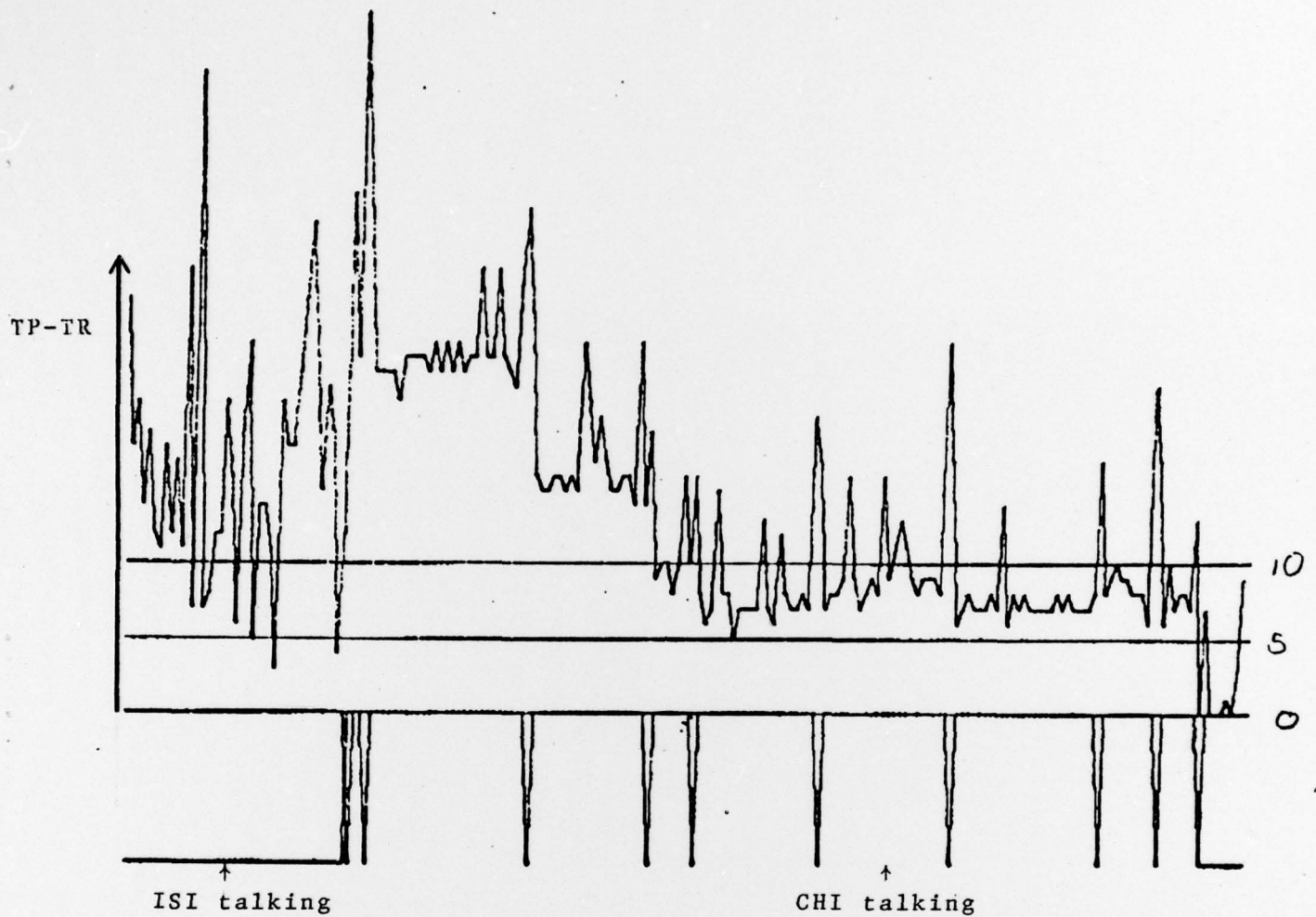


Figure 4
TP-TR for 200 messages with automatic delay adjustment

0,5 and 10 level lines indicated. Intervals when ISI was talking underlined. CHI talking during middle interval. Downward spikes indicate silence bit on for corresponding message (delay adjusted if preceding interval contained at least 20 messages). Min TP-TR approaches 5 ($\epsilon=5$) after two adjustments. Smooth playback throughout conference.

REFERENCES

1. Cohen, D., Issues in Transit Packetized Voice Communication, NSC Note 95, VSC Information Sciences Institute, Marina del Rey, California, July 1976.
2. Culler, G.J., McCammon, M., A Variable Frame Rule Network Voice System, Quarterly Technical Report, Speech Compression Research at CHI, Culler/Harrison, Inc., Goleta, California, September 1976.