DOCUMENT RESUME

ED 021 851 56 TE 000 672

By- Carroll, John B.; Cramer, H. Leslie

THE INTELLIGIBILITY OF TIME-COMPRESSED SPEECH. FINAL REPORT.

Harvard Univ., Cambridge, Mass. Lab. for Research in Instruction.
Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No- BR-5-0958
Pub Date Jul 68
Contract- OEC-7-31-0370-271
Note- 36p.
EDRS Price MF-$0.25 HC-$1.52
Descriptors- AUDIO EQUIPMENT, BLIND, COMMUNICATION (THOUGHT TRANSFER), COMPREHENSION, INFORMATION THEORY, INTONATION, *LANGUAGE RESEARCH, LECTURE, *LISTENING COMPREHENSION, RESEARCH METHODOLOGY, SENTENCES, *SPEECH, *SPEECH COMPRESSION, TAPE RECORDINGS, TIME

Time-compressed speech is now being used to present recorded lectures to groups at word rates up to two and one-half times that at which they were originally spoken. This process is particularly helpful to the blind. This study investigated the intelligibility of speech processed with seven different discard intervals and at seven rates from two to five times the original. Optimal parameters for processing speech at each of the rates are reported along with a comparison of intelligibility rendered by three different modes of presentation. Headphone presentation was found to be about 22% more intelligible on the average than presentation by loudspeaker. The commercial equipment available for processing time-compressed speech varies the sampling interval but maintains a constant discard interval. The findings indicate that the sampling interval should be held constant for rates two to five times normal. A sampling interval of 15 milliseconds was found to be optimal at all rates for a low-pitched man's voice. A higher pitched woman's voice would require an even shorter sampling interval. Processing and presenting time-compressed speech using the optimal parameters reported here should promote greater comprehension as well as intelligibility at all rates, intelligibility being a necessary prerequisite for comprehension. (Author)

BR-5-0958

5-0958

PA-56

FINAL REPORT
Contract No. OE-7-31-0370-271

THE INTELLIGIBILITY OF TIME-COMPRESSED SPEECH

July 1968

TE 000 672

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Contract No. OE-7-31-0370-271

The Intelligibility of Time-Compressed Speech

John B. Carroll
and
H. Leslie Cramer

Laboratory for Research in Instruction
Graduate School of Education
Harvard University

Cambridge, Massachusetts 02138

July, 1968

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

PREFACE

The following report was originally given as a paper, the 18th of 20 presented at the Library of Congress Conference on Time Compressed Speech, held at the University of Louisville, Kentucky, October 19-21, 1966. It was subsequently published as Chapter 18 (pgs. 126-148) in Foulke, E. (ed.), The Proceedings of the Louisville Conference on Time Compressed Speech, May, 1967, Center for Rate Controlled Recordings, University of Louisville, Kentucky. Since it was presented toward the end of the Conference, the explanation of the process of time-compression and definitions of terms was not included. The reader who is interested in this process is urged to refer to the complete Proceedings, which are available at $2.00 from the University of Louisville, or from ERIC. A much more detailed descrittion of the two studies contained in this report is available in Cramer, H.L. The Intelligibility of Time Compressed Speech, unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, Mass., 1968.

TABLE OF CONTENTS

	Page
PREFACE	ii
LIST OF TABLES	iii
LIST OF ILLUSTRATIONS	iv
TIME COMPRESSION PROCESS	1
INTELLIGIBILITY MEASUREMENT	1
METHODS	3
RESULTS AND FINDINGS	5
REFERENCES	26
ERIC REPORT RESUME	31

LIST OF TABLES

Table		
1.	Latin Square Design for Pilot Study	4
2.	Amount of Delay for Each of the 7 Conditions in the Pilot Study	4
3.	Analysis of Variance Table for Pilot Study	5
4.	Latin Square Design for Main Study	12
5.	Discard Intervals for Tape Series A to C	13

LIST OF ILLUSTRATIONS

Figure		Page
1.	Block Diagram of Playback Apparatus for Pilot Study	6
2.	Pilot Study Intelligibility Plotted vs. Speech Rate with 7½ MS Inter-aural Time Difference and No Inter-aural Time Difference in Presentation . . .	8
3.	Number of Errors in Pilot Study with Seven Amounts of Delay at Four Times Normal Speech Rate	9
4.	Oscillograph Tracing of the Beginning of the Word "John"	10
5.	Oscillograph Tracing of Figure Number 3-8 with the Same Tracing Offset on Itself by One Pitch Period	10
6.	Block Diagram of Audio Playback Apparatus for Main Study	14
7.	Intelligibility vs. Discard Interval for Stereophonic Mode of Presentation . .	15
8.	Intelligibility vs. Discard Interval for Monophonic Mode of Presentation . . .	16
9.	Intelligibility vs. Discard Interval for Loudspeaker Mode of Presentation . .	17
10.	Intelligibility vs. Sampling Interval for Stereophonic Mode of Presentation . .	18
11.	Intelligibility vs. Sampling Interval for Monophonic Mode of Presentation . . .	19
12.	Intelligibility vs. Sampling Interval for Loudspeaker Mode of Presentation . .	20
13.	Tempo Regulator Parameters Compared with Optimum Sampling Intervals	23
14.	Main Study Intelligibility vs. Speech Rate for both Monophonic and Loud Speaker Presentation	24

LIST OF ILLUSTRATIONS (continued)

Figure Page

15. Intelligibility vs. Plotted Speech Rate for Optimum Conditions in Pilot Study with Delay and Main Study Monophonic and Loudspeaker Presentation . . . 25

The Intelligibility Of Time-Compressed Speech

Time Compression Process

If we make a magnetic tape recording of speech at 7 1/2 inches per second (ips.) and play it at 15 ips., it can be heard in half of the original time in which it was recorded. All the frequencies double, however, giving a high pitched, so-called "Donald Duck" effect. Speech so processed and played back is rather unpleasant and practically unintelligible. With the development of magnetic tape recording, it has become possible to halve play back time by another method without the attendant frequency shift of the speeded-up play back. This other method involves cutting a tape recording every one quarter of an inch from beginning to end, and discarding alternate pieces. The remaining pieces are then spliced together to make a reconstituted tape that is half the length of the tape as originally recorded. The play back of a tape processed by this chop-splice method, will sound normal however, as far as the pitch of the speaker's voice is concerned, although it will obviously sound as though words were spoken rapidly.

Garvey and Henneman (1950) investigated word intelligibility of speeded-speech produced by this chop-splice method. Garvey (1953) reported that the intelligibility remains high, above 90 per cent, at a word per minute rate 2 and 1/2 times the original recording. Garvey further reports (1965) that after completing his thesis with the chop-splice method that he never wanted to see another tape or splicer; and it is fortunate for researchers that Fairbanks, Everitt, and Jaeger (1953, 1954, 1959) developed an electro-mechanical system for automatically discarding segments of speech from a recorded tape. Their system uses four playback heads mounted in a rotating drum to scan a magnetically recorded tape. The effective time length of each speech sample scanned and retained is equivalent to the speech time on each piece of tape spliced together in the chop-splice method, and is called the sampling interval. This sampling interval is determined by the revolutions per minute of the rotating head assembly. The time length of each speech sample eliminated is likewise similar to the speech time on each piece of tape discarded in the chop-splice method, and is called the discard interval. The discard interval is determined by the speed of the magnetic recording tape around the rotating head assembly. The output of Fairbanks' system is thus equivalent to the output produced by the chop-splice method.

Intelligibility Measurement

Previous studies dealing with the intelligibility of time-compressed speech have used phonetically balanced spondaic-word lists (lists of two syllable words with equal stress on each syllable, eg. horseshoe). (Fairbanks and Kodman 1957). A restricted list of fifty words is presented after time has been allowed for the listener to familiarize himself with the words in the list, both by studying the written lists, and by hearing the list in order while reading them. The lists are then presented many times in different randomized orders. There is no known previous investigation of the intelligibility of time-compressed words in context.

Fairbanks, Guttman, and Miron (1957) investigated comprehension of compressed speech by using 1500 word technical passages presented at high word-per-minute (wpm) rates,* tested by multiple choice type questions. However, it was not clear which of three possible causes might be attributed to a wrong answer: (1) the subject did not hear every speech sound because the length of the discarded sample was so long that whole sounds were dropped; (2) the distortion introduced by the interruption frequency of the compression equipment was interfering with the signal; or (3) there was a problem in perceiving at a rapid rate (i.e., difficulty in cognitive processing). Unfortunately, this work was done only with a discard interval of .02 seconds. The only speech compression apparatus commercially available is the Eltro Information Rate Changer manufactured by Telefonbau and Normalzeit, Frankfurt-am-Main, Germany. This equipment uses a .04 second discard interval at all compression ratios, and cannot be altered. Tests run to determine the comprehension of speeded speech by both blind pupils (Bixler, Foulke, Amster, & Nolan, 1961, Foulke and Bixler, 1963 & 1964, Foulke, Amster, Nolan, & Bixler, 1962) and sighted pupils (Orr and Friedman, 1964, Friedman, Orr, Freedle, and Norris, 1966, Friedman, Orr, and Norris, 1966, Voor, 1962, Wood, 1965) used this equipment for compressing their materials.

Fairbanks and Kodman (1957) tested word intelligibility as a function of time compression. Their curves suggest that the optimum rate of interruption and length of discard interval is not a constant for all compression ratios. At compression rates up to 75%, a discard interval of .01 sec. appears best; at 80 to 85%, .06 sec., but at 90%, .05 sec. When listening to connected discourse, a person has cues to the words from both the context and the grammar of the sentences (Miller, Heise and Lichten 1951, Goldman-Eisler, 1958 and 1961, Pollack 1964, Miller 1962, and Savin 1963.)

The research reported here attempts to test the middle ground between word intelligibility of words in isolation and comprehension of long passages by testing the word intelligibility of short sentences. This research was done in two parts, the first being a pilot study to see if a difference of 15 to 30 milliseconds (ms.) in the time of presentation of sentences to the two ears would improve the intelligibility of time-compressed speech. The second part was the main study and deals with the intelligibility of the same passages used in the pilot study. These were presented at seven different compression ratios, each processed with seven different discard intervals. The passages used were from the Harvard Psycho-Acoustic Laboratory (P.A L.) Auditory test no. 12, (Hudgins, et al., 1947) which consisted of seven passages of 28 questions each. The P.A.L. Test passages were preceded by the following introductory passage which was used to help motivate the students and also to allow a gradual increase of speech rate to approximately the starting rate of the lists. Here is the introductory passage and the first ten sentences.

> This is an experiment to determine the intelligibility of speeded speech. You will hear a passage, which has been specially processed, in less time than it took the reader to read it. Even these instructions to which you are now listening have been speeded by 20% as compared with the original The reader's voice sounds normal, however, rather than high-pitched, as it would if a phonograph record of it were being played at a higher speed than that at which it was recorded. Speech processed in this manner is already being used to present material to

*Since the average length of words may vary, Carroll (1964) proposes that the syllable-per-minute rate would perhaps be a better measure, with the speed of 265 syllables-per-minute as the average rate of normal speech production.

blind persons at rates up to 475 w.p.m. A normal speaking range varies from a low 125 w.p.m. to a rapid 200 w.p.m. or more. The blind high school student averages only 90 w.p.m. when reading braille while the average sighted high school student reads books at 200 w.p.m. Some blind people using both hands simultaneously to read braille can reach reading speeds as high as 225 w.p.m. However, this is exceptional, and less than 10% of the blind learn to read braille at all. This experiment is designed to test the intelligibil of compressed speech at rates up to 800 w.p.m., increasing by increments of 50 words per minute over a series of seven passages. I is contemplated that the results of this study will lead to improved methods of presenting verbal material to the blind. It also seems that such rapid speech can be useful in presenting tape recorded reviews of lectures at rates so high that the 50 minute lecture can be heard in as little as 10 to 12 minutes.

The following passages are from the Psycho-Acoustic Laboratory Test Number 12, published by Harvard University.

List Number 1

1. In what country is Chicago?
2. What letter comes after Y?
3. What is the color of grass?
4. What number comes before 7?
5. What part of the body do you write with?
6. What comes out of a kitchen faucet?
7. What number is between 8 and 10?
8. Which is wetter, water or sand?
9. Do you dig holes with a shovel or a rake?

Methods

For the pilot study this tape was presented by means of earphones to Radcliffe students, who heard it from seven different tapes, each in the same ascending order of compression, starting at 50% compression, (494 syllables pe minute or) 398 words per minute, with a 1.24 syllable-to-word ratio. The sub were screened to eliminate any who had more than 5 decibels hearing loss in e ear or more than 3 decibels difference between ears using the Central Institu for the Deaf (CID) Auditory Test W-2. (Davis, et al., 1964, pg. 535)

This pilot study which preceded the main study was performed to dete whether there was any difference in intelligibility when the speech was delay one ear by seven different amounts including zero delay. Table 1 shows the d for this study. Table 2 shows the amounts of delay which were obtained by ad of a micrometer head. A playback head was attached to the micrometer so that head could be moved along the tape on which the passages were recorded. The

TABLE 1

Latin Square design for the Pilot Study showing distribution of passages from the P.A.L. Auditory Test No. 12. Each subject heard 7 passages as shown in each row, in ascending order of compression ratios. All passages were compressed at .035 sec. discard interval.[1]

Speed-up Factor		2.0	2.33	2.67	3.0	3.33	3.67	4.0
Compression Ratio		50.00	57.15	62.40	66.60	70.00	72.75	75.00
Syllables/min.		494	576	658	740	822	905	987
Words/min.		398	464	531	597	663	730	796
Tape Numbers	Subjects Numbered							
1	01-07	2	3	4	5	6	7	8
2	08-14	4	8	7	2	5	3	6
3	15-21	8	2	6	3	7	5	4
4	22-28	7	6	3	4	2	8	5
5	29-35	5	7	2	6	8	4	3
6	36-42	3	5	[illegible]	7	4	6	2
7	43-49	6	4	5	8	3	2	7

Each Tape (numbered 1-7) was played 7 times. The first subject hearing each tape had the same material presented at the same time in each ear. Each of the six subsequent subjects to hear each tape had the material alternately delayed at one ear (alternately right and left on every question). The seven amounts of delay, including zero delay, were as follows:

TABLE 2

Amount of delay for each of the 7 conditions in the Pilot Study

Delay Condition No. & Subjects Number for Table number 1	1	2	3	4	5	6	7
Delay in Milliseconds	0.0	0.5	1.0	4.0	7.5	15.0	30.0
Micrometer Head Setting in Inches	0	.00375	.00750	.03000	.05625	.11250	.22500

[1]Reader averaged 199 words/min. or 247 syllables/min. across 7 lists for a syllable-to-word ratio of 1.24.

was calibrated in 10 thousandths of an inch to allow precise settings. A special channel switch was constructed which was activativated by a voice key so that the delayed presentation was automatically changed from the right to the left ear for every other sentence. Figure 1 is the block diagram of the instrumentation used for presentation of the compressed speech to the students.

Results and Findings

An IBM card was punched for each page of each subject's answer booklet. Each sentence was punched as 0 if wrong or 1 if correctly answered. The last two columns of the card contained the total number of the 28 sentences in each list which were correctly answered with no corrections for guessing The dependent variable for the analysis of variance (ANOVA) was the raw score for each individual on each list This IBM format for punching the scores allowed pooling of the data by means of list numbers, amount of delay, and compression ratio. The following table (Table 3) shows the results of analysis of variance.

TABLE 3-3

ANALYSIS OF VARIANCE TABLE

Source	Degrees of Freedom	Sums of Squares	Mean Squares	F Ratio
I (list)	6	253.6289	42.2715	22.80*
J (delay)	6	43.2598	7.2100	3.88**
K (compression ratio)	6	3822.5254	637.0876	344.0*
I x J	36	66.8066	1.8557	
I x K	36	239.8320	6.6620	
J x K	36	121.3457	3.3707	
I x J x K	216	962.0313	4.4538	
Total	342	5509.4297		

*F = 3.90, significant > .001
**F = 3.535, significant > .0025

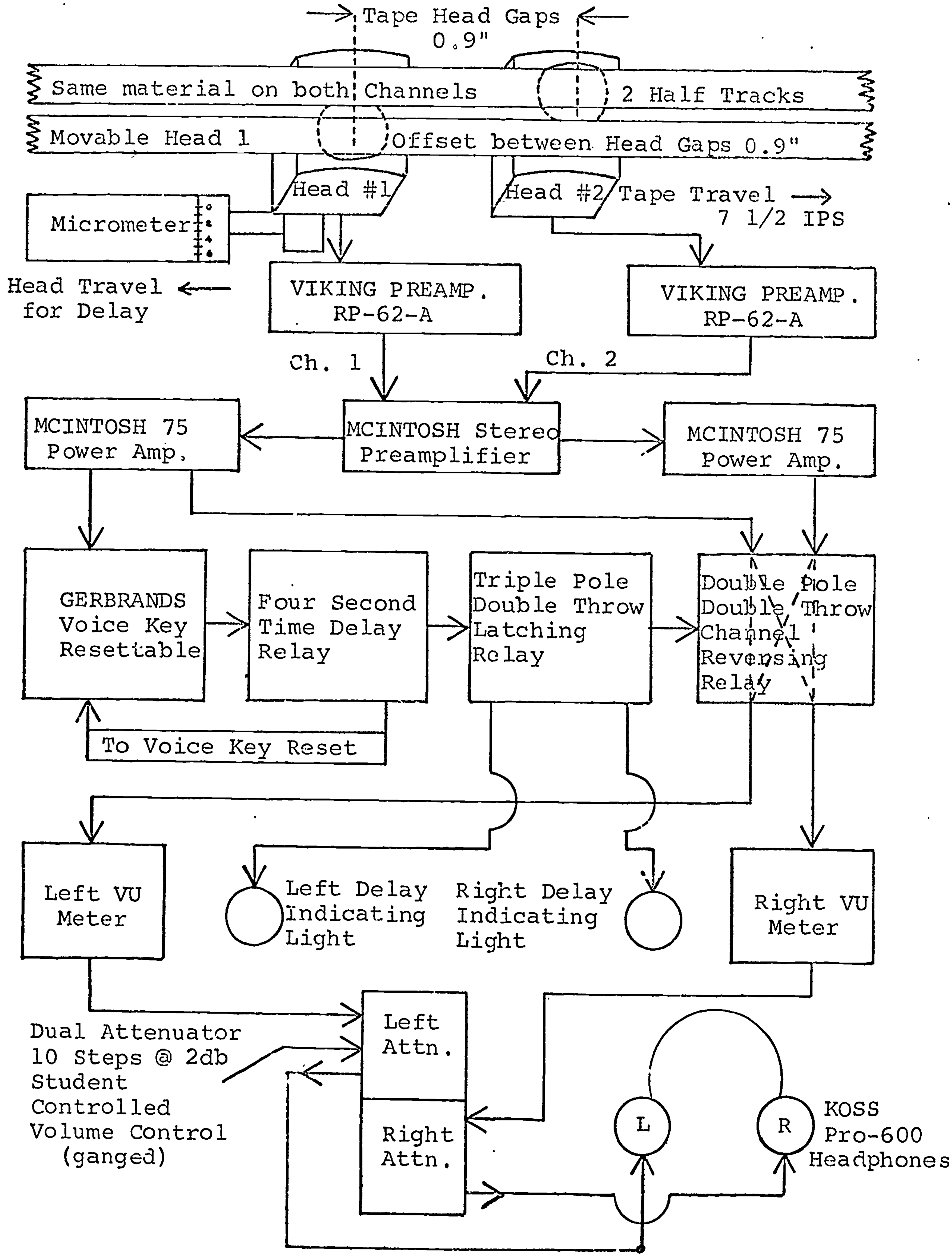


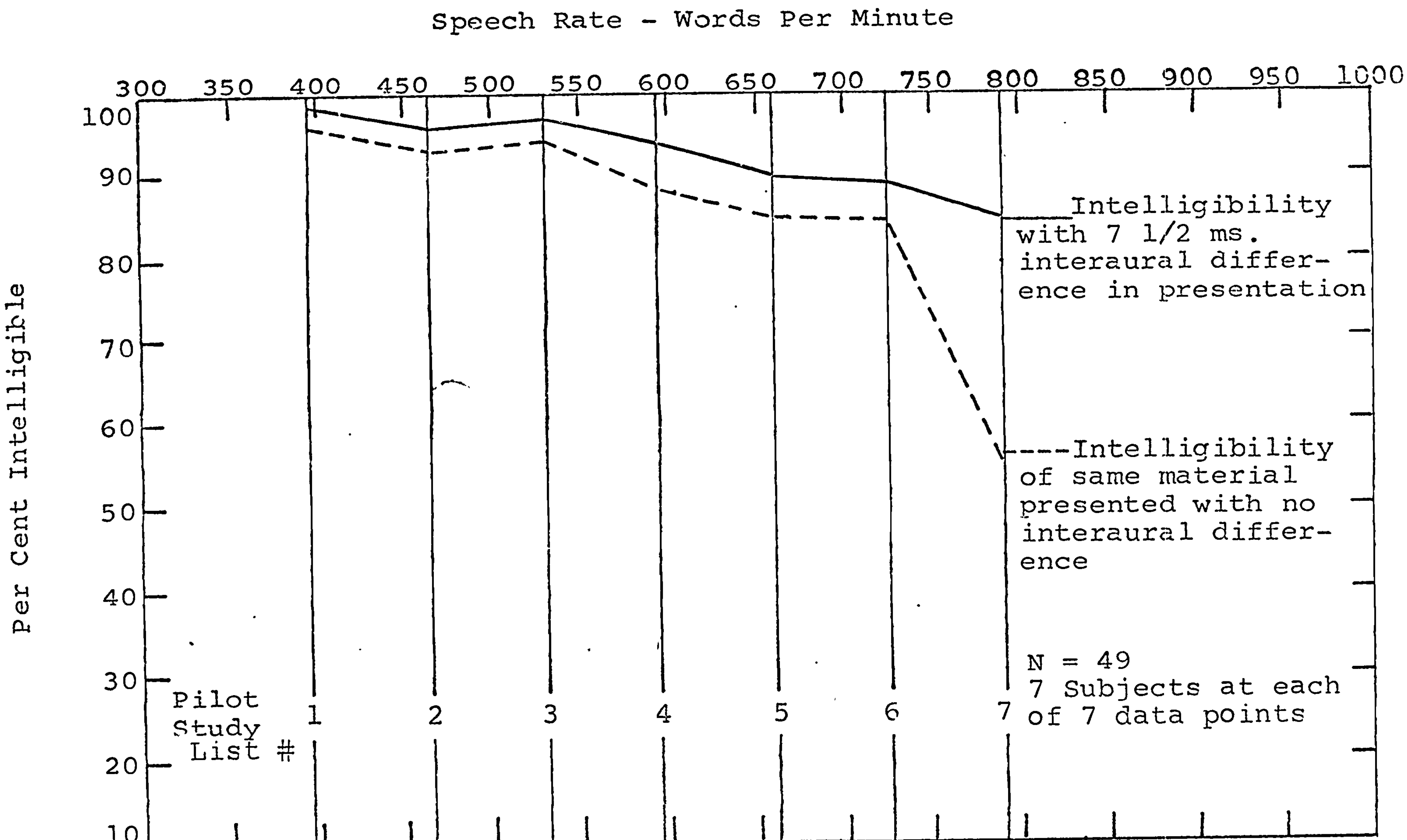
Figure 1 . Block Diagram of Playback Apparatus for Pilot Study

It had been originally hypothesized that a delay in one direction might be helpful for some students, while the opposite direction might be detrimental. It was thought that the direction of favorable delay if found, might have a relationship to the handedness of the individual, as the speech center in the brain is generally in the opposite cerebral hemisphere from the motor control center (Kimura, 1961a, 1961b, 1963, 1964). The results showed no significant differences in intelligibility as measured by the P.A.L. Test between even and odd items delayed to the left and right ears respectively. This negative finding does not, of course, preclude the possibility that there may be differences not detectable with this technique. Figure 2 shows the results of the pilot study for the 5th delay condition of 7 1/2 milliseconds (ms) which was the major effect found (Cramer, 1965). All of the amounts of delay other than 30 ms. (which produced speech less intelligible than that produced by no delay) showed a slight increase in intelligibility over the no delay condition It should be obvious that when the intelligibility is near 100%, little can be done to improve it In other words, speech has to be somewhat unintelligible in order for an inter-aural (difference in time of) presentation to facilitate an improvement

This writer has named this increase in intelligibility produced by delaying the speech at one ear the binaural redundancy effect. Figure 3 shows this effect when plotted as a curve The sharper drop of the curve at 7 1/2 ms. than at 15 ms. raises the speculation that the maximum effect would occur around 10 ms as shown by the dotted line This happens to be the average pitch period of the voice of Mr. Paul Clark, the reader of the passages The binaural redundancy effect suggests that the brain of a listener is able to correlate the inputs to his ears. Figure 4 shows an oscillograph tracing of a vowel sound It will be noted that successive pitch periods are highly similar. When we displace one tracing on the other by the amount of one pitch period (Figure 5) the match is excellent

Incidentally, it has come to our attention that the Air Force has contracted with Melpar, Inc in Falls Church, Virginia, to test the effect of delaying speech to one ear to improve the intelligibility of normal speech in the presence of noise (Reddinger 1966). It was found that a 7 1/2 ms. inter-aural difference in time of presentation of a voice signal allowed the introduction of almost twice as much noise as normally present with no loss in intelligibility.

In the main study the same passages were read by the same speaker, with the same introductory passage. However, since speech became seriously distorted only at the 75% compression ratio (equal to 4 times normal or 987 syllables per minute or 796 words per minute). it was decided to drop the 50% compression ratio (twice normal speed, 499 syllables per minute or 398 words per minute) from the study and start with a compression ratio of 57.19 (2.33 times normal speed, 576 syllables per minute, or 464 words per minute). The next rate was then omitted as used in the pilot study, namely, 62.4% compression (2.67 times normal speed, 658 syllables per minute or 531 words per minute) This allowed the addition of two higher rates at the end of the series, beyond the compression ratio of 75% (4 times normal speed) used in the pilot study. Each of the seven passages at seven compression ratios was then processed at seven different discard intervals



Compression Ratios	50.0	57.2	62.4	66.6	70.0	72.7	75.0
Speed-up Factor	2.00	2.33	2.67	3.00	3.33	3.67	4.00
Words Per Minute	398	465	531	597	663	730	796
Syllables Per Min.	499	576	658	740	822	905	987

Figure 2. Pilot Study Intelligibility Plotted Versus Speech Rate with 7 1/2 ms. Inter-aural Time Difference and No Inter-aural Time Difference in Presentation.

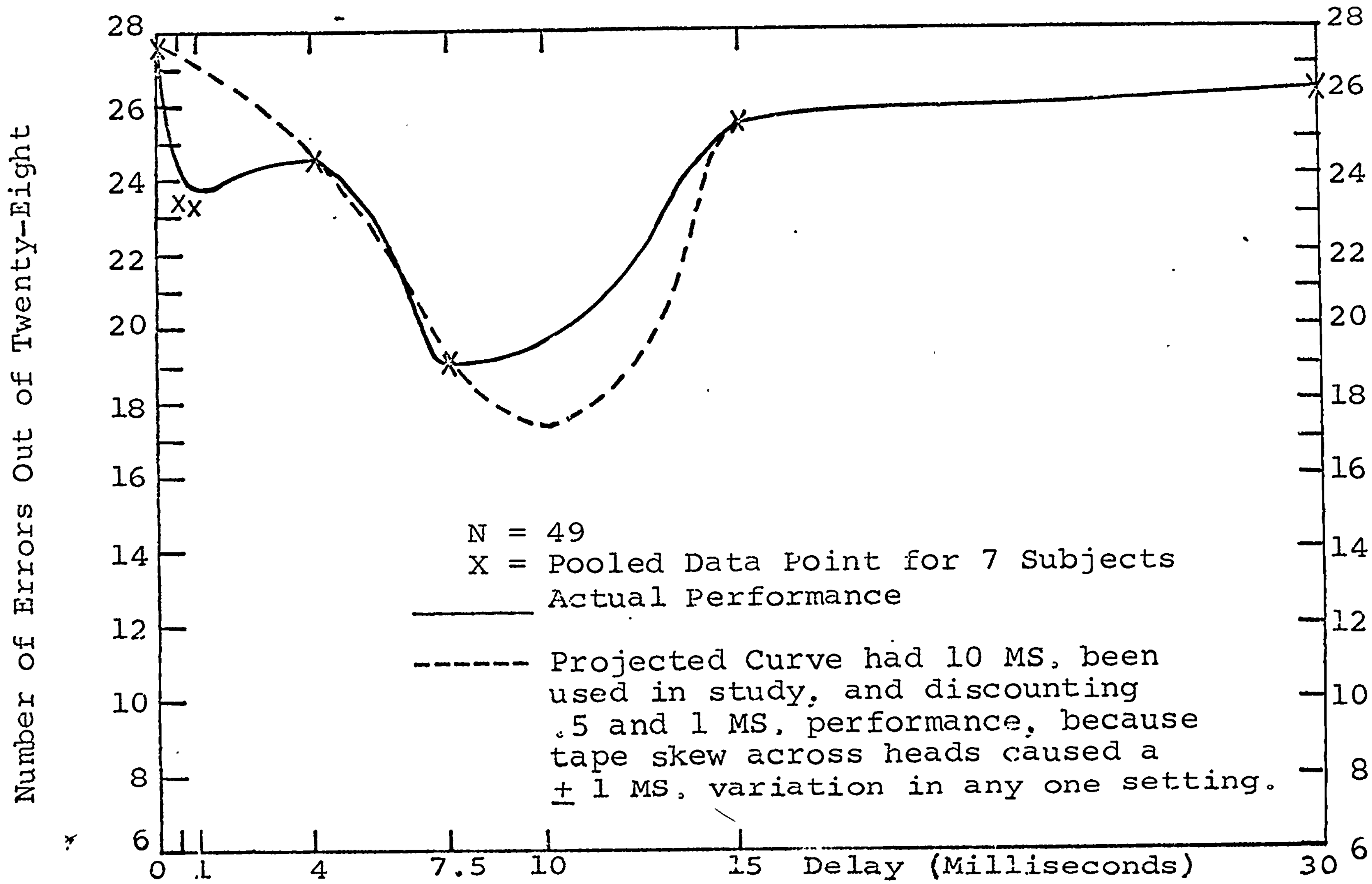


Figure 3 Number of Errors in Pilot Study with Seven Amounts of Delay with Four Times Normal Speech Rate (75% Compression. 795 WPM or 987 Syllables Per Minute)

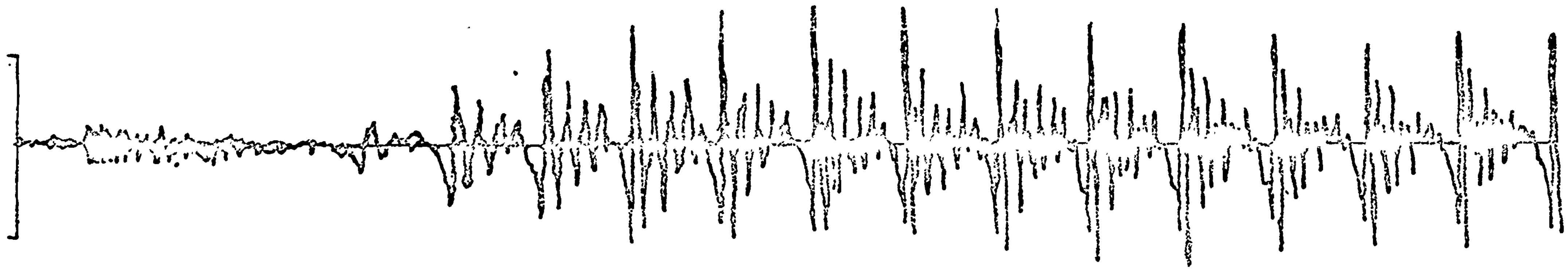


Figure 4 Oscillograph Tracing of the Beginning of the Word JOHN (This includes only about one half of the pitch periods of the /a/ sound)

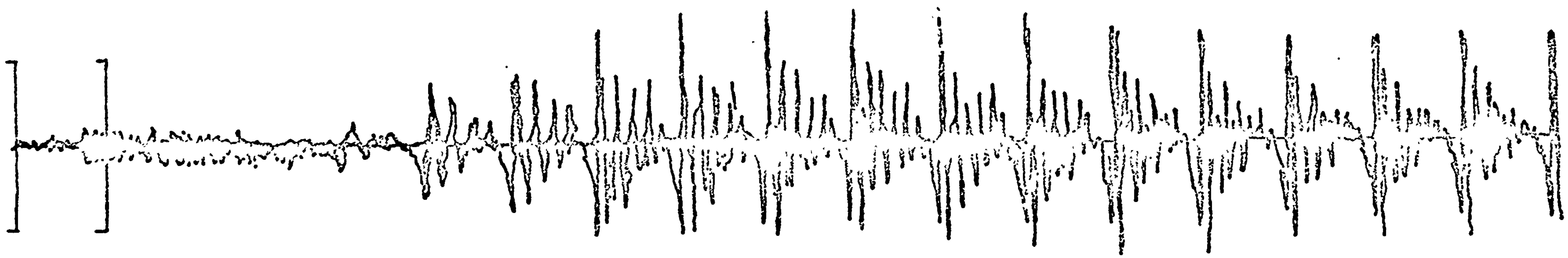


Figure 5 . Oscillograph Tracing of Figure Number 4 with the Same Tracing Offset on Itself by One Pitch Period

from 10 to 95 ms. The design for this part of the study is shown in Tables 4 and 5.

This design was used to determine the optimal discard interval for maximal intelligibility of speech at each of the seven rates shown in table number 3. Also to be tested was whether a 10 ms. inter-aural time difference of presentation would significantly improve intelligibility. A third condition, the use of a loudspeaker in a large reverberant classroom was also to be tested.

Voor (1962) expressed concern that the rapidity and short duration of each speech sound in highly compressed speech could cause a masking by reverberation when such sounds are presented by means of a loudspeaker in an ordinary classroom. To avoid this possibility, he presented all his material by means of earphones. Orr and Friedman (1964a, 1964b, 1964c), Friedman, Orr, Freedle, & Norris, (1966), Friedman, Orr, & Norris, (1966) used loudspeaker presentation, however, because they felt it represents a closer approximation to what is practical in the typical classroom. The equipment was set up as shown in Figure 6.

In the pilot study, the answers given by each student were scored as either right or wrong for each question by the correspondence to the correct answers as published with the P.A.L. Test Number 12. There were a few modifications necessary, for instance the question, "How many states are there in the Union?" is now correctly answered by 50 rather than 48, and the color of the cloth on a pool table is not always green today, decorators having entered the field. In one case the question, "Which is larger, a dog or a horse?" was misread "house", so that house had to be allowed as a correct answer. There were several questions, however, where an answer could be given without assurance that the question was correctly perceived. Such an example is "What do you use to unlock a door?" The answers included key, knob, and hand. The latter answer could also have been given if the subject thought the question was, "What do you use to knock a door?" (Asher, 1958). Because of such uncertainties, in this final part of the study, subjects were asked to write each utterance as they heard it, rather than answering the question as done in the pilot study. Their responses were then scored on a word-for-word basis. At first a syllable-for-syllable scoring basis was proposed. When we found responses such as "Which is Dr. Michael Day?" for the question "Which is darker, night or day?", "Does California grow the orange?" for the question, "Does a cow have kittens or horns?" and "What is a Catholic worth?" for "What does a cat lick with?", it became obvious that such scoring would give spuriously high scores, even when the response bore little relation to the original in meaning. We did allow plurals for singulars and "come" for "comes" and vice versa.

The additional time required for subjects to write what was heard would have lengthened the test time to more than one and one half hours; therefore, each list was cut in half arbitrarily. When the sheets were all collected and scored, the data were first plotted by discard interval. It appeared that something peculiar was occuring when the curves were plotted with intelligibility versus discard interval (as in Figures 7, 8, & 9). Because of the shapes of the tails of the curves and the shift of the optimum points to the right as the speed increases the intelligibility was then plotted as a function of the sampling interval. When the raw data were fitted by a least squares solution for a cubic equation, we obtained the smoothed curves shown in each case (Figures 10, 11, and 12). The broken line shows the raw data.

It will be observed that the optimum points (highest intelligibility point on each curve) indicated by arrow heads lie nearly in a straight line from top to bottom, showing that the optimum is a function of the sampling interval rather than the discard interval. Fairbanks was a phonetician and was concerned with the

TABLE 4

Latin Square Design for the Main Study showing order of PAL lists 2-8, Renumbered 1-7 respectively for this study. Blocks B to G were arranged by permuting the numbers in each cell as designated by Winer (1962, 1965).

Block A

Tape Series A-1 to A-7 I_d = 10 MS[1]		List numbers as announced on the tape heard by students						
		1	2	3	4	5	6	7
Compression Ratio		57.15	66.67	70.00	72.75	75.00	76.90	78.57
Syllables/min. Rate		576	740	822	905	987	1069	1152
Words/min. Rate		464	597	663	730	796	862	928
Speed-up Factor		2.33	3.00	3.33	3.67	4.00	4.33	4.67
Tape[2] Number	Subjects[3] Numbered							
A-1	01-03	6	7	1	2	3	4	5
A-2	04-06	1	5	4	6	2	7	3
A-3	07-09	5	6	3	7	4	2	1
A-4	10-12	4	3	7	1	6	5	2
A-5	13-15	2	4	6	3	5	1	7
A-6	16-18	7	2	5	4	1	3	6
A-7	19-21	3	1	2	5	7	6	4
		The above numbers are PAL Auditory Test #12, Lists 2-8, renumbered 1-7 respectively.[4]						

[1]The list numbers for each subsequent block from A as shown here to block G were cyclically permuted by increasing the passage numbers in block A by 1, except that number 7 became 1 as each was changed to render the next block in alphabetical order (See Appendix G pg. 215 for complete explication of Blocks A through G).

TABLE 4 footnotes--(continued)

[2]Each tape was heard by three subjects: the first with headset and interaural time delay, the second monophonically with headset, and the third by means of loudspeaker in a reverberant classroom.

[3]The subject numbers for each row of each subsequent block after A were increased by 21 so that subject numbers for tape B-1 would start 22, 23, etc., and tape C-1, 43, 44, etc.

[4]Reader averaged 199 WPM or 247 syllables per minute across 7 lists for a syllable-to-word ratio of 1.24.

TABLE 5

Discard Intervals for Tape Series A to G with 7 tapes in each series designation.

Tape Series Designation	A	B	C	D	E	F	G
Discard Interval in Milliseconds	10	20	35	50	65	80	95
Listened to by Subjects Numbered	1-21	22-42	43-63	64-84	85-105	106-126	127-147

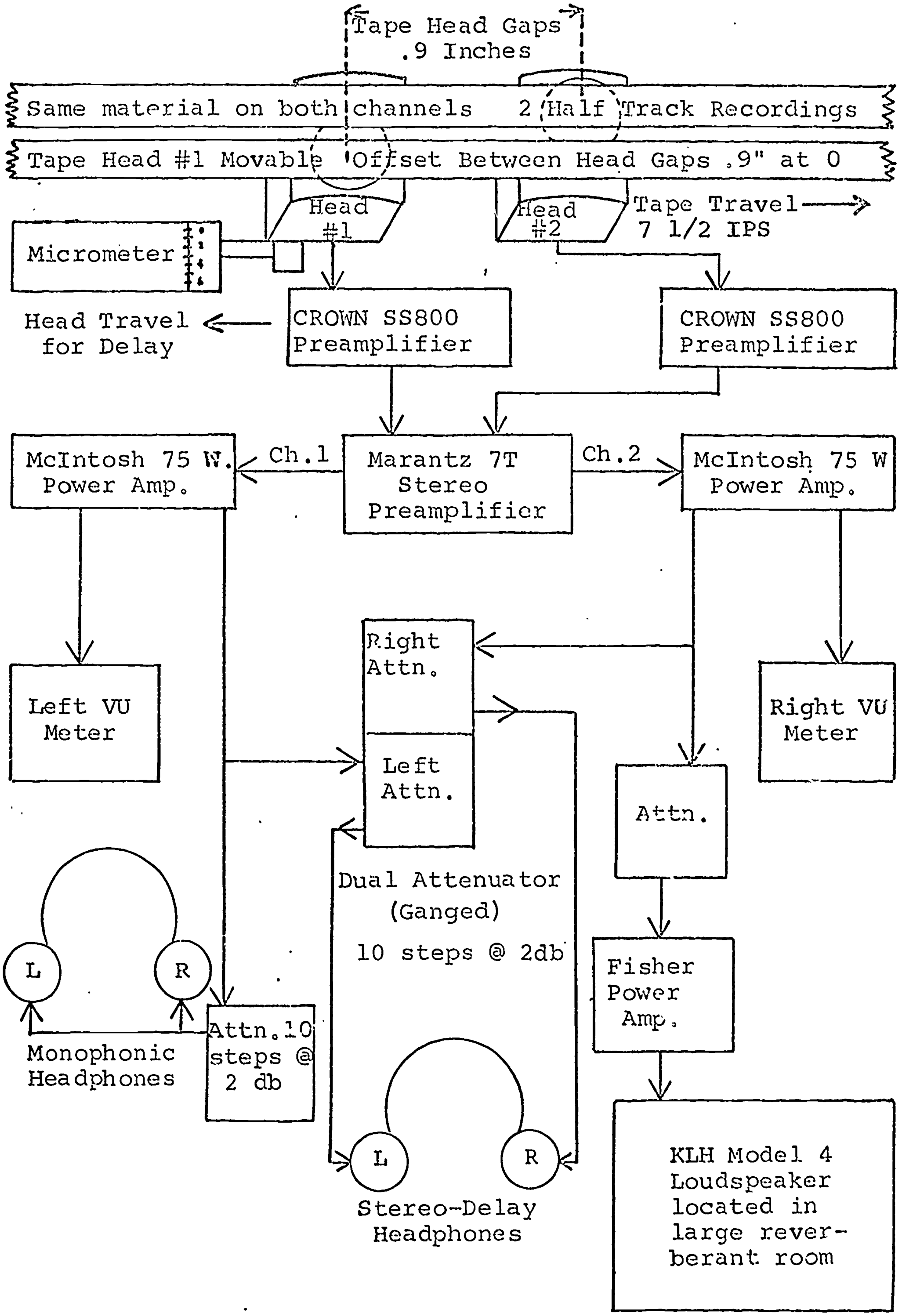
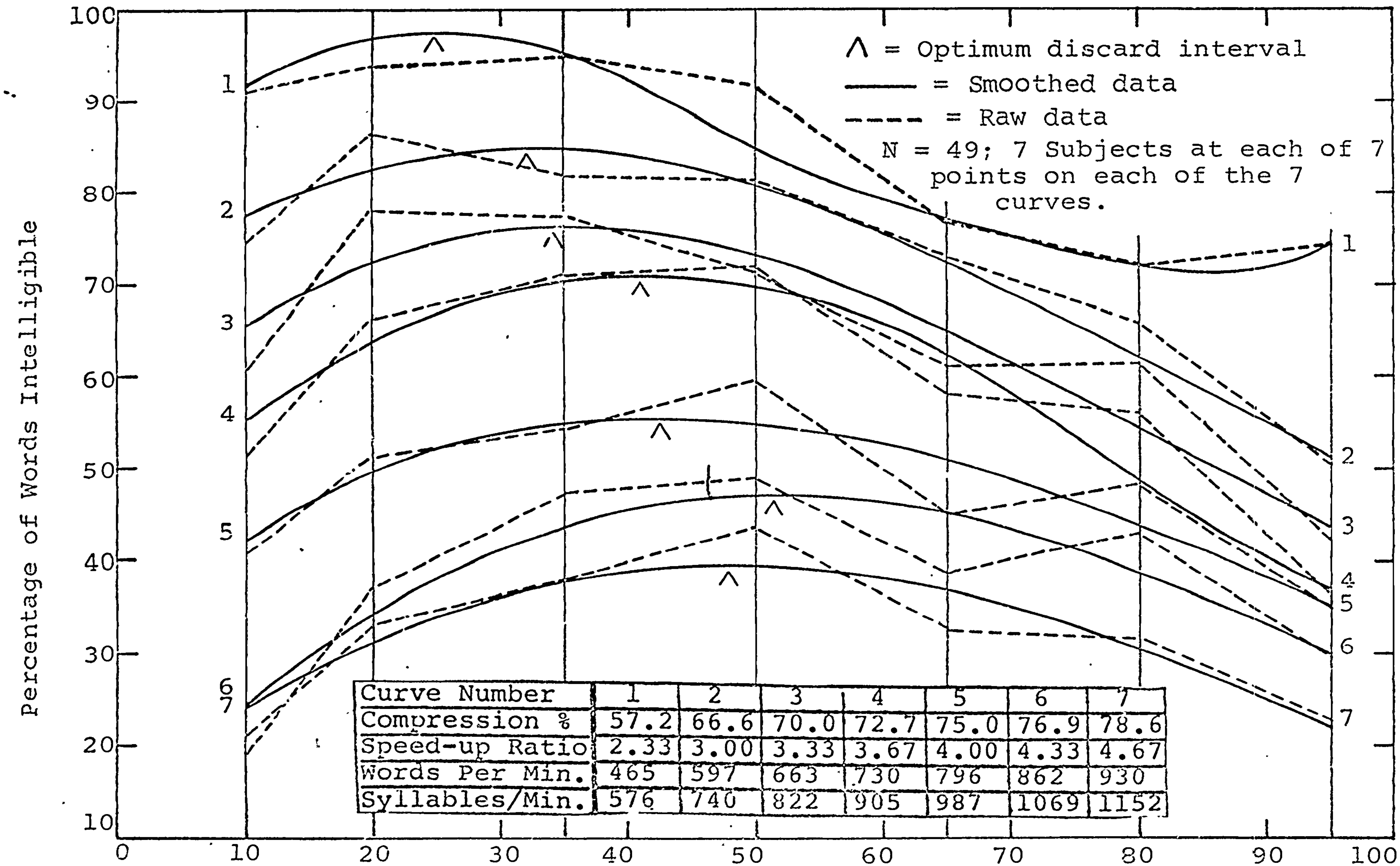


Figure 6 . Block Diagram of Audio Playback Apparatus for Main Study



Curve Number	1	2	3	4	5	6	7
Compression %	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Ratio	2.33	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Min.	465	597	663	730	796	862	930
Syllables/Min.	576	740	822	905	987	1069	1152

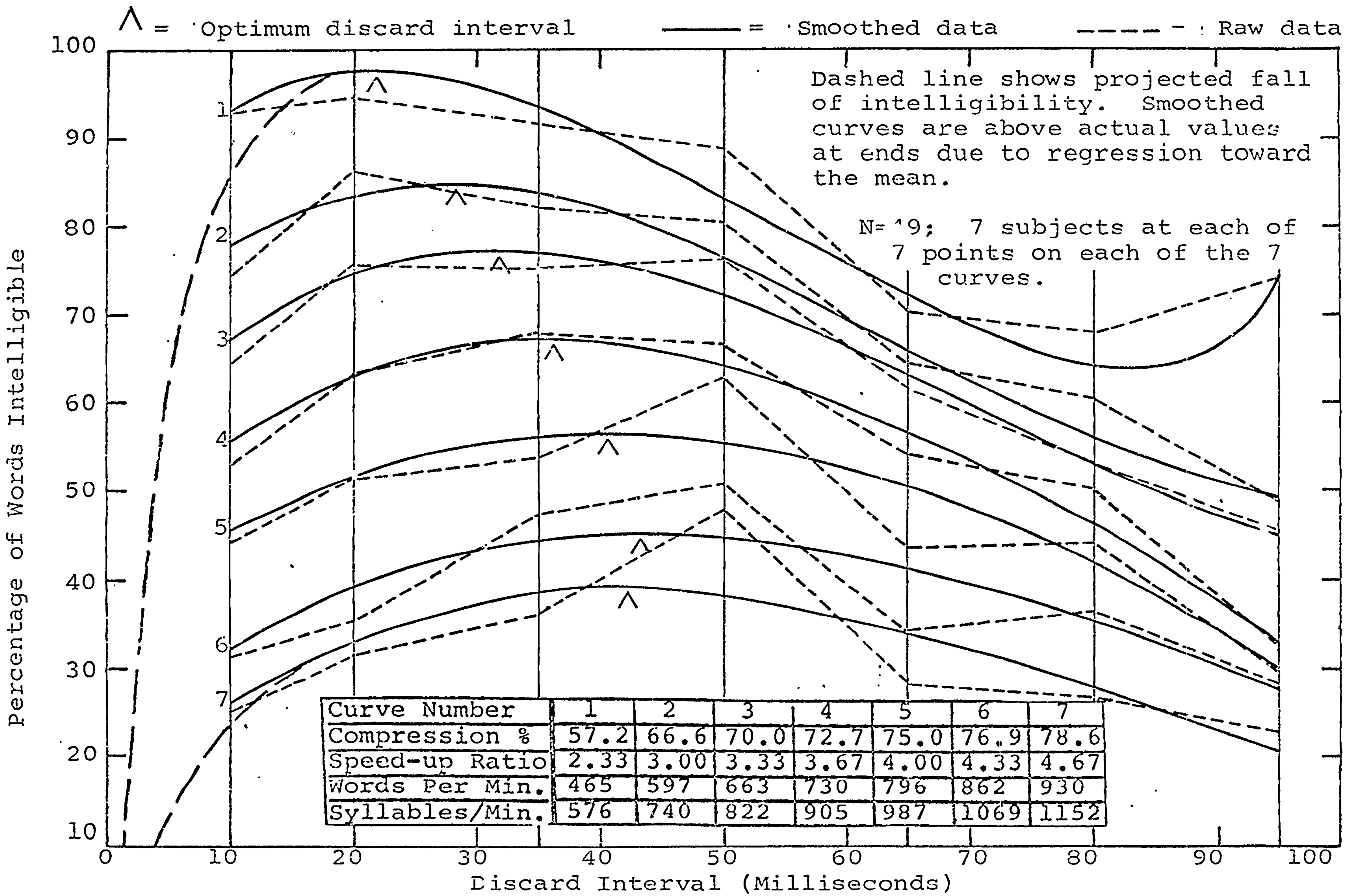
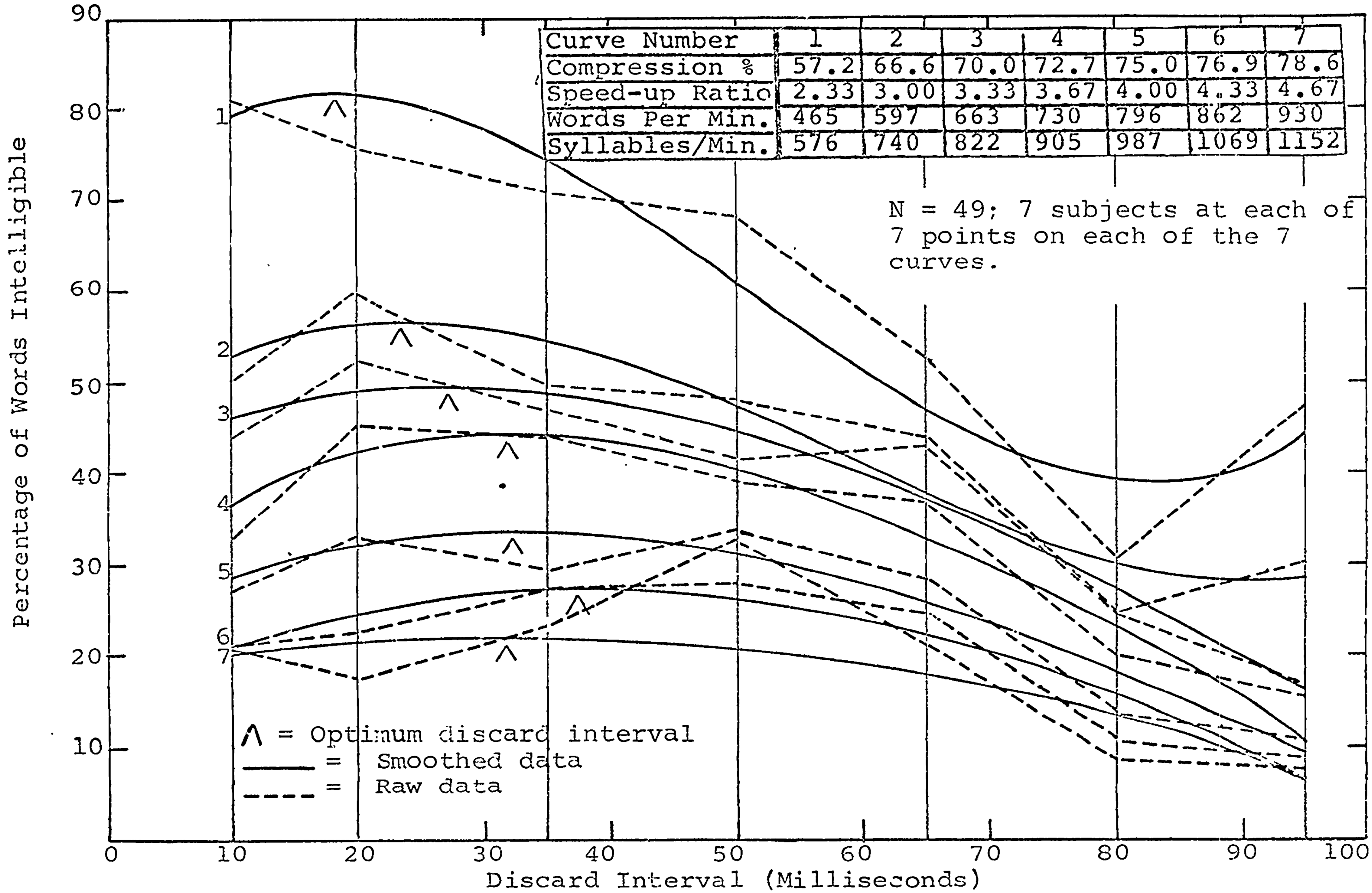


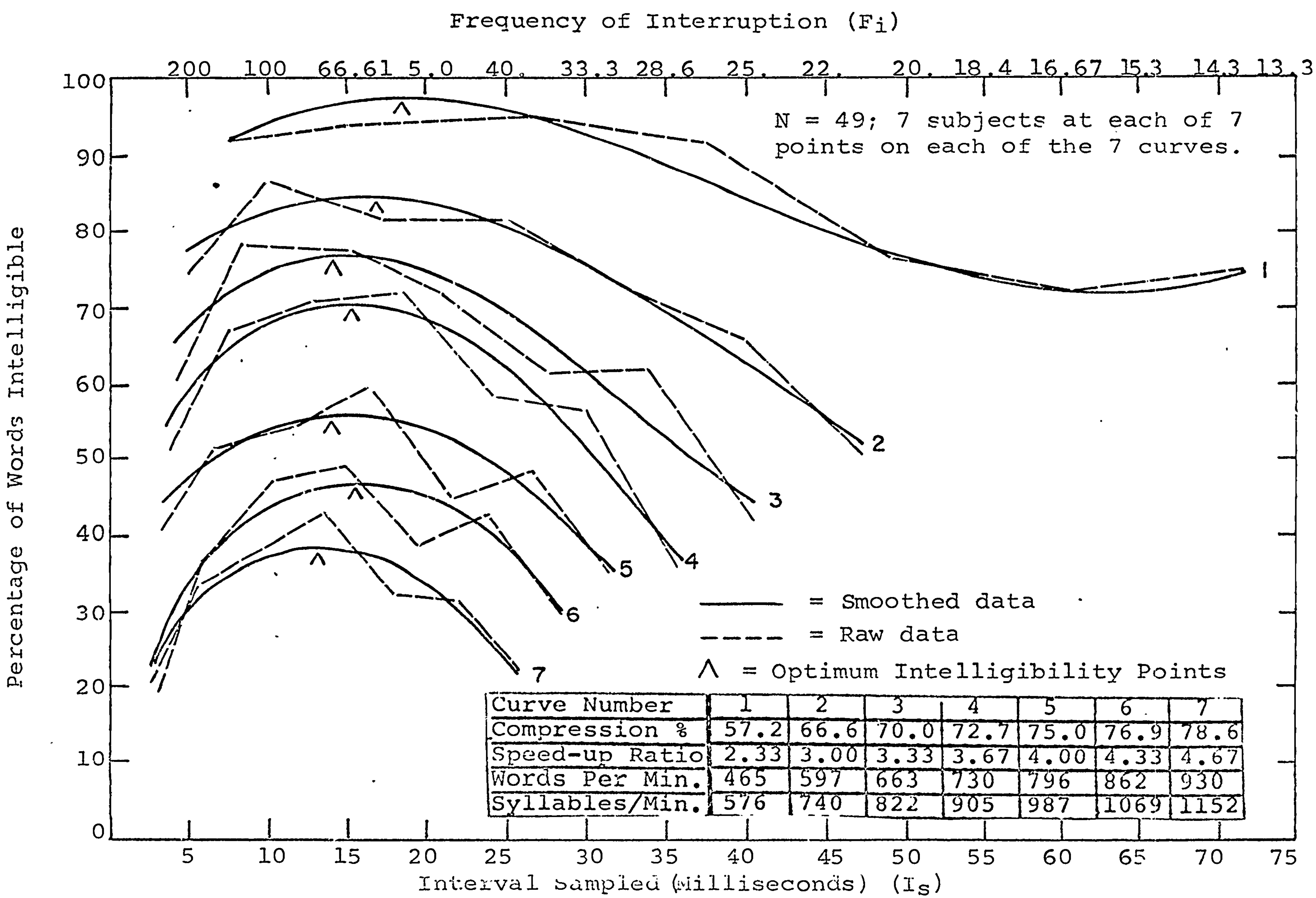
Figure 8 Intelligibility vs. Discard Interval for Monophonic (M) Presentation

Curve Number	1	2	3	4	5	6	7
Compression %	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Ratio	2.33	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Min.	465	597	663	730	796	862	930
Syllables/Min.	576	740	822	905	987	1069	1152



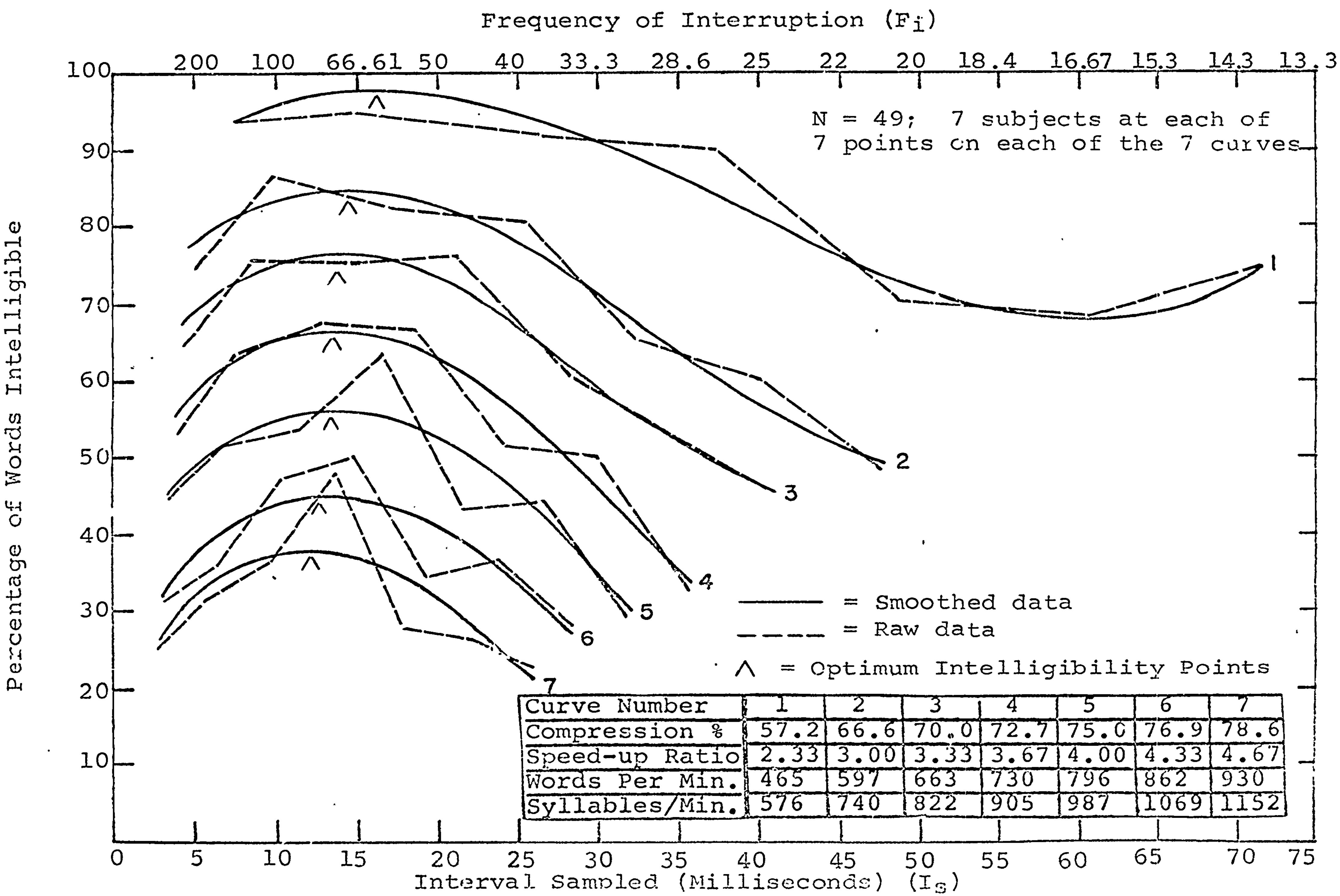
Curve Number	1	2	3	4	5	6	7
Compression %	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Ratio	2.33	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Min.	465	597	663	730	796	862	930
Syllables/Min.	576	740	822	905	987	1069	1152

Figure 9 . Intelligibility vs. Discard Interval for Loudspeaker (L) Presentation



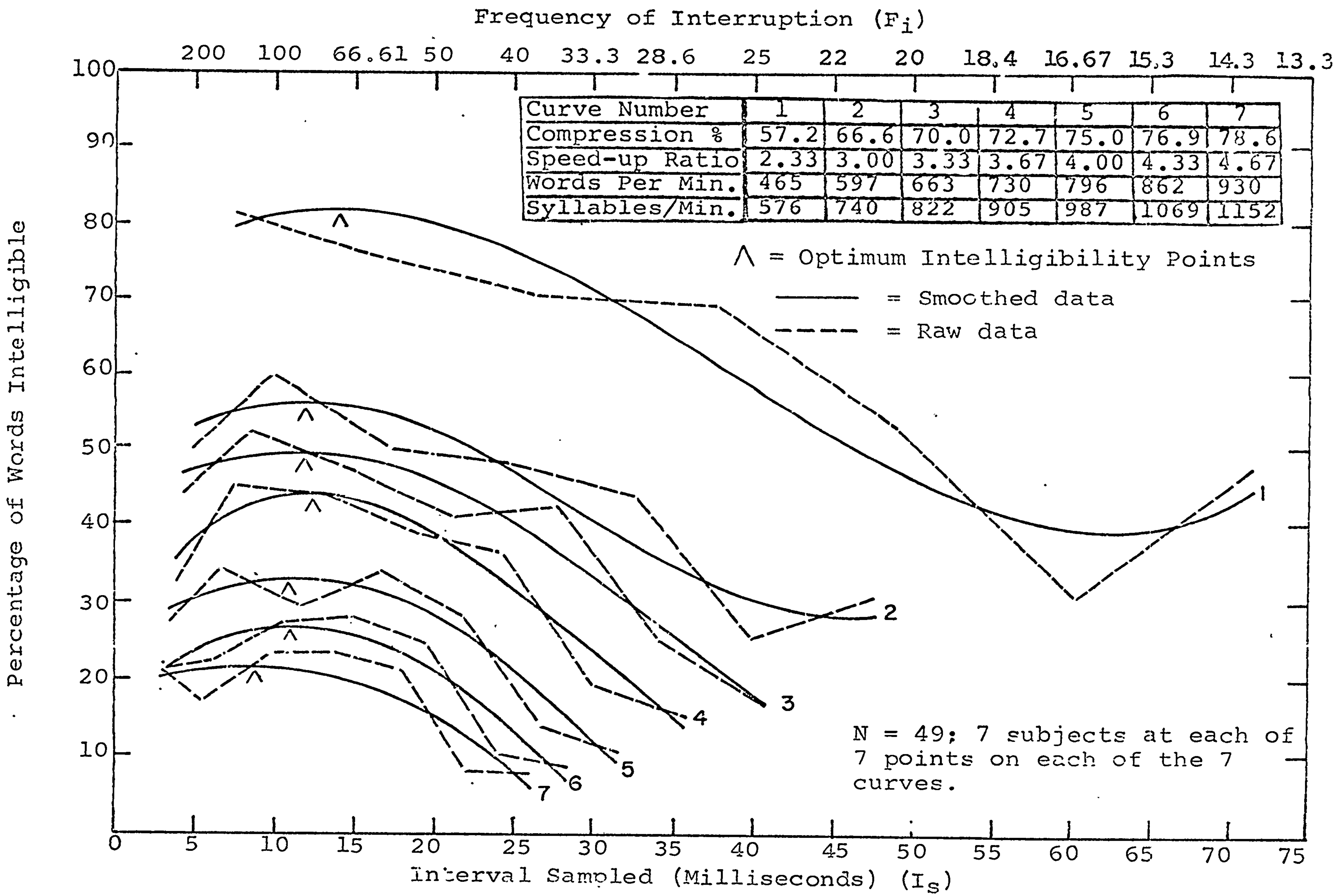
Curve Number	1	2	3	4	5	6	7
Compression %	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Ratio	2.33	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Min.	465	597	663	730	796	862	930
Syllables/Min.	576	740	822	905	987	1069	1152

Figure 10. Intelligibility of Speech Presented with an Interaural Delay of 10 ms. Stereophonically (S) by Means of Earphones at Each of 7 Compression Ratios and at 7 Different Sampling Intervals.



Curve Number	1	2	3	4	5	6	7
Compression %	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Ratio	2.33	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Min.	465	597	663	730	796	862	930
Syllables/Min.	576	740	822	905	987	1069	1152

Figure 11. Intelligibility of Speech Presented Monophonically (M) by Means of Earphones at Each of 7 Compression Ratios, and at 7 Different Sampling Intervals.



Curve Number	1	2	3	4	5	6	7
Compression %	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Ratio	2.33	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Min.	465	597	663	730	796	862	930
Syllables/Min.	576	740	822	905	987	1069	1152

Figure 12 Intelligibility of Speech Presented by Loudspeaker (L) in a Large Reverbrant Classroom, at Each of 7 Compression Ratios and at 7 Different Discard Intervals.

redundancy of phonemes. He was interested in seeing how much could be discarded and still have intelligible speech, and therefore he plotted his curves as a function of the discard interval. From a perceptual point of view, however, it should be obvious that the listener has no knowledge of what has been deleted, and can only perceive what is present. When the optimum sampling interval at each compression ratio is examined, it can be observed that approximately 15 ms. is the optimum sampling interval for every compression ratio This, incidentally, is what is required to insure that, on the average, there is at least one complete pitch period of the voice in each sample of speech.

In veiw of all the work being done with The Harmonic Compressor (Golden, 1966) and the Haskins Pattern play back equipment (Cooper et al., 1951 & 1952, Delattre et al., 1956, Liberman et al., 1957)where pitch periods are not considered, it seemed somewhat puzzling that the pitch period appeared so important. At first, it seemed that this might be a coincidence. However, the delay to one ear did seem related to pitch periods, and it is fairly easy to understand the rationale for this.

During the past summer, some very short sampling intervals less than a pitch period, were derived by means of the Adage Ambilog 200-Hybrid Analog-Digital computer, (Grandine and Hagan, 1965). With Fairbanks' compressor, it is not possible to get samples shorter than a pitch period except at very short discard intervals of 10 ms. and at rates above 50% compression. For such conditions, Fairbanks & Kodman (1957) complained that the frequency of interruption obtruded on the first formant of the speech frequencies From our work, it appears that samples of a half a pitch period will simply double the pitch of the voice so processed, just as one can do by playing a recording of a voice at twice normal speed. All these findings tend to indicate that speech is processed mentally over theperiod of a pitch period, and the pitch that is perceived is determined by the length of time between fundamental cycles irrespective of half periods added to a fundamental. A soprano voice will have almost twice the number of pitch periods in a given time as an average male. This has led telephone engineers to comment that women convey half as much information in a given time as men, because their voices are twice as redundant. However, short wave radio stations capitalize on this, because a woman's voice has twice as many chances (twice as many pitch periods) to get through static intact as a man's voice. In view of this phenomenon it would seem that a woman's voice might be expected to be more intelligible when compressed at high ratios, because twice as many pitch periods can be discarded, leaving an equal number to be perceived as with the average male voice. Comparisons made by Foulke, et al., (1962) and others have found that the speaker you have heard on our tape was the most comprehensible, when compressed, of a number of speakers including both men and women. Our research suggests at least two reasons for this.

When the pitch period record of the voice of our speaker was examined it was found that in twenty-five vowels sampled, his fundamental frequency never went below 95 Hertz* (cycles per second) nor above 105 Hertz. This is ± 5% variation. A similar study of Vice-President Humphrey's voice, however, gives a range from a low of 112 Hertz to a high of 187 Hertz, ± 25%. His voice does not compress well.

The women who read for the American Printing House for the Blind are quite dramatic in their voicing, and consequently have a very large dynamic pitch range.

*Hertz is the new designation for cycles per second proposed and used by the American Institute of Physics.

These are the readers used by Foulke et al., (1962) and others. For the rotating head type of compression, using a periodic discard, it would seem that a woman with a high-pitched voice but limited dynamic range would give superior results.

Earlier it was mentioned that this research has suggested two reasons why previous research has found a low-pitched man's voice superior to any of several women's voices. The first reason was that no known study has included the voice of a woman with a narrow dynamic pitch range. The second reason is that the current system for compressing speech, namely the rotating head assembly of the Tempo Regulator, has a fixed discard interval of 40 ms., regardless of the compression ratio. This means that at 50% compression, or 2 times the normal rate the sampling interval is likewise 40 ms., almost 3 times the optimum for a low-pitched man's voice, but seven to nine times the optimum for a high-pitched woman's voice. What is needed is a compressor that has the features of automatically adjusting the sampling time length to the optimum sampling length of the voice being processed. Rather than a strictly periodic device, what seems needed is a pitch-synchronous model such as those that recently patented by Denis Gabor (1950) (1965). His speech compressor has 8 different sampling intervals within an octave, with automatic switching to match the optimum sample length. With such equipment it should be possible to achieve improved intelligibility at higher compression ratios than is now possible, and consequently, a higher percentage of comprehension. Figure 13 shows the sampling interval of the tempo regulator at each speed-up ratio from 10% to 50% compression ratio.

The data were plotted from the optimum points, indicated by arrow heads, in Figures 11 and 12 as intelligibility versus word-per-minute rate in Figure 14.

Superimposing Figure 2 from the Pilot Study on Figure 14 of the main study renders Figure 15. The differences in the intelligibility of the two studies reflect the lower intelligibility scores when the P.A.L. is scored by a word-for-word basis rather than by correct answers.

In concluding it should be made clear that the intelligibility at the high word rates reported in this research does not by any means infer that comprehension of long connected continuous discourse can also be obtained at comparable rates. Continuous passages have also been processed at the optimum rates for maximum intelligibility, and at about 2.5 times normal (70% compression) they become meaningless, even though some words are intelligible. This is probably due in part to lack of time for cognitive processing. It is also due to a "double-take" effect; that is, one word is transformed into another, for example, the word "thriving" is heard as "fighting" at 2.5 times normal speed. This appears due to the /th/ being confused for an /f/ when the /r/ in thriving is clipped short, and the /v/, when clipped short, sounding like a /t/. In a long continuous passage, such a confusion can interfere with the comprehension of subsequent sentences presented at a rate two to three times normal. It would seem that such a confusion is analogous to the confusion resulting from losing one's place while watching a rapidly paced reading film.

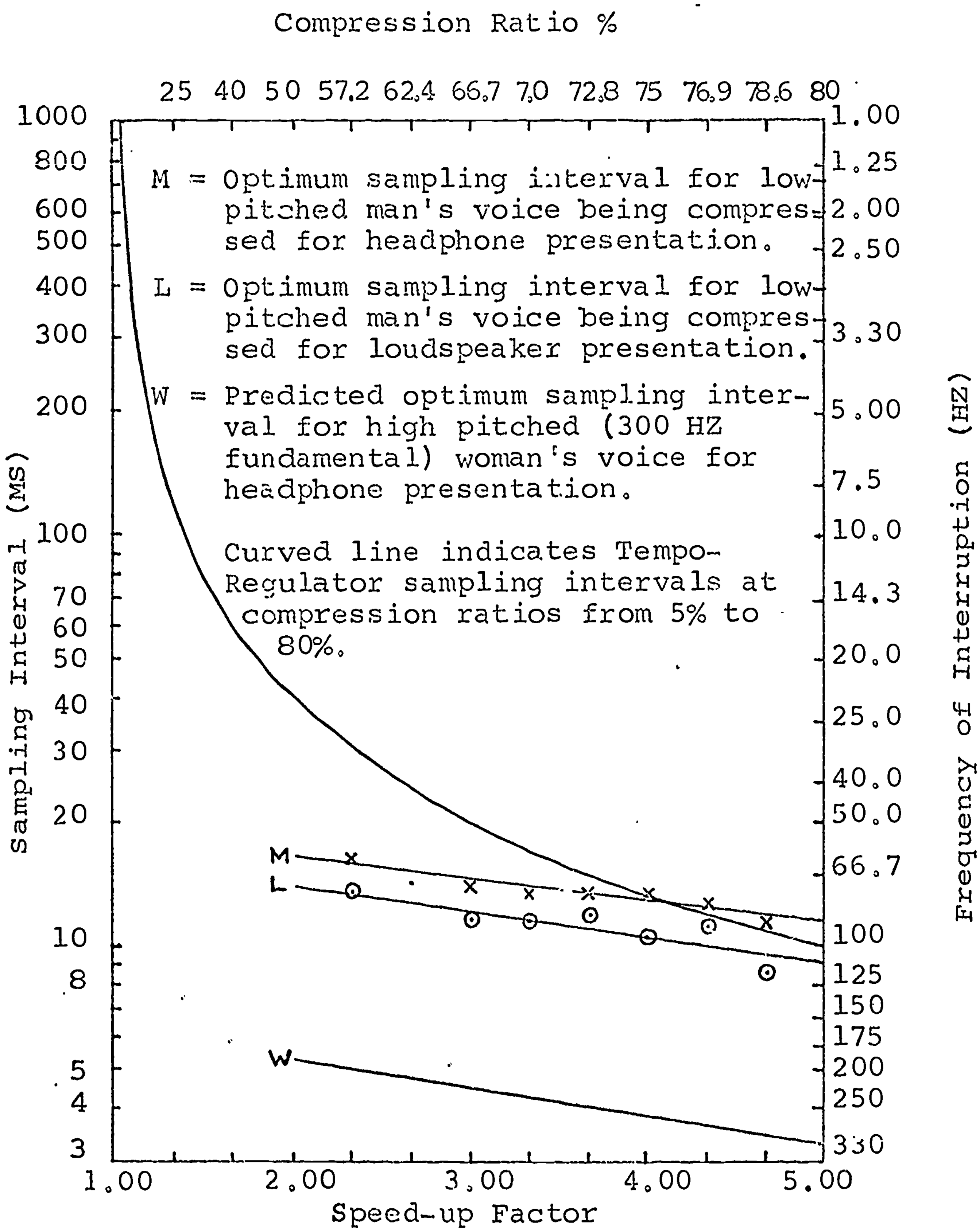
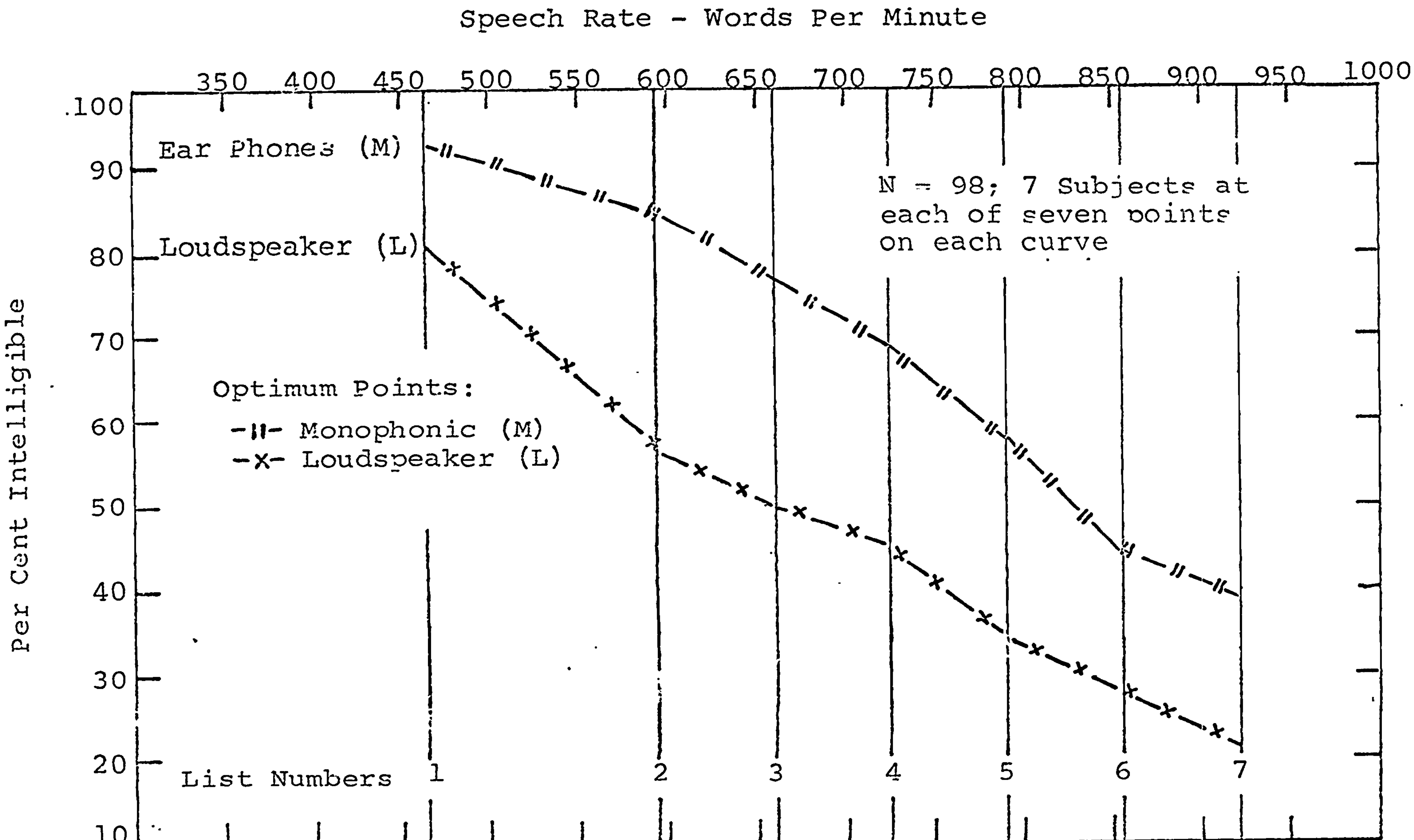
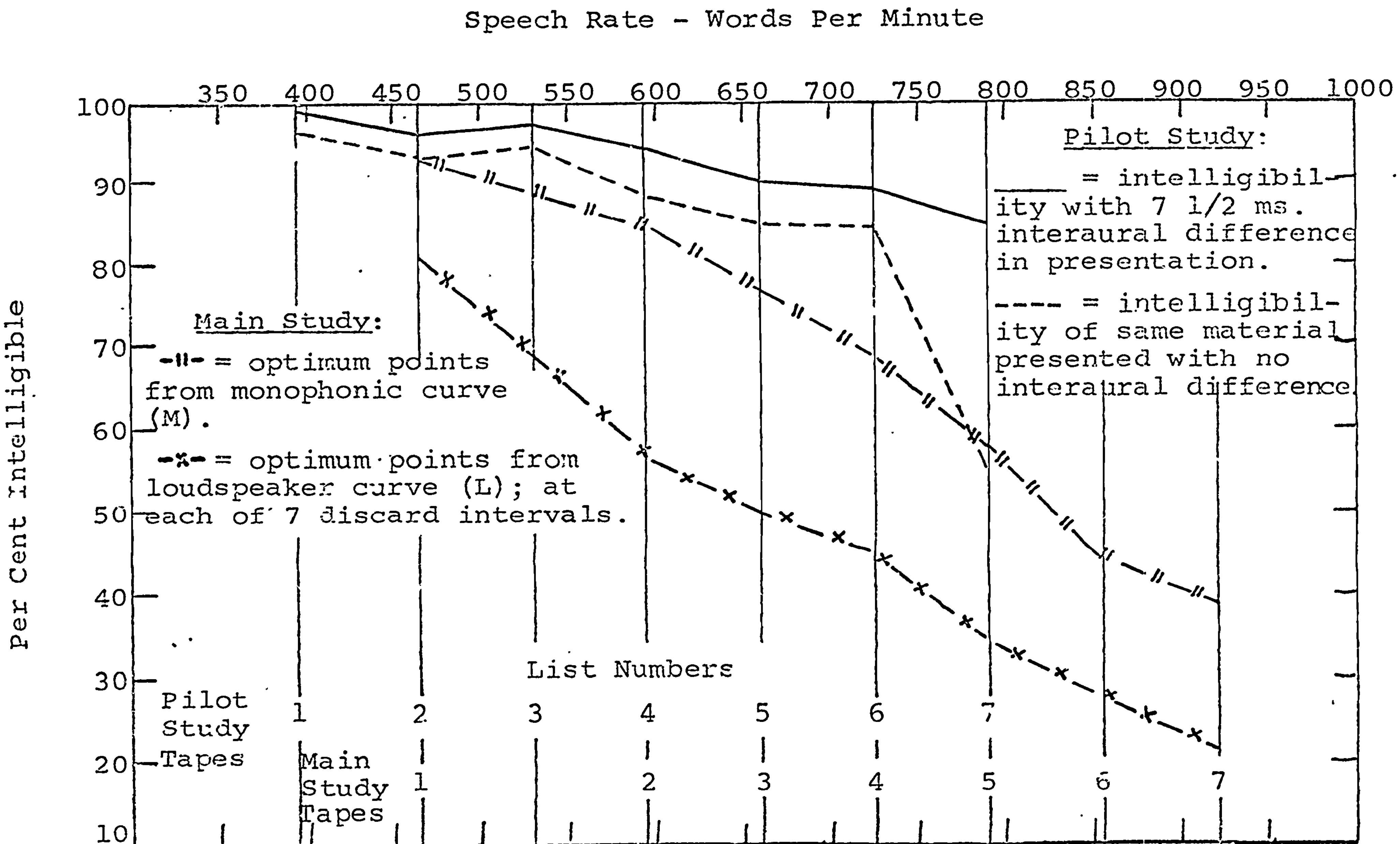


Figure 13. Tempo Regulator Parameters Compared with Optimum Sampling Intervals for a Low-Pitched Man's Voice and a Predicted Optimum Sampling Interval for a High-Pitched Woman's Voice



Compression Ratios	57.2	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Factor	2.33	3.0	3.33	3.67	4.0	4.33	4.67
Words Per Minute	465	597	663	730	796	862	930
Syllables Per Min.	576	740	822	905	987	1069	1152

Figure 14. Main Study Intelligibility of Time-Compressed Speech Versus Speech Rate for Both Monophonic Earphone and Loudspeaker Presentation.



Compression Ratios	50.0	57.2	62.4	66.6	70.0	72.7	75.0	76.9	78.6
Speed-up Factor	2.00	2.33	2.67	3.00	3.33	3.67	4.00	4.33	4.67
Words Per Minute	298	465	531	597	663	730	796	862	930
Syllables Per Min.	499	576	658	740	822	905	987	1069	1152

Figure 15. Intelligibility Versus Plotted Speech Rate from Pilot Study and Main Study for Monophonic Earphone with No Delay, Earphone with 7 1/2 MS. Inter-aural Difference in Time of Presentation, and Loudspeaker Presentation in a Reverberant Classroom.

REFERENCES

Asher, J. W. Intelligibility tests: a review of their standardization, some experiments, and a new test. Speech Monographs, 1958, 25 (1), 14-28.

Bixler, R.H., Foulke, E., Amster, C.H., & Nolan, C.Y. Comprehension in method of rapid speech: part 1. Louisville: University of Louisville, 1961.

Carroll, J. B. Language and thought. Englewood Cliffs, N. J.: Prentice-Hall, 1964.

Carroll, J. B. Problems of measuring speech rate. In Foulke, E. (ed), Proceedings of the Louisville conference on time compressed speech, Oct. 19-21, 1966. Louisville: Center for Rate Controlled Recordings, University of Louisville, 1967. Pp. 88-94.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. Journal of the Acoustical Society of America, 1952, 24, 597-606.

Cooper, F. S., Liberman, A. M., & Borst, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proceedings of the National Academy of Sciences, 1951, 37 (5), 318-325.

CRCR Newsletter. (News bulletin of the Center for Rate Controlled Recordings) 1, (1). American Institutes for Research, 8555 Sixteenth St., Silver Spring, Md. 20910. (Center for Rate Controlled Recordings located at University of Louisville, Louisville, Ky. 40208.)

Cramer, H. L. Intelligibility of compressed speech as a function of degree and direction of delay of presentation from one ear to the other. Paper presented at American Psychological Association, Chicago, Sept., 1965. Div. 15 Symposium: Recent Research on the Comprehension of Time-Compressed (Speeded) Speech.

Cramer, H. L. Braided-speech technique for time-compressing speech. Paper presented at the annual meeting of the Institute of Electrical and Electronic Engineers, New York, March 20, 1967. (a)

Cramer, H. L. The intelligibility of time-compressed speech. In Foulke, E. (ed.), Proceedings of the Louisville conference on time-compressed speech, Oct. 19-21, 1967. Louisville: Center for Rate Controlled Recordings, University of Louisville, 1967. Pp. 126-148. (b)

Cramer, H.L. The intelligibility of time-compressed speech. Unpublished doctoral dissertation, Graduate School of Education, Harvard University, 1968.

Davis, H. & Silverman, C.R. (eds.) Hearing and deafness. (Rev. ed.) New York: Holt, Rinehart and Winston, 1960.

Delattre, P., Cooper, F. S., Liberman, A. M., & Gerstman, L. J. Speech synthesis as a research technique. Proceedings of the VIIth International Congress of Linguistics (1952). London: Titus Wilson, 1956. Pp. 545-561.

Fairbanks, G., Everitt, W. L. & Jaeger, R. P. Method for time or frequency compression-expansion of speech. Convention Record of the Institute of Radio Engineers National Convention, part 8, Information Theory, 1953, 120-124.

Fairbanks, G., Everitt, W. L., & Jaeger, R. P. Method for time or frequency compression-expansion of speech. Transactions of the Institute of Radio Engineers Profession Group on Audio, Jan. and Feb. 1954, 7-12.

Fairbanks, G., Everitt, W. L., & Jaeger, R. P. Recording devices. Patent No. 2,886,650. Washington, D. C.: U. S. Patent Office, May 12, 1959. (Filed Oct. 22, 1952).

Fairbanks, G., Guttman, N., & Miron, M. S. Auditory comprehension in relation to listening rate and selective verbal redundancy. Journal of Speech and Hearing Disorders, 1957, 22, 23-32.

Fairbanks, G., & Kodman, F., Jr. Word intelligibility as a function of time compression. Journal of the Acoustical Society of America, 1957, 29, 636-641.

Foulke, E. (ed.) Proceedings of the Louisville conference on time compressed speech, Oct. 19-21, 1966. Louisville: Center for Rate Controlled Recordings, University of Louisville, 1967.

Foulke, E., Amster, C. H., Nolan, C. Y., & Bixler, R. H. The comprehension of rapid speech by the blind. Exceptional Children, 1962, 29, 134-141.

Foulke, E., & Bixler, R. H. Speech compression. The New Beacon, 1963, 47, 241-242. (a)

Foulke, E. & Bixler, R. H. The comprehension of rapid speech by the blind III. Appendix A proj. #2430. University of Louisville, Dec., 1963. (b).

Friedman, H.L., Orr, D.B., Freedle, R.O. & Norris, Cynthia M. Further research on speeded speech as an educational medium. Washington: American Institutes for Research, Jan. 1966.

Friedman, H.L., Orr, D.B., & Norris, Cynthia M. Further research on speeded speech as an educational medium: the use of listening aids. Washington: American Institutes for Research, July, 1966.

Gabor, D. Improvements in and relating to the transmission of speech. British Patent No. 645-542, London: The Patent Office, Nov. 1, 1950.

Gabor, D. Bandwidth compression and expansion system for transmission of speech. U. S. Patent No. 3,183,310. Washington, D. C.; U. S. Patent Office, May 11, 1965.

Garvey, W. D. The intelligibility of speeded speech. Journal of Experimental Psychology, 1953, 45, 102-108.

Garvey, W. D. Johns Hopkins Medical School, Baltimore, Md. Personal communication, 1965.

Garvey, W. D. & Henneman, R. H. Practical limits of speeded speech. Air Force Technical Report No. 5917, May, 1950.

Golden, R.M. Time compression and expansion of recorded speech. Paper presented at the meeting of the Acoustical Society of America, Los Angeles, Nov. 1966.

Goldman-Eisler, Frieda. Speech production and the predictability of words in context. Quarterly Journal of Experimental Psychology, 1958, 10, part 2, 96-106.

Goldman-Eisler, Frieda. Continuity of speech utterance, its determinants, and its significance. Language and Speech, 1961, 4, 220-231. (a)

Grandine, J. D. & Hagan, T. G. A parallel/sequential, stored-program hybrid signal processor. Simulation, Jan., 1965.

Hudgins, C. V., Hawkins, J. E., Karlin, J. E., & Stevens, S. S. The development of recorded auditory tests for measuring hearing loss for speech. The Laryngoscope, 1947, 57, 57-59.

Kimura, Doreen. Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 1961, 15 (3), 166-171. (a)

Kimura, Doreen. Some effects of temporal-lobe damage on auditory perception. Canadian Journal of Psychology, 1961, 15, (3), 156-165. (b)

Kimura, Doreen. A note on cerebral dominance in hearing. Acta Oto-laryngology, 1963, 56, 617-618. (a)

Kimura, Doreen. Speech lateralization in young children as determined by an auditory test. Journal of Comparative and Physiological Psychology, 1963, 56, (5), 899-902. (b)

Kimura, Doreen. Left-right differences in the perception of melodies. Quarterly Journal of Experimental Psychology, 1964, 16, Part IV, 355-358.

Liberman, A.M. Some results of research on speech perception. Journal of the Acoustical Society of America, 1957, 29, 117-123.

Miller, G. A. Decision units in the perception of speech. Transactions of the Institute of Radio Engineers Professional Group on Information Theory, 1962, IT-8, 81-83.

Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. Journal of Experimental Psychology, 1951, 41, 329-335.

Orr, D. B. & Friedman, H. L. Research on speeded speech as an educational medium. Washington, D. C.: American Institutes for Research, 1964. (a)

Orr, D. B., & Friedman, H. L. Research on speeded speech as an educational medium - project report. Grant No. 7-48-7670-203, Office of Health, Education and Welfare, 1964. Washington, D. C.: 1964. (b)

Orr, D. B., & Friedman, H. L. Research on speeded speech as an educational medium - proposed extension of subject grant. Washington, D.C.: American Institutes for Research, 1964. (c)

Pollack, Irwin. Message probability and message reception. Journal of the Acoustical Society of America, 1964, 36, 937-945.

Riedinger, Patricia A., Senior Physicist, Melpar Inc., Falls Church, Va., Personal communication, 1966.

Riedinger, Patricia. Internal technical note on delayed listening test. In Cramer, H.L. The intelligibility of time-compressed speech. Unpublished doctoral dissertation, Graduate School of Education, Harvard University, 1968. Pp. 281-293.

Savin, H. B. Word-frequency effect and errors in the perception of speech. Journal of the Acoustical Society of America, 1963, 35, 200-206.

Voor, J. B. The effect of practice upon the comprehension scores of time compressed speech. Unpublished master's thesis, University of Louisville, 1962.

Voor, J. B. & Miller, J. M. The effect of practice upon the comprehension of time-compressed speech. Speech Monographs, 1965, 32, 452-454.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Winer, B. J. Personal communication, 1965.

Wood, C. D. Comprehension of compressed speech by elementary school children. Final Report, Title VII number 1300. Grant No. 7-24-0210-262. Washington, D.C., U. S. Dept. of Health, Education & Welfare, 1965. (Work done at A-V Ctr., Indiana University, Bloomington, Indiana.)