Washington University School of Medicine

Digital Commons@Becker

Independent Studies and Capstones

Program in Audiology and Communication
Sciences

1947

The reliability and validity of monitored connected discourse as a test of the threshold of intelligibility

George Alexander Falconer

Follow this and additional works at: https://digitalcommons.wustl.edu/pacs_capstones

Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Falconer, George Alexander, "The reliability and validity of monitored connected discourse as a test of the threshold of intelligibility" (1947). *Independent Studies and Capstones*. Paper 363. Program in Audiology and Communication Sciences, Washington University School of Medicine.
https://digitalcommons.wustl.edu/pacs_capstones/363

This Thesis is brought to you for free and open access by the Program in Audiology and Communication Sciences at Digital Commons@Becker. It has been accepted for inclusion in Independent Studies and Capstones by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.

53
WASHINGTON UNIVERSITY

Central Institute for the Deaf

THE RELIABILITY AND VALIDITY OF
MONITORED CONNECTED DISCOURSE
AS A TEST OF THE THRESHOLD OF
INTELLIGIBILITY

by

George Alexander Falconer

A dissertation presented to the Board
of Graduate Studies of Washington
University in partial fulfillment
of the requirements for the
degree of Master of Science
in Education

June, 1947

Saint Louis, Missouri

TABLE OF CONTENTS

CHAPTERS

	Page
I. Introduction.....	1
II. Electro-Acoustic Apparatus and Recorded Material.....	3
A. Description of Apparatus.....	3
B. Calibration of Earphones.....	5
C. Calibration of Loudspeaker.....	5
D. Description of Recorded Connected Discourse.....	6
E. Description of Harvard University Auditory Test #9.....	7
III. Preliminary Study.....	10
A. Subjects.....	10
B. Procedure.....	12
C. Data and Discussion.....	13
IV. Description of Major Study.....	21
A. Subjects.....	21
B. Procedure.....	22
V. Data and Discussion of Major Study.....	26
A. Group I.....	26
1. Range.....	26
2. Test-retest (Reliability).....	31
3. Relationship to Control Test (Validity).....	33
B. Group II.....	34
1. Range.....	35
2. Test-retest.....	35
3. Relationship.....	35
C. Group III.....	38
1. Range.....	38
2. Test-retest.....	40
3. Relationship.....	40

	Page
D. Comparison of Normal and Hard-of-Hearing Subjects.....	43
1. Range.....	43
2. Test-retest.....	46
3. Relationship.....	46
4. Face Validity of TICD Test.....	50
5. Time Required for TICD.....	50
6. Fatigue.....	51
VI. Conclusions.....	53
Bibliography.....	56

LIST OF TABLES

	Page
1. Audiometric Data for Group I.....	11
2. Threshold of Intelligibility for Group II.....	14
3. Means and Medians of the Settings of Group I for all the Sessions.....	15
4. Thresholds of Intelligibility for Group III....	20
5. Mean of the Means of the Sessions Compared with the Mean of the Medians. Group I.....	17
6. Group I. Individual Settings for Threshold of Intelligibility for Connected Discourse.....	28
7. Test-retest Results for Group I Showing Devia- tion of the TICD from the #9 Threshold (Right Ear).....	29
8. Test-retest Results for Group I Showing Devia- tion of the TICD from the #9 Threshold (Left Ear).....	32
9. Group II. Individual Settings for Threshold of Intelligibility for Connected Discourse.....	36
10. Test-retest Results for Group II Showing Devia- tion of the TICD from the #9 Threshold.....	37
11. Group III. Individual Settings for the Threshold of Intelligibility for Connected Discourse.....	39
12. Test-retest Results for Group III Showing Deviation of the TICD from the #9 Threshold (Unoperated Ears).....	41
13. Test-retest Results for Group III Showing Deviation of the TICD from the #9 Threshold (Operated Ears).....	42
14. Distribution of the Ranges of the 6 Settings of the Normal and Hard-of-Hearing Listeners.....	45

LIST OF FIGURES

	Page
1. Medians of the Medians of the Sessions of Group I.....	19
2. Block Diagram of Apparatus.....	20

CHAPTER ONE

INTRODUCTION

Current articulation tests to determine the threshold of intelligibility employ lists of either nonsense syllables, numbers, words, or sentences. The thresholds of such tests are usually defined as the lowest intensity level at which the listener hears correctly 50 per cent of the test material. The lists are very often long and, consequently, time consuming and fatiguing. Workers who have developed the various forms of articulation tests seem to have purposely avoided one very obvious method of determining the threshold of intelligibility, i. e., of using as test material a sample of actual everyday connected discourse and letting the listener determine for himself where his threshold of intelligibility is. This can be done by letting the listener vary the intensity of the speech to the point where he can just hear and understand what is being said.

We felt that by using this technique a quick measure of the threshold of intelligibility might be obtained from connected discourse. Such a test would simulate everyday listening conditions. It might be more interesting than lists of words and sentences, because the listener himself would be making the test. The technique might also eliminate some of the emotional factors involved when an anxious

listener begins missing some of the test items. The listener, varying the intensity himself, would not become concerned about not hearing because he would be able to restore intelligibility immediately by adjusting his attenuator. It was the observation of this emotional factor that brought up the question of the ability of the individual to select his own threshold of intelligibility for connected discourse. Of course, such a test speech sample must be "monitored," that is, the intensity level of the speech must be kept constant. This would eliminate the chance of error due to a change of the level of speech rather than an error of the listener's judgment.

It is not our intent to advocate that connected discourse can replace refined articulation tests where exact discriminations are desired. We are here primarily interested in learning whether the normally hearing or hard-of-hearing individual can reliably select a threshold of "intelligibility" and, if so, whether his judgment is a valid measurement of that threshold.

CHAPTER II

ELECTRO-ACOUSTIC APPARATUS AND RECORDED TEST MATERIAL

Description of Apparatus¹

The apparatus was capable of reproducing known sound pressure level under an earphone from 0 db to approximately 145 db r.m.s. (re 0.0002 dynes/cm²) at 1000 cps. For the field measurements an appropriately matched WE 753-B loud speaker was substituted for the earphone and was capable of delivering up to 115 db sound pressure level (re 0.0002 dynes/cm²) at 1000 cps six feet on the axis. Figure 1 is a block diagram of the complete testing system. The laboratory in which the equipment was housed consists of a large room containing a sound treated room and an adjacent control room. The testing room has an absorption coefficient of .83 as calculated by the Sabine formula. The ambient noise is 35 db sound pressure level as measured on the WE R-330 sound level meter with the selector set to flat. The meter had been checked for accuracy by the Bureau of Standards.

The subject was seated comfortably in a chair in the testing room and listened to the speech through a single PIR-10 earphone mounted on a double headband. A 6B cushion and dummy headphone covered the ear which was not be-

¹ O.S.R.D. Report No. 6303 "Tolerance for Pure Tones and Speech in Normal and Hard-of-Hearing Ears," July, 1946, pp. 8-13.

ing tested. The PCH-10 earphone was fed through an appropriately matched and calibrated 110 db Hewlett-Packard attenuator. Of course, the earphones were removed for field testing and the subject faced the test loud speaker. To insure relatively uniform position of the head with respect to the sound source, an adjustable headrest was used.

The two attenuators used by the subjects were Daven PT-320-P, 10 watt, at 2 db per detent, and were housed in a 7" by 5" by 4" wooden box. One attenuator was inserted between the output transformer of the power amplifier and the matching transformer of the speaker. The other attenuator was inserted between the output transformer of the power amplifier and the matching transformer of the earphone.

A talkback microphone was suspended from the ceiling of the testing room. An instruction microphone was available to the operator for communication with the subject through the earphone or the loud speaker. This arrangement assured convenient and accurate rapport with the subject.

The remainder of the apparatus was mounted in a cabinet convenient to a window between the dead room and the control room. A resistance-tuned audio-oscillator was used as the source for calibration by pure tones. The source

for the speech was an electrical transcription reproducer. The recorded material was made according to the engineering standards set by the National Association of Broadcasters for transcription recordings and reproduction.

Calibration of Earphones

The pressure response of the earphones was individually determined at the Electro-Acoustic Laboratory of Harvard University by methods now adopted as standard. The pressure response of the earphone used was flat ± 2 db up to 3000 cps, from 3000 to 8000 it deviated ± 5 db. The PDR-10 earphone was also capable of delivering instantaneous peaks of pressure, such as occur in speech, at levels of at least 155 db instantaneous sound pressure (re 0.0002 dynes/cm²) into the 6 cc coupler. The power required for such sound pressure was about 6 watts.

Calibration of Loud Speaker²

The chart below indicates the sound pressure levels produced by the loud speaker at a distance of six feet on the axis in the dead room. A fixed frequency of 1000 cps was maintained. The electrical power fed to the speaker was measured with a standard VU meter bridged across the 500 ohm primary of the loud speaker matching transformer as shown in the diagram. The sound pressure level was

²O.S.R.D. Report No. 3105, "Response Characteristics of Interphone Equipment IV," Jan., 1944.

measured with the WE-330 sound level meter. This method of calibration has been adopted by the American Standards Association.

Watts Approx.	VU (dbm) 0-VU .001 watt into 500 ohm load	SPL (re 0.0002 dynes/cm ²)
.000076	-11	60
.000237	- 6	65
.00076	- 1	70
.00237	+ 4	75
.0076	+ 9	80
.0237	+14	85
.0759	+19	90
.2372	+24	95
.759	+29	100
2.371	+34	105
7.590	+39	110
24.000	+44	115

Above 24 watts amplitude distortion begins to appear.
Maximum speaker capacity 25 watts.

Description of Recorded Connected Discourse

The speech sample of connected discourse was a newscast by Fulton Lewis, Jr., transcribed from the radio network lines. It was chosen because it contained the qualities of evenness, simplicity, and interest, and was conveniently obtained, having been used previously in a study of speech tolerance.³ The evenness of the speech had been varied by an experienced observer who adjusted an attenuator

³0.S.R.D. Report No. 6303. op. cit., p. 13.

of 1 db steps so as to keep the speech just intelligible to him. To insure reliability the test was repeated. The settings of both tests lay within a total range of 3 db. Similarly, the observer adjusted the attenuator to maintain a definite and consistent tickle in his ear on nearly all of the stressed syllables. All of the thirty-seven settings were within a total range of 3 db. The two criteria, evenness of peaks and evenness of the general level of speech, indicate that the record is suitable as a test for intelligibility.

A calibrating tone of 1000 cps was recorded on the disc. The average speech level, as measured by the VU meter, throughout the 4 minutes and 45 seconds playing time lay within 2 db of the level of calibrating tone.

Vinylite pressings were made by the Allied Record Company, Hollywood, California. Each record was played about 50 times.

Description of Harvard University Auditory Test No. 9

Each disc contains a list of 42 familiar spondaic words (two long syllables of equal stress, such as row-boat and cookbook) recorded by male voice at 6 second intervals. The words are divided into seven groups of 6 words each, and each group is recorded at progressively

lower levels of intensity 4 db apart. Thus the range on the disc is 24 db. A carrier phrase, "Number One," "Number Two," etc., precedes each word; and, at the lower levels of recording, the intensity of the carrier phrase is higher than that of the test word. On each disc is a 1000 cps calibrating tone.⁴

The recording scheme is presented in the following table:

Words	No.	Relative Level of Carrier Phrase	Relative Level of Test Words
1	- 6	0 db	0 db
7	- 12	0	-4
13	- 18	0	-8
19	- 24	-2	-12
25	- 30	-6	-16
31	- 36	-10	-20
37	- 42	-14	-24

⁴ Hudgins, C.V., J.E. Hawkins, J.E. Karlin, and S.S. Stevens, "The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech," Laryngoscope, 57:64-67, Jan., 1947.

Figure 1. Block Diagram of Apparatus

CHAPTER III

PRELIMINARY STUDY

A preliminary study was made to explore the characteristics of the responses of normal and hard-of-hearing individuals and thereby establish a general procedure of study which would most directly examine the several questions involved. We wished to determine the influence of (1) experience, (2) fatigue, and (3) familiarity with the test material. We also hoped to (4) develop an accurate description of the threshold sought so that each subject would use the same criterion. The study comprised three groups for reasons of convenience.

Subjects

Group 1 was made up of 7 individuals, 6 of whom heard speech at or near normal threshold when tested with Test #9. However, 2 of the 6 (C.S. and P.T.) revealed a loss of approximately 50 db at 4096 c.p.s. when tested on the audiometer. At 2048 c.p.s. both of them were within the normal range of hearing. For our purposes any speech threshold was considered normal if it was within 3 times the standard deviation of the norm. The seventh subject (A.R.) was hard of hearing, having a 40 db loss for speech. Audiometry showed an abrupt high frequency loss beginning at 512 c.p.s. Table 1 contains pertinent information concerning the above subjects.

TABLE 1 Audiometric Data for Group I.

Subject	Age	Sex	Air Conduction Audiogram	Speech Threshold (#9) (re: 0.0002 dynes/cm)	
				Free Field	Monaural
J.V.	20	F	R. * L. *	17 **	R. 20 *** L. 19
A.M.	39	F	R. * L. *	14	R. 18 L. 17
P.O.	33	F	R. * L. *	19	R. 19 L. 17
V.A.	39	F	R. 0 0 0 0 1 0 0 **** 26 L. 0 0 0 1 0 1 0	26	R. 26 L. 17
C.S.	28	M	R. 0 0-0-0-0 4 3 L. 0-0-0-0 3-0	19	R. 23 L. 21
P.T.	36	M	R. 1 0 1 0 0 3 0 L. 1 1 1 1 2 0	27	R. 30 L. 27
A.R.	21	F	R. 1 1 2 6 9 X X L. 1 1 1 5 6 9 7	65	R. 67 L. 65

* Heard all tones at 10 db on audiometer by sweep frequency test.

** Normal threshold 15 db, S.D. 3 db.

*** Normal threshold 22 db, S.D. 2.63 db.

(Normal thresholds and S.D.'s. subsequently determined on the control group of the main experiment)

**** Audiograms represented numerically. Each digit represents the hearing loss in bels. A dash below a digit indicates loss of an additional half bel. A dash before a digit represents hearing more acute than the zero level of the audiometer.⁵

⁵O.S.R.D. Report No. 6303, op. cit., p. 7.

Group II consisted of 20 hard-of-hearing subjects who were candidates for the fenestration operation and had come to the Hearing Clinic at Central Institute for a series of diagnostic tests.

Group III included 20 hard-of-hearing subjects who had undergone the fenestration operation and had come to us for an evaluation of the results. Since the natural function of one of the ears of each subject had been disturbed, it is proper to focus attention on the unoperated ear. Thus, for this group, we deal entirely with single ears.

Procedure

The subject sat in a comfortable chair in the testing room. His speech threshold was first obtained in a free field, using Test #9. Then the attenuation box was placed in his lap and verbal instructions were given for seeking the threshold of intelligibility for connected discourse. The subject adjusted the attenuator to his threshold and called off the dial reading to the observer. The setting was repeated twice more and the dial readings were recorded. (A more detailed discussion of the procedure is given later.) The method for the monaural tests was the same; except, of course, the subject listened through an earphone. The right ear was always tested first.

Group I sat for a series of 5 sessions one week apart. The first session included the tests described in the preceding paragraph; the succeeding 4 sessions included only

the experimental test with connected discourses. The subjects soon grasped the technique of the test; so it was not necessary, as a rule, to repeat the instructions from test to test or from session to session. The dial reading for each setting was recorded. The time required for each setting was measured and recorded in seconds. Time started at the moment the subject began hearing the speech and ended when he called off the dial reading.

Groups II and III came for only one session and were tested in the manner described above. The data was gathered over a period of three months and arranged in chronological order for study.

Data and Discussion

The preliminary study revealed several pertinent facts which opened the problems of the main study to direct attack. We found that a comparatively brief test does almost as well as a larger one; therefore, it was practical to give the test to a larger number of people, thus increasing the statistical reliability. An adequate set of instructions to the subject had been developed which increased both the validity and reliability of the test. Finally, repeated use of a single recording of monitored connected discourse proved to be satisfactory.

The only evidence that experience had any influence on the results of the test is demonstrated most clearly by Tables 2 and 3. The ranges for the tests in the left ear

Table 2 Threshold of Intelligibility for Group II
(Decibels re 0.0002 dynes/cm²)

Sub- ject	Settings									Range			#5 Test		
	PF			RE			LE			PF	R	LE	PF	RE	LE
1.	52	56	56	60	60	56	68	68	70	4	4	2	45	44	53
2.	56	72	75	68	91	83	86	86	89	19	6	3	65	70	70
3.	66	66	70	72	72	76	78	82	80	4	4	4	65	58	78
4.	70	68	70	72	66	72	80	80	78	2	6	2	72	78	85
5.	58	56	56	68	70	72	56	56	58	2	4	2	59	68	58
6.	66	59	59	62	54	60	65	59	71	7	8	6	53	53	61
7.	58	60	56	58	60	64	62	64	64	4	6	2	55	58	62
8.	74	70	78	84	88	86	92	92	94	8	4	2	70	80	87
9.	68	68	68	76	78	80	78	80	78	0	4	2	66	70	71
10.	58	64	60	70	72	68	60	62	62	6	4	2	56	62	63
11.	64	66	68	64	68	66	76	80	78	4	4	4	54	62	70
12.	64	66	68	70	72	72	86	88	98	4	2	2	58	56	87
13.	72	72	70	82	92	88	78	76	76	2	4	2	66	84	76
14.	66	66	62	84	88	84	72	76	72	4	4	4	70	85	70
15.	74	70	70	86	88	90	90	92	90	4	4	2	73	80	65
16.	68	76	70	72	76	74	72	76	72	8	4	4	63	66	65
17.	72	74	70	80	80	80	76	70	74	4	0	6	73	88	72
18.	70	64	68	74	82	84	80	84	82	6	10	4	68	78	78
19.	68	66	66	74	70	68	80	80	80	2	6	0	65	70	78
20.	64	72	70	80	82	78	80	86	86	8	4	6	68	79	78
Averages	65.4	66.5	66.5	74.1	75.5	75.5	76.3	77.0	77.6	5.15	4.8	3.01	63.3	70.4	72.4

Table 3. Means and Medians of the Settings of Group I
for all the Sessions.

Sub- ject	Ses- sion	Free Field		Right Ear		Left Ear		Range		
		Mean	Med.	Mean	Med.	Mean	Med.	FF	RE	LE
J.V.	I	23.33	26	28.66	28	29.33	30	6	2	2
	II	22.00	22	27.33	28	27.33	28	4	2	6
	III	24.66	24	29.33	30	29.33	30	2	6	2
	IV	23.33	24	28.00	30	30.66	30	2	6	2
	V	24.00	24	27.33	28	27.33	26	4	6	2
A.M.	I	29.66	30	36.66	36	35.33	36	1	6	2
	II	25.33	26	31.33	30	27.33	28	6	6	2
	III	23.33	24	25.33	26	23.33	22	2	2	4
	IV	20.00	20	24.66	24	20.66	20	4	6	2
	V	18.00	18	24.66	22	26.00	26	0	6	8
F.O.	I	23.66	24	28.66	28	27.33	27	1	4	3
	II	25.33	24	34.00	34	27.33	28	4	4	2
	III	26.00	26	35.33	36	32.33	33	0	2	6
	IV	26.00	26	31.33	32	30.00	30	0	2	4
	V	28.66	28	34.66	34	32.00	32	2	2	0
V.A.	I	25.66	25	32.00	33	31.00	31	8	3	2
	II	29.33	28	31.33	32	26.00	26	4	2	4
	III	28.66	28	32.66	32	30.66	30	6	6	2
	IV	26.66	26	30.66	32	27.33	26	2	4	4
	V	25.33	26	31.33	32	31.33	32	2	2	2
C.S.	I	25.66	27	32.33	33	31.66	31	6	4	4
	II	23.33	24	22.66	22	22.66	22	2	2	2
	III	29.33	30	30.00	30	30.66	30	2	2	2
	IV	27.33	28	35.33	34	31.33	32	2	4	2
	V	24.00	22	32.66	30	29.33	30	6	8	2
P.T.	I	30.66	32	40.66	40	41.33	40	8	6	8
	II	31.33	32	41.33	42	44.00	44	6	6	0
	III	30.66	30	40.66	40	37.33	36	2	6	4
	IV	32.00	32	37.33	36	38.66	38	2	4	2
	V	32.00	32	40.66	40	40.66	40	2	2	2
A.R.	I	49.33	50	54.00	54	50.66	50	2	0	2
	II	50.66	52	58.66	58	52.66	52	4	6	2
	III	49.33	50	55.33	56	53.33	54	2	6	2
	IV	49.33	48	58.66	58	52.00	52	4	2	4
	V	48.00	48	60.66	62	56.66	56	0	4	2
Average Range 3.14 4.06 2.66										

were smaller than for the tests in either the preceding free field or right ear. Eliminating the one highly erratic setting of Subject 2 of Group II, the average range for the free field is changed from 5.15 db to 4.4 db. Such a change makes the average range for the two groups more nearly alike. There was a trend, then, toward narrowing the range as the subject gained experience. The first setting was likely to be less reliable than subsequent settings, but the difference was not significant. Apart from this rather logical characteristic, there were no other trends evident. In Group I we found little or no difference between the first setting of the first session and the last setting of the last session. An equal amount of variation could exist within a single session of three trials. (See Fig. 2) The exception was A.H. whose median setting was consistently lower from session to session, as Table 3 illustrates. The explanation may be that she used a different criterion for the first two sessions, becoming more reliable in the last three sessions. The order of testing was not significant. The settings in the free field of any session were equally as reliable as the settings in the subsequent monaural settings.

There was little reason to expect fatigue to influence the results of the test, since the time spent on each test was never more than a few minutes. However, the factor of fatigue was given consideration and found to be of no consequence. This is borne out significantly by the general ab-

sence of trends associated with the complete random scatter of the scores within each session.

Familiarity with the test material had no effect on the scores. The lack of trends verifies this statement.

There was no significant difference between the group mean of the three settings and the group median of the three settings. (See Tables 4 and 5) This indicates that there were few erratic responses. We can conclude, however, that the median score of the three settings is of greater statistical value because of the small sampling involved and the ease with which the whole numbers can be handled. The median proves to be a quick and accurate method of determining the threshold for connected discourse. Table 5 compares the mean of the means and the mean of the medians of Group I.

Table (5) Mean of the Means of the Sessions compared with the Mean of the Medians. Group I.

Sex	Normals N = 6		Hard of Hearing N = 1		Group I N = 7	
	Mean	Median	Mean	Median	Mean	Median
FF	26.24	25.25	49.53	49.0	29.54	29.0
HE	31.06	31.0	59.46	59.0	35.30	35.77
LE	30.60	30.40	55.01	54.0	34.16	35.04

The time required for each setting was generally of short duration. There were, of course, individual differences which varied greatly--from as little as 15 seconds to over 2 minutes. The average time for each setting was approximately

45 seconds. The first setting usually required more time. Most of the subjects reduced their time as they gained in experience, but a few remained slow and deliberate to the very last setting.

Of final interest, the preliminary study revealed that the characteristics of the responses of the hard-of-hearing individuals were essentially the same as those of the normal group. An examination of the responses of A.R. of Group I (Table 3) will show that she was more reliable than most of the normals of that group.

It became apparent after a time that the instructions to the subject needed to be revised. We were getting fair reliability with each series of 3 settings and from series to series, but we were disturbed by the comparatively poor validity of the test. The deviations from the #0 threshold were scattered over a range of 22 db. Obviously, such a range could not be tolerated in any valid articulation test. However, the reliability of the test encouraged us; because it seemed to indicate that regardless of what criterion the subject took as his threshold of intelligibility, he retained it from test to test. A more accurate description of the threshold was needed in order to pin the threshold to a definite point on the intensity scale. By careful listening and experiencing the test many times, we were able to find a positive clue which we thought represented a definite threshold. The original instructions were, "the lowest point at which you can hear the speech without effort." This was

changed to "the lowest point at which you can understand everything the speaker is saying. You may reach a point where some of the words seem to drop out. By turning the dial back a point or two those words will come in more distinctly." The results proved more satisfying as a consequence. Subjects 13 of Group II and M of Group III initiated the changed directions. An examination of Tables 2 and 4 will reveal that the amount of scatter from the #9 threshold was considerably reduced. The range of settings, however, remained more or less unchanged. Group I used the original instructions throughout the entire series of five sessions. We can conclude, then, that the instructions to the subject are critical in producing a test of high validity and reliability from monitored connected discourse.

Figure 2. Medians of the Medians of the Sessions of Group I. Note the absence of any systematic trend such as might be caused by learning.

Table 4. Thresholds of Intelligibility for Group III
(Decibels re 0.0002 dynes/cm²)

Sub- ject	Settings		Range	Mean	Median	Test #9	Deviation of Median from #9	
A	92	96	94	4	94.00	94	0	
B	86	80	88	4	88.00	88	6	
C	80	82	84	4	82.00	82	8	
D	Old Instructions	66	70	70	4	68.66	70	-4
E		68	68	68	0	68.00	68	0
F		80	82	86	6	82.66	82	13
G		78	76	76	2	76.66	76	4
H		58	58	62	4	59.33	58	-4
I		68	68	68	0	68.00	68	-1
J		72	74	76	4	74.00	74	11
K		78	78	78	0	78.00	78	5
L		70	72	72	2	71.33	72	12
.....								
M	New Instructions	74	74	76	2	74.66	74	6
N		74	76	80	6	76.66	76	9
O		78	76	72	6	75.33	76	-2
P		80	80	80	0	80.00	80	9
Q		82	82	84	2	82.66	82	5
R		84	86	86	2	85.33	86	6
S		68	70	70	2	69.33	70	8
T		102	104	104	2	103.33	104	2
.....								
Averages		76.9	78.1	78.7		77.9	77.9	73.0

CHAPTER IV

DESCRIPTION OF MAJOR STUDY

The procedure established for the major study evolved from the preliminary study reported in the preceding pages.

Subjects

Both normal and hard-of-hearing subjects were chosen for the experiment on the reliability and validity of monitored connected discourse as a test of the threshold of intelligibility. The subjects were divided into 3 groups described as follows:

Group I consisted of 25 individuals whose two ears were normal for hearing pure tones and speech. The ages of the group ranged from 17 years to 39 years. The pure tone thresholds were determined by means of an Audio-Development Company audiometer. A subject's threshold was regarded as normal if it varied no more than 10 db from 0 intensity on the audiometer from 256 cps to 4096 cps. Acuity for speech was assessed by administering the Harvard Test #9. Any speech threshold was considered normal if it was within 2 times the standard deviation of the norm.⁶ The norm, subsequently determined from the group, was 22 db (re 0.0002 dynes/cm²). The standard deviation for the group was 2.6 db.

⁶J.C.Steinberg, H.C.Montgomery, and M.B.Gardner, "Results of the World's Fair Hearing Tests," Bell Sys. Tech. Pub., Monograph B-1256, pp. 27-29 (1940).

Group II contained 15 hard-of-hearing subjects who were candidates for the fenestration operation and had come to the Hearing Clinic at Central Institute for a series of diagnostic tests.⁷ An otologist determined clinically the type of impairment for each subject as either pure conductive or mixed deafness. The ages ranged from 17 years to 41 years.

Group III included 32 hard-of-hearing subjects who had undergone the fenestration operation and had returned to the clinic for evaluation of the results. Since the natural function of one of the ears of each subject of this group had been disturbed, the data will of necessity be treated in a manner different from the other groups.

Procedure

The testing procedure for the 3 groups was as follows:

Group I -- The subject was first tested on the audiometer to determine if his hearing was normal for pure tones. This, of course, was essentially a screening test. If his hearing was thus determined normal for both ears, he was taken into the sound treated chamber and seated in a comfortable chair. The PDR-10 earphone was placed over the right ear and the #9 test was administered to determine his threshold for speech. If the subject's threshold for speech was normal, the experimental test with connected discourse followed. An attenuator box was placed on his lap, and he was handed a card bearing typewritten instructions for the test. The subject read as follows:

~~W. J. L. S. R. Silverman~~ S. R. Silverman, "Diagnosis and Evaluation of Fenestration," Laryngoscope, 56: 536-538, Sept., 1946.

You are going to hear a man talking. It will be loud; but, by turning the dial away from zero, you can make the speech softer and softer. Please turn it to the lowest point where you can understand everything the speaker is saying. You may reach a point where some of the words seem to drop out. By turning the dial back a point or two, those words will come in more distinctly. That is the point we are seeking. Call the dial reading off to me. Turn the dial back to zero and begin again. Repeat the setting twice -- three times in all. The tests will not all start at the same loudness; therefore, do not expect to arrive at the same number on the dial each time.

Be quick but be accurate.

The dial reading for each setting was recorded. The time required for each setting was measured in seconds and recorded. Time began the moment the subject began hearing the speech and ended when the subject called off the dial reading.

To make certain that the subject would not use his initial setting as a reference for the next two settings, the observer changed the level of loudness of the speech by means of the attenuators outside the testing room. The observer attempted to keep the starting loudness within a range of from 10 db to 30 db above the estimated threshold of intelligibility of the subject. This allowed the subject plenty of room to play up and down the intensity scale without confining him to either the first or last few detents on his attenuator. The subject's attenuator was capable of attenuating 55 db in 2 db steps. The starting level of the speech was always kept at an even number of db so as to conform with the even numbers on the subject's attenuator.

The procedure used in testing the right ear was repeated for the left ear. By this time the subject had become sophisticated enough to disregard the written instructions. If,

however, there was any question about the interpretation of the instructions, he was asked to reread them. It was rarely necessary to make verbal elaborations.

If both ears proved normal for hearing speech, the tests were continued by retesting each ear. The right ear was tested first. The #9 lists were varied from test to test. The same recording of connected discourse was used repetitiously, since the preliminary experiment had shown that familiarity with the material had no apparent effect on the results. Below is an outline of the order of the tests for Group I:

1. Harvard Test #9, right ear
2. TICD test, 3 settings, right ear
3. Harvard Test #9, left ear
4. TICD Test, 3 settings, left ear
5. Retest #9, right ear
6. TICD, 3 settings, right ear
7. Retest #9, left ear
8. Retest TICD, 3 settings, left ear

Groups II and III -- the procedure for these groups will be defined jointly, since the method employed for both was identical. The procedure differs somewhat from that used for Group I. The TICD test was incorporated into the regular testing routine of the fenestration patients in such a way as to disrupt the routine as little as possible and at the same time to simulate closely the procedure followed with Group I. Since the #9 tests were given initially, it was an easy matter to shift quickly to the TICD test. The TICD test was the only test that was repeated.

An outline of the order of the tests for Groups II and III is as follows:

1. Harvard Test #9, right ear
2. TIOB test, 3 settings, right ear
3. Harvard test #9, left ear
4. TIOB test, 3 settings, left ear
5. Retest TIOB, right ear
6. Retest TIOB, left ear

CHAPTER V

DATA AND DISCUSSION

In discussing the reliability and validity of connected discourse as a test for the threshold of intelligibility, we can best investigate the problems by keeping the following questions in mind:

1. What is the range of the settings? Do they lie within the limits of accepted tests of thresholds of hearing.
2. Are the test-retest values of the TICD related to the test-retest values of the threshold for spondees?
3. What is the relationship between the TICD and the threshold for spondees.

The first two questions are directed toward the reliability of the TICD test. The third question examines the validity of the test. These questions will be considered for each group and then comparisons will be made between the normal and hard-of-hearing groups.

Group I

Range: The subject obtained his threshold for connected discourse by adjusting an attenuator to a point which he judged to be the lowest intensity at which he could hear everything the speaker was saying. A series

of 3 judgments, or settings, were made in quick succession. The right ear was tested in the first series, followed by a series of 3 settings for the left ear. Retests were made for the right and left ears respectively, repeating the 3 settings. The session, then, consisted of 4 series of 3 settings each, 6 settings being made for each ear. All the settings are given in Table 6.

The average range for the 6 settings in one ear was 5.0db (see Table 7).

The average range for a series of 3 settings was 2.82 db.

The average range for the first series of 3 settings was 0.4 db greater than the average range for the 3 settings of the final (fourth) series. This difference is too small to be indicative of a trend. It may, however, involve a slight learning factor.

58% of the 6 settings were within a range of 4 db.

60% of the 3 settings were within a range of 2 db.

That the range for the 6 settings is greater than the range for the 3 settings is not surprising when we recall that the average intensity level of the recording varies from moment to moment over a range of 3 db. We must also appreciate that the subject can not adjust the intensity level in steps of less than 2 db. He may actually be making an accurate judgment according to the intensity level of the recorded speech at that particular moment of listening. The factor of human error must also

Table C. Group 3. Individual Settings for Threshold of
Intelligibility for Connected Discourse.
N = 20

The columns are arranged in the order of testing.
(Decibels re 0.0001 dynes/cm²)

Sub- ject	First Series (Right Ear)			Second Series (Left Ear)			Third Series (Retest, RE)			Fourth Series (Retest, LE)		
M.B.	28	28	26	24	24	22	24	26	26	24	26	22
L.B.	26	26	26	30	26	30	26	24	26	26	26	26
E.L.	28	32	32	26	26	24	30	30	32	26	24	24
M.W.	22	24	23	24	24	26	24	22	22	24	26	28
H.W.	24	22	22	22	22	24	20	22	22	24	22	24
M.T.	24	18	18	24	24	24	26	28	26	22	20	26
K.M.	26	22	24	22	22	20	20	22	20	22	22	24
P.H.	24	22	26	22	24	24	20	22	22	22	24	22
J.D.	20	24	24	20	18	18	24	24	20	22	18	20
J.S.	28	28	32	20	20	24	30	26	26	22	24	22
R.A.	32	30	28	24	24	24	26	24	24	24	22	24
P.N.	24	20	22	32	28	26	24	22	22	26	26	28
K.G.	22	24	26	24	26	28	26	32	28	26	26	28
H.B.	30	28	30	26	26	26	26	26	30	26	26	28
W.S.	26	26	26	20	20	24	24	22	26	20	20	18
D.L.	14	16	14	18	16	14	16	16	16	18	16	20
B.D.	28	22	24	22	22	18	22	20	24	20	22	24
C.L.	26	20	18	22	24	24	20	18	20	26	26	26
D.S.	26	20	20	18	20	18	18	22	12	20	22	20
H.B.	18	18	20	16	18	18	22	16	14	16	16	14
I.T.	16	16	18	20	18	24	18	22	22	24	20	22
M.P.	22	24	22	22	22	22	28	22	24	24	24	26
M.A.	22	22	24	22	22	20	22	22	22	22	26	20
J.C.	20	22	24	26	26	26	24	22	22	22	22	22
D.H.	26	26	26	24	26	26	20	22	26	22	22	20

Table 7. Test-retest results for Group I showing deviation of the TIOD from the #9 threshold.
N = 25 Right Ears. (Normals).

Subject	Test		Retest		Range of the 6 settings
	db Above	db Below	db Above	db Below	
M.B.	2		2		4
L.W.	0	0	0	0	2
E.N.	6		4		4
M.M.		1		3	6
H.M.	0	0	0	0	4
M.T.		6	2		10
K.N.		1		2	6
P.N.	0	0		2	6
J.D.	2		2		4
J.S.	1		1		4
R.M.	2			4	8
F.M.	0	0	0	0	4
K.G.	0	0	4		10
W.H.	6		1		4
W.M.	6		4		4
D.L.		8		6	2
B.D.	0	0		4	8
O.L.		6		5	8
D.S.	2		0	0	10
H.B.		4		6	8
I.T.		4	0	0	6
M.P.	4		6		6
M.A.		2		2	2
J.O.	3		1		4
D.M.	6		2		8
Averages	+.3 db		-.2 db		5.68 db

be considered when we are working with such small units of sound intensity. His chances of making a misjudgment of 2 db are doubled when he increases from 3 settings to 6 settings. One might reasonably expect a greater variation.

Only one individual had a range of 10 db for a single series of 3 settings. This was D. S. who gave an erratic setting of 12 db (re 0.0002 dynes/cm²). This is close to the threshold of detectability. It is possible that he made a 10 db error in reading the number on the dial of his attenuator. Such an assumption is permissible when we note in Table 6 that his 11 other settings ranged from 18 db to 22 db. The real setting was probably 22 db. However, the erratic setting had no marked effect on the median score which compared favorably with the median scores of his other 3 series.

If we consider range as a good measure of reliability, there is little question that the responses of the normal group were reliable. Witting and Hughson,⁸ in a kind of test-retest evaluation of responses by audiometry, found that 57% of a group of 116 hard-of-hearing patients were within a range of 5 db for the frequencies 512, 1024, and 2048. 88% were within a range of 10 db for the same frequencies. These percentages were obtained by comparing

⁸Witting, E.G., and Hughson, W., "Inherent Accuracy of a Series of Repeated Clinical Audiograms", Laryngoscope, 50: 259, 1940.

the first audiogram of every patient with the average of 10 or more audiograms made on that patient over an average period of 28 months. Although comparison with the Witting & Hughson study is not exact, it does illustrate that the ranges of the TICD are relatively small. Actually, they approximate the standard deviations at all frequencies of 126 audiograms taken on 7 normal ears.⁹

Test-retest. The differences between the average score of the first test and the average score of the second test for the #9 and TICD tests were essentially the same. The score for the #9 test was obtained by the tabular method described by Hudgins, et al. The score is a db value corresponding to the subject's threshold, defined as the level at which he is able to understand 50% of the words. The score for the TICD test was the median setting of a series of 3 settings. The median of the 3 settings was chosen rather than the mean because it was less likely to be affected by extreme variations. The average difference for the test-retest of the #9 for the right ear was 2.28 db compared with a difference of 2.16 db for the TICD (Table 7). The test-retest scores for the left ear were slightly less for both. This may be attributed to a slight learning factor. The differences for the #9 test and the TICD test for the left ear were 1.88 db and 1.6 db respectively (Table 8). In both

⁹Ibid., p. 262.

Table 8. Test-retest results for Group I showing deviation of the TIGD from the #9 threshold.
N = 28 Left Ears. (Normals).

Subject	Test		Retest		Range of the 6 settings
	db Above	db Below	db Above	db Below	
M.B.	0	0	0	0	6
L.W.	6		2		4
E.M.	4		2		4
M.M.		2	0	0	4
H.W.	0	0	0	0	2
M.T.	2		4		4
K.M.	1		1		4
P.M.	3		1		2
J.D.		4		2	4
J.S.		1	1		4
R.M.	2		2		2
P.M.	4		2		6
K.O.	6		6		4
W.H.	2		2		2
D.L.		3		1	6
B.D.	2		2		6
C.L.		2	0	0	4
D.S.	0	0	2		4
H.B.		6		7	4
I.T.		2	0	0	6
M.P.	6		7		4
M.A.		1		1	6
J.C.	2			2	4
D.M.	6		1		6
W.M.	2		2		6
Averages	+1.04 db		+.98 db		4.32 db

cases the test-retest values for the TIOD were less than the test-retest values of the #9 test.

Combining the ears, the test retest scores were 2.08 db for the #9 test and 1.88 db for the TIOD test. Conceding that approximate test-retest scores are a function of reliability in a test, the normal group gave evidence that the TIOD test is a reliable test.

Relationship to Control Test: There is a definite relationship between the TIOD score and the #9 score. The mean for the TIOD for the group was .77 db higher than the mean for the #9 test.

The mean for the TIOD was 23.23 db and the mean for the #9 test was 22.46 db (re 0.0002 dynes /cm²).

The standard deviations (sigmas) were 3.77 db (TIOD) and 2.63 db (#9). The sigma for the #9 test compares favorably with the standard error of measurement reported by C.V. Hudgins, et al,¹⁰ who obtained from 30 normal subjects under very favorable conditions a standard error of 2 db. A standard error of measurement under less favorable conditions was 2.4 db.

From the sigma of the #9 test, we know that at least two thirds of the #9 scores were within a range of ± 2.63 db of the mean. For the purpose of comparison we will use the #9 sigma as a gauge for measuring the positive and negative deviations of the TIOD from the #9 threshold for each sub-

¹⁰Op. cit., p. 70.

ject. The mean of the TICD was approximately 1 db above the mean of the #9 test; therefore, 1 db above the #9 score will represent the mid-point from which we shall measure the positive and negative deviations. Any TICD score within 4 db above the #9 score or within 2 db below the #9 score will be within the range in which two thirds of the #9 scores lay. 74% of the TICD scores were within this area. Only 8% of the TICD scores were beyond twice the sigma of the #9 test. All were below the area. The three erratic subjects were D.L., H.B., and C.L., all of whom probably used the threshold of perceptibility as a criterion rather than the threshold of intelligibility. D.L. and C.L. gave more satisfactory scores when the left ear was tested. H.B. persistently chose a low level throughout the entire session. Actually, D.L. could not repeat the thoughts expressed in the speech sample when he was asked to listen for a time at the intensity level he had chosen as his threshold of intelligibility. This was done as a check only after the experimental tests were completed.

The close relationship of the #9 score and the TICD score for the normal group is a point of evidence in favor of the validity of monitored connected discourse as a test of the threshold of intelligibility.

Group II

Group II was composed of 15 hard-of-hearing subjects who were candidates for the fenestration operation. (30 ears were studied.)

Range. The average range for the 6 settings in one ear was 5.26 db. All the settings appear in Table 9.

The average range for a series of 3 settings was 3.0 db.

The average range for the first series of 3 settings was 0.54 db smaller than the average range for the last series of 3 settings. This is the reverse of what appeared for the normal group and substantiates the lack of trend.

47% of the 6 settings were within a range of 4 db.

83% of the 6 settings were within a range of 6 db.

The greatest range for the 6 settings was 8 db, which included 17% of the 6 settings.

55% of the 3 settings were within a range of 2 db.

Test-retest. The section on the testing procedure explained that only the TICD test was repeated for the hard-of-hearing subjects. The #9 test was done only once for each ear.

The difference between the average score of the first TICD test in the right ear and the second TICD test was 2.8 db.

For the left ear the average difference between the first TICD test and the second TICD test was 2.1 db (See Table 10).

The average difference for both ears combined was 2.5 db. This is 0.6 db greater than the average difference for the normal listeners.

Relationship. The relationship between the #9 score and the TICD score which we noted for the normal group existed also for the hard-of-hearing subjects of Group II. The aver-

Table 2. Group II. Individual Settings for Threshold of Intelligibility for Connected Discourse.
N = 15
Arranged in order of magnitude of δ_9 scores for Right Ear.
(Decibels re 0.0002 dynes/cm²)

Subject	Right Ear			Left Ear		
	Test	Retest		Test	Retest	
A.E.	68 66 68	62 62 60		84 86 83	82 82 82	
F.E.	66 66 64	66 66 68		86 86 88	90 90 88	
A.E.	68 68 72	70 72 70		76 80 80	72 72 74	
H.R.	70 74 72	76 66 68		74 72 76	70 72 74	
A.D.	74 80 82	70 70 74		72 68 74	66 66 70	
E.C.	74 72 76	80 78 80		82 82 82	82 86 84	
E.L.	74 72 72	68 70 70		74 74 76	72 72 76	
S.W.	84 86 86	88 86 88		88 70 72	76 72 70	
R.B.	80 78 78	78 76 78		74 76 76	78 72 74	
J.D.	78 76 78	80 76 82		78 78 74	78 78 76	
M.R.	78 80 78	76 78 76		62 62 60	62 62 66	
S.P.	82 82 82	80 78 80		48 50 48	50 50 52	
W.H.	84 80 80	84 86 86		76 76 74	78 76 74	
D.R.	82 80 84	80 84 82		74 70 72	72 72 76	
S.Z.	92 92 90	92 88 90		90 92 90	88 86 92	

Table 10. Test-retest results for Group II showing
Deviation of the TCD from the #9 Threshold.
N = 30 Hard of Hearing Ears (15 right, 15 left)

Subject	Test		Retest		Range of the 6 Settings
	Db Above	Db Below	Db Above	Db Below	
Right	A.E.	4		2	6
	P.E.			1	2
	A.D.	1	3		4
	W.R.	4	0	0	8
	A.D.	9		1	8
	E.C.	4	6		6
	E.L.			5	6
	S.W.	9		11	4
	R.B.	1	1		4
	J.D.	0	2		6
	M.R.	0		2	4
	S.R.	3	1		4
	W.H.	1	7		6
	D.R.	3	3		4
	S.Z.			6	4
Averages		+2.06 db	-.33 db		5.2 db
Left	A.E.	0		2	4
	P.E.	9	11		4
	A.D.	7		1	8
	W.R.	0		2	6
	A.D.	5		1	8
	E.C.	4	6		4
	E.L.			3	4
	S.W.	6	6		8
	R.B.	5	3		4
	J.D.	4	4		4
	M.R.			3	6
	S.R.		1		4
	W.H.			1	4
	D.R.			3	6
	S.Z.			9	6
Averages		+1.6 db	+.53 db		5.33 db

age deviation of the TIOD from the #9 threshold was 0.96 db, a positive deviation very similar to the 0.77 db positive deviation of Group I. However, fewer of the deviations of Group II were within the limits of ± 3 db from 1 db above the #9 threshold. Only 60% of the scores were within this area, compared to the 74% of the scores for Group I. This will be discussed in greater detail later when the responses of the normal and hard-of-hearing subjects are compared.

Group III

Group III consisted of 32 hard-of-hearing subjects who differed from Group II in that one of the ears had been fenestrated. The ears were grouped as unoperated ears and operated ears. The sequence of testing one particular ear first was of necessity abandoned. Discussion of the actual threshold scores for the operated ears will be excluded, since this information is being reserved for publication at a more appropriate time.

Range: The average range for the 6 settings in the unoperated ears was 4.19 db. (See Table 11 for all the settings.)

The average range for the 6 settings in the operated ear was 4.2 db.

The average range for a series of 3 settings for the unoperated ears was 2.34 db.

The average range for a series of 3 settings for the operated ear was 2.44 db.

Table 11. Group III. Individual Settings for the Threshold of Intelligibility for Connected Discourse.

N = 32 Unoperated Ears (Hard of Hearing)
(Decibels re 0.0002 dynes/cm²)

Sub- ject	Test	Retest	Sub- ject	Test	Retest
A.K.	68 70 72	70 70 70	B.M.	68 68 68	68 70 70
R.E.	66 66 64	68 64 66	L.C.	74 76 76	76 74 74
D.E.	58 58 58	58 58 60	J.E.	92 92 94	92 92 92
M.C.	72 72 70	76 74 74	A.O.	78 74 80	80 78 80
L.C.	70 68 68	76 70 70	L.W.	82 82 80	80 80 82
R.K.	72 70 72	72 72 72	I.B.	82 84 82	82 82 80
M.E.	68 66 72	70 72 72	W.E.B.	80 82 82	78 80 82
C.J.	66 66 66	66 66 64	S.G.	80 78 78	76 76 78
M.M.	72 74 76	74 74 72	M.O.	80 80 80	86 80 82
J.P.	78 76 72	76 76 74	C.M.	84 84 84	84 82 84
R.O.	78 78 78	80 80 80	E.G.	86 84 86	84 84 84
K.D.	72 70 66	66 70 70	P.S.	78 78 76	78 76 76
E.T.	74 74 72	82 80 80	MMo.	90 90 92	88 90 88
A.F.	68 68 70	68 72 70	E.B.	94 92 94	94 96 96
G.H.	76 74 72	72 70 74	S.S.	90 90 92	94 90 90
R.S.	84 84 86	82 82 82	W.B.	96 98 96	94 98 96

69% of the 6 settings for the unoperated ears were within a range of 4 db; 94% were within a range of 6 db.

66% of the 6 settings for the operated ears were within a range of 4 db; 91% were within a range of 6 db.

72% of the 3 settings for the unoperated ears were within a range of 2 db.

67% of the 3 settings for the operated ears were within a range of 2 db.

The greatest range for the 6 settings was 10 db. This occurred only once in 64 series of 6 settings for both the operated and unoperated ears. Only 8% of the ranges for the 6 settings for both ears were as great as 8 db.

Test-retest: The difference between the average score of the first TICD test and the average score of the second TICD test for the unoperated ear was 1.43 db. (See Table 12).

For the operated ear the average difference between the first and second TICD test was 1.5 db. (See Table 13).

The average difference when the scores were combined for both ears was 1.46 db. This was 0.4 db less than the average difference for the normal listeners.

Relationship: The average deviation of the TICD from the #9 threshold was greater for Group III than for

Table 12. Test-retest results for Group III showing deviation of the TICD from the #9 threshold.
N = 32 Unoperated Ears.

Subject	Test		Retest		Range of the 6 Settings
	Db Above	Db Below	Db Above	Db Below	
A.K.	15		15		4
R.E.	3		3		4
D.F.		7		7	4
M.C.	6		8		6
L.C.	2		4		9
R.K.	6		6		2
H.B.	0	0	4		9
C.J.		3		3	2
M.M.	5		5		4
J.F.	7		7		6
R.O.	7		9		2
K.D.		1		1	6
E.T.	2		8		10
A.H.		4		2	4
G.H.	2		0	0	6
R.S.	11		9		4
B.M.		7		5	2
L.C.	1			1	2
J.E.	16		16		2
A.O.	2		4		6
L.W.	5		3		2
I.E.	5		5		4
W.E.	4		2		4
G.G.	0	0		2	4
M.O.	1		3		6
C.M.	4		4		2
E.G.	2		0	0	2
P.S.		7		9	2
M.M.	4		2		6
H.B.	7		9		4
S.S.	1		1		4
W.B.	0	0	0	0	4
+ 2.78 db			+ 3.03 db		4.19

Table 13. Test-retest results for Group III showing
Deviation of the TICD from the #0 threshold.
N = 32 operated ears.

Subject	Test		Retest		Range of the 6 Settings
	Db Above	Db Below	Db Above	Db Below	
A.K.	13		7		6
R.E.		4	0	0	4
D.E.		7		5	6
M.E.	8		8		2
L.C.C.		1		1	2
R.K.	7		5		2
M.L.		1	1		2
C.J.		1		1	6
M.M.	3		3		6
J.P.	5		7		6
R.O.	1		3		4
K.D.		1		1	0
E.T.	4		4		4
A.H.	0	0	0	0	2
G.H.	8		6		4
R.S.	7		7		4
B.M.		1		1	2
L.C.	1		3		4
J.E.	15		15		4
A.O.	1		3		6
L.W.	5		5		4
I.B.	3		3		6
WEE	6		4		2
O.G.	2			2	8
M.O.	1		3		8
C.M.	0	0		2	6
E.G.	0	0	0	0	4
P.S.		3		1	2
M.Mo.	14		16		4
E.B.	2		8		8
S.S.	2		2		2
W.B.	2		0	0	4
Averages	+ 2.84 db		+ 3.09 db		4.2 db.

Group II, but the percentage of scores within the limits of ± 3 db from the level of 1 db above the #9 threshold was much the same.

52% of the TICD for the unoperated ears were within the area ± 3 db from the level 1 db above the #9 threshold.

62% of the TICD scores for the operated ears were within this area.

57% of the TICD scores for both ears were within the area.

The average deviation of the TICD from the #9 threshold for the unoperated ear was 2.9 db.

The average deviation for the operated ear was 2.96 db.

The average deviation for both ears was 2.93 db. This is 2.15 db greater than the average deviation for the normal group and 1.97 db greater than the average deviation for the other hard-of-hearing group. The greater deviation is accounted for by the extreme deviations of subjects J.E., A.K., and M.M. These deviations will be discussed later.

The data reveals no significant difference between the responses for the unoperated and operated ears; therefore, the operated ears will be treated hereafter as hard-of-hearing ears.

Comparison of Normal and Hard-of-Hearing Subjects

Range: The following table compares the ranges of the normal and hard-of-hearing subjects:

Number of Ears	Average Range for the 6 Set- tings in 1 ear	Average Range for a Series of 3 Settings	% of the 6 Settings within a Range of 4 db.	% of the 3 Settings within a Range of 2 db.
50 Normal	5.0 db	2.82 db	58%	60%
94 H.O.H.	4.55 db	2.59 db	61%	65%

From every comparison of the ranges we see that the hard-of-hearing subjects as a group were more reliable in repeating a setting than the normal group. This is a point in favor of the reliability of TICD test; since, as a rule, the responses of the hard of hearing are less reliable than the responses of normals.

Comparing maximum ranges we find that the hard of hearing subjects did better than the normal subjects. 3 of the normals had a range of 10 db for the 6 settings in one ear; 1 hard-of-hearing subject had an equal range. 1 normal had a range of 10 db for a series of 3 settings; none of the hard of hearing had a range of over 8 db. A distribution of the ranges of the 6 settings of the Normal and H.O.H. is shown in Table 14.

The data from normal and hard-of-hearing subjects used in this study show that the responses to the TICD test are equally reliable or more reliable than the responses to pure tone audiometry as reported in Witting and Hughson,¹¹

¹¹Op. cit., p. 268.

Table 14. Distribution of the ranges of the 6 settings of the normal and hard of hearing listeners.
 N = 50 Normal Ears, 25 Listeners.
 62 Hard of Hearing Ears, 47 Listeners.

Range in Decibels	Percentage Distribution	
	Normal	Hard of Hearing
0	0	0
2	14	18
4	44	44
6	26	25
8	10	11
10	6	2

who state that there is an "inherent error of slightly less than 5 db" by audiometry.

Test-retest: The average difference between the score of the first #9 test and the score of the second #9 test for the normal group was 2.08 db. (A retest of the #9's was not made on the hard-of-hearing subjects).

The average difference between the first TICD score and the second TICD score for the normal subjects was 1.88 db compared to the average difference score of 1.98 db for the hard-of-hearing subjects, a value slightly less than that obtained from the #9 test. This is an indication that the TICD test is comparatively reliable, falling well within the limits of 2.8 db which Hudgins, et. al.,¹² concluded was the standard error of measurement for the #9 test for both normal and hard-of-hearing individuals. They state: that

"a single administration of the Test 9 to normal and hard-of-hearing subjects will, two-thirds of the time, give a score that falls within 2.8 db of the subject's true score on that form of test."

Relationship of the TICD to Test #9: The relationships presented in the table below indicate that the TICD test is more valid as a test of the threshold of intelligibility for normally hearing individuals than for the hard of hearing.

¹²Op. cit., p. 70

Table showing the Relationship of the TICD to Test #9

	50 Normal Ears	94 HON Ears
#9 Scores within 3 db of the Mean.	77%	NT
TICD Scores within 3 db of a level 1 db above the #9 Threshold.	74%	58%
TICD Scores within 6 db of a level 1 db above the #9 Threshold.	92%	80%

The scores of the normal listeners tend to follow the laws of normal frequency distribution. There are present the occasional large deviations from the average which are expected in any normal curve. These few deviations can further be explained by the fact that there are several clearly defined thresholds lying relatively close to each other on the intensity scale. These are (1) the threshold of detectability, the intensity level at which the listener is able to detect the presence of speech sounds about half the time without being able to identify any of the sounds themselves; (2) the threshold of perceptibility, the level at which the listener is able to understand with considerable effort the gist of the speech; (3) the threshold of intelligibility, the level at which he can understand the speech with little effort;¹³ and (4) the less clearly defined

¹³ OERD Report No. 3802, "Articulation Testing Methods II," pp 58-59, Nov., 1944.

threshold of ~~comfort~~, the level at which he can understand the speech most easily. This is sometimes called the level of preference. Unless the threshold of intelligibility is clearly described to the listener, he is likely to select one of the related thresholds. This explains why a subject can give reliable responses even when there is a great discrepancy between the TICD score and the #9 score.

There was a greater percentage of extreme deviates among the hard-of-hearing subjects. The type of deafness does not seem to account for the more frequent discrepancies of the hard-of-hearing group. One of the two most extreme positive deviates (TICD score above #9 score) was diagnosed by an otologist as having conductive deafness. The other extreme positive deviate was diagnosed as having a mixed deafness with predominant nerve involvement. On the other hand, another deviate who had one of the most extreme negative deviations was diagnosed as having a mixed deafness with predominant nerve involvement; a second extreme negative deviate had a pure conductive type deafness. Each extreme is represented by cases of both conductive and mixed nerve deafness. (There were no pure nerve deafness cases in the study. They, of course, are not candidates for the fenestration operation and would be eliminated before being sent to the clinic for the diagnostic tests).

An examination of the medical histories of the hard-of-hearing subjects who showed the greatest deviations from

the #9 score revealed no general characteristics that would differentiate them in any way from the larger group.

The TICD test depends on subjective judgment; therefore, there are apt to be more individuals who vary exceptionally from the true threshold of intelligibility than from a more objective articulation test. It seems that some individuals select an intensity that is higher than what they really need in order to meet a high standard of the threshold of intelligibility. Others may swing to the other extreme and are satisfied with too little intensity. Where this is suspected the operator might correct the listener by reviewing the instructions and making proper elaborations. This uncontrolled element is best illustrated in this study by D.L., C.L. and H.B. of the normal group and by A.K., R.B., J.H., and S.W. of the hard-of-hearing subjects. A positive deviation of 15 and 16 db from the #9 threshold is represented by 2 hard-of-hearing subjects as the greatest deviations. Large deviations in any hearing test are not unusual for the hard of hearing, whose responses may be affected by tinnitus or some other contributing factor of deafness.

The Harvard University Test #9 is an accepted test of the threshold of intelligibility. We have pointed out that two-thirds of the time a single administration of that test will give a score within 2.3 db of the subject's true threshold for spondaic words. We have discussed the relationship of the TICD test to the #9 test and noted that the TICD

scores of 74% of the normals and 58% of the hard of hearing were within ± 3 db of an intensity level 1 db higher than the individual #9 scores. 92% of the normals and 30% of the hard of hearing were within ± 6 db of a level 1 db above the individual #9 scores. It is evident, then, that there is a close relationship between the scores of the 2 tests. Because of its subjectivity, the TICD test is not as exact a measure of a threshold of intelligibility as the #9 test; but it will, 58% of the time, yield a score within 3 db of the threshold of intelligibility for connected discourse.

Face Validity of the TICD Test: Social adequacy from the standpoint of hearing rests on the extent to which an individual can hear ordinary conversation. The connected discourse used in this study is a sample of everyday speech taken directly from the radio networks. The high face validity of the TICD test, then, is the characteristic which gives it value as an actual test of the type of speech one listens to day by day. By means of the TICD test, it is possible to determine on an intensity scale the degree of loudness a given individual requires in order for him to get along adequately in his social environment.

Time Required for TICD Test: One of the minor objectives of the study was to verify our presupposition that the threshold of intelligibility for connected discourse would be less time-consuming than ordinary articulation tests. This was

true for the great majority of subjects.

The time required for the individual settings ranged from 10 seconds to 2 minutes. The large majority of subjects made the settings in from 20 to 40 seconds each. Usually the first setting required more time, but the subjects learned quickly and thereafter were able, as a rule, to make a single setting in about 20 seconds. Those few individuals who required a minute or more were no more accurate than those who took much less time. In fact, the slow individual was likely to vary more from setting to setting than the average individual.

The time required, then, for the TICD test was much less than the 4 to 5 minutes required for ordinary articulation tests, such as the Western Electric 40 test, and the Harvard University spondee, sentence, and phonetically balanced word tests. The differences of the time spent on each setting by the various subjects were individual differences and not the result of a hearing deficiency. There was no difference between the time required by the normals as a group and the time required by the hard of hearing as a group. Time was not recorded on all the subjects throughout the entire study, because it was observed that the majority made the settings in about equal time.

Fatigue. Another presupposition was that the TICD test might be less fatiguing. No special attempt was made to

measure fatigue (mental fatigue, not auditory fatigue), but the subjects were asked how they liked the test in comparison with the word test. Many of them liked the idea of the subject testing himself and found it interesting and stimulating. On the whole the test was accepted with an unusual amount of enthusiasm. No definite conclusions can be made concerning the factor of fatigue, and it is probably of little importance. The fact that the test is less time-consuming and more interesting than many articulation tests makes it safe to assume that it is to some degree less fatiguing.

CHAPTER VI

CONCLUSIONS

1. Normal and hard-of-hearing individuals can reliably select a threshold of intelligibility for monitored connected discourse. This is demonstrated by the narrow range of the settings made by both normal and hard-of-hearing listeners. The average range for 6 settings in one ear was 5.0 db for the normals and 4.55 db for the hard of hearing. 58% of the 6 settings of the normals and 61% of the 6 settings of the hard of hearing were within a range of 4 db.

Another point in favor of the reliability of the threshold of intelligibility for connected discourse (TICD) is the small difference occurring between the first TICD score and the second TICD score. The average difference was 1.88 db for the normals and 1.98 for the hard of hearing. The average difference between the score of the Harvard University Test #9 (control test) and the second #9 test was 2.08 db (normals). The average test-retest difference, then, of the TICD for both groups was slightly less than the average difference of scores of the #9 test.

2. There is a close relationship of the TICD to the threshold of intelligibility for spondaic words (Test #9). 92% of the TICD scores of the normals were within ± 6 db of an intensity level 1 db above the #9 score. 74% were

within ± 3 db of this level. 80% of the TICD scores of the hard of hearing were within ± 6 db of an intensity level 1 db above the #9 score. 58% were within ± 3 db of this level. This is interpreted as evidence for the validity of the TICD as a measure of "intelligibility."

3. The high face validity of the TICD test is a characteristic which gives it value as an actual test of the type of speech one listens to day by day. By means of the test, it is possible to determine on an intensity scale the degree of loudness a given individual requires in order for him to get along adequately in his social environment.

4. The judgments of the hard of hearing do not differ significantly from those of the normals. There were fewer extreme deviates among the normals, but the hard of hearing were more reliable in repeating their judgments of the threshold of intelligibility.

5. A threshold of intelligibility can, as a rule, be obtained more quickly from connected discourse than from ordinary articulation tests. The average time required for the first judgment is about 40 seconds. As the listener becomes experienced, the average approaches 20 seconds. This is quicker than the 4 to 5 minutes required for most articulation tests.

6. It became evident early in the study that the instructions to the subject were critical in producing a test of good validity. The presence of several related thresholds lying near the threshold of intelligibility may be confusing. It is important to give the listener some definite clues so that he can pick out the proper threshold quickly and accurately. More satisfying results were obtained when we instructed the listener to attenuate the speech quickly to the point where some of the words seemed to "drop out." Then by increasing the intensity slowly, he could bring those words in more distinctly and "hear everything the speaker was saying." These instructions brought the subjective threshold of the TICD test into more definite relationship with the objective threshold of the #9 test.

BIBLIOGRAPHY

Hudgins, C.V., J.E. Hawkins, J.E. Karlin, and S.S. Stevens, "The Development of Recorded Auditory Tests for Measuring Loss for Speech," Laryngoscope, 57: 57-89, Jan., 1947.

O.S.R.D. Report No. 3105, "Response Characteristics of Interphone Equipment," Jan., 1944.

O.S.R.D. Report No. 3802, "Articulation Testing Methods II," Nov., 1944.

O.S.R.D. Report No. 6303. "Tolerance for Pure Tones and Speech in Normal and Hard-of-Hearing Ears," July, 1946.

Steinberg, J.C., H.C. Montgomery, and J.B. Gardner, "Results of the World's Fair Hearing Tests," Bell System Technical Publication, Monograph B-1256, 1940.

Walsh, Theo. E., and S. Richard Silverman, "Diagnosis and Evaluation of Fenestration," Laryngoscope, 56: 536-555, Sept., 1946.

Witting, E.G., and W. Hughson, "Inherent Accuracy of a Series of Repeated Clinical Audiograms," Laryngoscope, 50: 259-269, March, 1940.

BIBLIOGRAPHY

- Rudgins, C.V., J.E. Hawkins, J.E. Karlin, and S. S. Stevens, "The Development of Recorded Auditory Tests for Measuring Loss for Speech," Laryngoscope, 57: 57-89, Jan., 1947.
- O.S.R.D. Report No. 3105, "Response Characteristics of Interphone Equipment," Jan., 1944.
- O.S.R.D. Report No. 3802, "Articulation Testing Methods II," Nov., 1944.
- O.S.R.D. Report No. 6303, "Tolerance for Pure Tones and Speech in Normal and Hard-of-Hearing Ears," July, 1946.
- Steinberg, J.C., H. C. Montgomery, and J. B. Gardner, "Results of the World's Fair Hearing Tests," Bell System Technical Publication, Monograph B-1256, 1940.
- Walsh, T.E., and S.R. Silverman, "Diagnosis and Evaluation of Fenestration," Laryngoscope, 56: 536-555, Sept., 1946.
- Witting, E.G., and W. Hughson, "Inherent Accuracy of a Series of Repeated Clinical Audiograms," Laryngoscope, 50: 259-269, March, 1940.