EFFECTS OF EXPERIMENTER BIAS UPON PURE-TONE AND SPEECH AUDIOMETRIC TEST RESULTS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY

Nicholas M. Hipskind

1968

This is to certify that the

thesis entitled

EFFECTS OF EXPERIMENTER BIAS UPON
PURE-TONE AND SPEECH TEST RESULTS

presented by

Nicholas M. Hipskind

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Audiology and Speech Sciences

William F. Rintelmann
Major professor

Date Nov. 7, 1968

O-169

ABSTRACT

EFFECTS OF EXPERIMENTER BIAS UPON PURE-TONE AND SPEECH AUDIOMETRIC TEST RESULTS

by Nicholas M. Hipskind

The purpose of this investigation was to determine whether certain types of previous audiometric information causes the tester to influence pure-tone and speech audiometric results. Specifically, this study was concerned with the possibility of testers' expectancy effects on audiometric responses obtained from normal hearing adults, hard-of-hearing adults having either conductive or sensorineural hearing impairments, and normal hearing children having either normal or defective articulation.

Two groups of testers were used that varied considerably in terms of amount of training in clinical audiology skills. One of the groups consisted of four predoctoral students majoring in audiology at Michigan State University and the other group was composed of four undergraduate students who were majoring in audiology and speech sciences at Michigan State University.

The study was composed of four experiments that were designed to study the phenomena of experimenter

ABSTRACT

EFFECTS OF EXPERIMENTER BIAS UPON PURE-TONE AND SPEECH AUDIOMETRIC TEST RESULTS

by Nicholas M. Hipskind

The purpose of this investigation was to determine whether certain types of previous audiometric information causes the tester to influence pure-tone and speech audiometric results. Specifically, this study was concerned with the possibility of testers' expectancy effects on audiometric responses obtained from normal hearing adults, hard-of-hearing adults having either conductive or sensorineural hearing impairments, and normal hearing children having either normal or defective articulation.

Two groups of testers were used that varied considerably in terms of amount of training in clinical audiology skills. One of the groups consisted of four predoctoral students majoring in audiology at Michigan State University and the other group was composed of four undergraduate students who were majoring in audiology and speech sciences at Michigan State University.

The study was composed of four experiments that were designed to study the phenomena of experimenter

(tester) bias and effect as related to pure-tone and speech reception threshold measurement and speech discrimination. The variables were as follows: age of subjects, audiometric tests employed, articulation of subjects, status of subjects' hearing mechanism, sophistication of testers, and types of previous audiometric information.

Experiment I

In this experiment four sophisticated clinical audiologists were employed as testers to study experimenter bias and effect upon pure-tone air-conduction thresholds and speech reception thresholds of eight normal hearing adults.

Experiment II

In this experiment the same four testers were utilized to explore the bias and effect phenomena as related to pure-tone air and bone-conduction thresholds and speech audiometry scores (speech reception thresholds and speech discrimination scores) of eight adults with sensorineural hearing disorders and eight adults with conductive hearing impairments.

Experiment III

In this experiment the same four testers were used to investigate the bias and effect phenomena as related to pure-tone air-conduction thresholds and speech audiometry scores (speech reception thresholds and speech discrimination scores) of sixteen normal hearing children. Half of the children had defective articulation and the remaining eight had normal articulation.

Experiment IV

In this experiment four unsophisticated testers were used to study experimenter bias and effect as related to pure-tone air-and bone-conduction thresholds of sixteen hard-of-hearing adults. Half of these subjects had a sensorineural hearing loss and half exhibited a conductive hearing disorder.

The clinical and statistical results of this study indicated that there were no significant differences between audiometric scores obtained by testers as a result of having either correct or erroneous previous test information. The results showed that each tester under each experimental condition (actual test results, better than actual test results, poorer than actual test results, and no test results) obtained audiometric results that were approximately the same. The results also indicated that previous audiometric information was no more infulential in eliciting an observable bias for speech audiometry than for pure-tone audiometry. Finally, the amount of sophistication of testers did not significantly affect the test results.

EFFECTS OF EXPERIMENTER BIAS UPON PURE-TONE
AND SPEECH AUDIOMETRIC TEST RESULTS

By

Nicholas M. Hipskind

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Sciences

1968

ACKNOWLEDGMENTS

This research was supported in part by a fellowship from the Vocational Rehabilitation Administration.

Special gratitude is given to Dr. William F. Rintelmann, Associate Professor of Audiology and Speech Sciences, for the interest he displayed and the guidance he gave as thesis advisor.

A very special thanks to my wife, Sandy, and daughter, Heidi, for their incomparable understanding, encouragement, and faith during the past three years.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page
XXI.	Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Articulation (C), and Frequency (D) for Sixteen Normal Hearing Children by Pure-Tone Air-Conduction Audiometry	163
XXII.	Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), and Articulation (C) for Sixteen Normal Hearing Children for Speech Reception Thresholds	164
XXIII.	Summary of Analysis of Variance Comparing Differences of Tester (A), Conditions (B), and Articulation (C) for Sixteen Normal Hearing Children for Speech Discrimination Scores	165
XXIV.	Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Frequency (C), and Pathologies (D) for Sixteen Hard-of-Hearing Adults by Pure-Tone Air-Conduction Audiometry	166
XXV.	Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Frequency (C), and Pathologies (D) for Sixteen Hard-of-Hearing Adults by Pure-Tone Bone-Conduction Audiometry . . .	167

LIST OF FIGURES

Figure		Page

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

In the field of audiology, research has been conducted regarding reliability and validity of audiometric tests. Research on threshold reliability has been concerned with inherent stability of electronic equipment used in the measurement of auditory thresholds or the inherent stability of the human auditory threshold itself. There has been very little experimental evidence reported regarding the reliability and validity of speech audiometry. Also, there are no reported studies related to how the experimenter or audiologist may affect the reliability and validity of audiometric tests. A few articles, however, suggest that a certain amount of error may be attributed to preconceived expectations of the tester. Thus, according to the literature reviewed, there is a dearth of information concerning experimenter biases and effects on clinical audiometric results.

Purpose of the Study

The major purpose of this investigation was to determine whether certain types of previous audiometric information cause the tester to influence obtained pure-tone and speech audiometric results. Specifically this study was concerned with the possibility of testers' expectancy effects on audiometric responses obtained from normal hearing adults, hard-of-hearing adults having either conductive or sensorineural hearing impairments, and normal hearing children having either normal or defective articulation.

Two groups of testers were used that varied considerably in terms of amount of training in clinical audiology skills. They were pre-doctoral students in audiology and undergraduate students who were taking an introductory course in audiology. The basic question posed prior to the investigation was as follows: does previous audiometric test information cause the tester to influence significantly audiometric scores as a function of the following variables:

1. degree of sophistication or amount of training of the tester
2. the auditory sensitivity of the subject (normal hearing, conductive hearing loss or sensorineural hearing loss)
3. pure-tone air-conduction versus bone-conduction thresholds
4. speech reception thresholds
5. speech discrimination scores
6. pure-tone threshold versus speech reception thresholds
7. defective articulation versus normal articulation on speech audiometric scores
8. adults versus children subjects

Based on the variables listed above the following research questions were posed.

1. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining pure-tone air-conduction thresholds from normal hearing adults?

2. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining pure-tone air-conduction thresholds from adults having a sensorineural hearing disorder?

3. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiolists when obtaining pure-tone air-conduction thresholds from adults having a conductive hearing impairment?

4. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining pure-tone air-conduction thresholds from normal hearing children?

5. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining pure-tone bone-conduction thresholds from adults having a sensorineural hearing disorder?

6. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining pure-tone bone-conduction thresholds from adults having a conductive hearing impairment?

7. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech reception thresholds from normal hearing adults?

8. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech reception thresholds from adults having a sensorineural hearing disorder?

9. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech reception thresholds from adults having a conductive hearing impairment?

10. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech reception thresholds from normal hearing children having normal speech articulation?

11. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech reception thresholds from normal hearing children having defective speech articulation?

12. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech discrimination scores from adults having a sensorineural hearing disorder?

13. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech discrimination scores from adults having a conductive hearing impairment?

14. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech discrimination scores from normal hearing children having normal speech articulation?

15. Does previous audiometric test information (true or erroneous) influence significantly sophisticated audiologists when obtaining speech discrimination scores from normal hearing children having defective speech articulation?

16. Does previous audiometric test information (true or erroneous) influence significantly unskilled testers when obtaining pure-tone air-conduction thresholds from adults having a sensorineural hearing disorder?

17. Does previous audiometric test information (true or erroneous) influence significantly unskilled testers when obtaining pure-tone air-conduction thresholds from adults having a conductive hearing impairment?

18. Does previous audiometric test information (true or erroneous) influence significantly unskilled testers obtaining pure-tone bone-conduction thresholds from adults having a sensorineural hearing disorder?

19. Does previous audiometric test information (true or erroneous) influence significantly unskilled testers obtaining pure-tone bone-conduction thresholds from adults having a conductive hearing impairment?

The primary intent of this research was to determine whether in a clinical audiometric evaluation previous audiometric information contributes significatnly to obtaining erroneous audiometric configurations and therefore, misdiagnosis. As stated above, the literature shows that researchers in audiology have ignored, experimentally, the influences of experimenter biases in recording audiometric test results. Several articles make reference to the fact that auditory scores are sometimes erroneously reported because of the otologist's past experiences, but these statements are not based upon experimental evidence. According to Sataloff[1], a serious mistake made when obtaining audiometric results is for the audiologist or otologist to have preconceived ideas of the patient's auditory sensitivity at various frequencies and conclude

---

[1]Joseph Sataloff, "Pitfalls in Routine Hearing Tests," Archives of Otolaryngology, 73 (1961), pp. 717-726.

that a specific audiometric configuration always means a specific pathology. A variety of variables such as physiological, psychological, and methodological are considered to be the major sources of error in audiometry.[1]

About forty years ago, psychologists began to investigate experimentally the personal effects that examiners had on the test scores obtained from their clients.[2] The literature reveals that studies on experimenter bias and experimenter effects have been conducted in many different disciplines. However, none of the reported investigations were conducted involving the sense of hearing nor have any of these studies been concerned with threshold measurement. As previously stated, however, studies have been designed in the field of audiology to determine the major sources of errors encountered in audiometry. Factors contributing to errors include faulty calibration of equipment, ambient noise, and the use of different psychophysical methods to obtain thresholds.

---

[1]Wallace S. High, Aram Glorig, and James Nixon, "Estimating the Reliability of Auditory Threshold Measurements," Journal of Auditory Research, 1 (1961), pp. 247-262.

[2]Robert Rosenthal, "Experimenter Attributes as Determinants of Subjects' Responses," Journal of Projective Techniques, 27 (1963), pp. 324-331.

In the interest of accurate audiometric assessment, it is important to know whether previous audiometric information does cause the audiologist to influence significantly the patients' audiometric scores. Because of the paucity of information in this area of audiology, this study was proposed.

Definition of Terms

The following terms have been used extensively in this investigation. For clarification and convenience, they are defined below.

Experimenter Bias.--Occurs when the experimenter obtains results from the subject that he expects to obtain.[1]

Experimenter Effect.--Occurs when different experimenters obtain different data from the same subjects.[2]

Normal Hearing.--For the purpose of this study, this term means that the individual's pure-tone air and bone-conduction thresholds are interweaving (± 5 dB) and are no poorer than 25 dB (ISO 1964 standard) for the octaves between 250 and 8000 Hz.

Conductive Hearing Loss.--For the purpose of this study, this term means that the hearing loss is due to an

---

[1]Neil Freidman, Daniel Kurland, and Robert Rosenthal, "Experimenter Behavior as an Unintended Determinant of Experimental Results," Journal of Projective Techniques, 29 (1965), pp. 479-490.

[2]Ibid.

impairment in the outer or middle ear or both in the presence of a normal inner ear. The condition is observed clinically when bone conduction thresholds are normal and air-conduction thresholds are depressed. A person was considered to have a conductive hearing loss whenever his bone-conduction thresholds for 250, 500, 1000, 2000, and 4000 Hz were at least 10 dB more sensitive than the same frequencies tested by air-conduction. Also, the sensitivity of the bone-conduction thresholds could be no poorer than 25 dB (ISO 1964 standard).

Sensorineural Hearing Loss.--For the purpose of this study, this term means that the hearing loss is attributed to malfunction of the inner ear. The condition is observed clinically when air and bone-conduction thresholds are interweaving (± 5 dB) and a hearing loss is present. A person was considered to have a sensorineural hearing loss when thresholds were poorer than 25 dB (ISO 1964 standard) for at least the frequencies of 2000, 4000, and 8000 Hz.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter is concerned with the reliability of pure-tone and speech audiometric thresholds and speech discrimination scores. It is also concerned with research conducted regarding the phenomena of experimenter bias and experimenter effect. Specifically, the chapter reveals the diversified areas in which these phenomena have been studied and the dearth of information existing in the area of audition.

Threshold Measurements

Man has been interested in the measurement of hearing sensitivity for the past four centuries.[1] However, this interest long preceded the development of standardized test methods and equipment that adequately and systematically measured auditory sensitivity.

Electronic equipment has been developed that is thought to measure precisely auditory thresholds via pure-tone and speech stimuli. Scientists have not,

---

[1]John J. O'Neill and Herbert J. Oyer, Applied Audiometry (New York: Dodd, Mead & Company, Inc., 1966), pp. 36-51.

however, developed an instrument to determine auditory thresholds without having a human relate instructions, manipulate dials, interpret results, or make recommendations. Even auditory thresholds obtained by so-called automatic audiometry cannot be isolated from some type of human influence. Thus, when attempting to define auditory thresholds, a number of variables, including the human variable, must be taken into account.

Dixon Ward[1] pointed out auditory threshold is not a simply defined phenomenon. According to Ward, the long accepted definition that threshold is a constant energy barrier (". . .if the signal energy exceeds this level, it will be perceived; if the energy is less than this critical value, it will not") is no longer adhered to by authorities of threshold measurement. Persons sophisticated in areas pertaining to auditory threshold measurement are well aware of the inadequacy of this definition. It is presently known that numerous variables must be controlled when attempting to define or measure the human auditory threshold. Some of these variables include: the spectrum and duration of the stimulus, the psychophysical method employed when conducting the measurement, practice effects, instructions given to the subject, the

---

[1]Dixon Ward, "Auditory Fatigue and Masking," ed. James Jerger, Modern Developments in Audiology (New York: Academic Press, 1963), pp. 242-243.

emotional state of the subject, age of the subject, listener fatique, medical history of the subject, transducer placement, calibration and stability of the equipment, sophistication of the examiner, and ambient noise in the test environment.

Authorities in audiology and otology are in agreement with Ward in believing that auditory thresholds cannot be defined as a static phenomenon that is only dependent on the stimulus being presented. Many of the above variables that influence auditory thresholds have been enumerated by High, Glorig, and Nixon.[1]

Sataloff[2] also has listed various pitfalls common in audiology and otology that influence auditory thresholds. One of the most consequential is that the audiologist or otologist or both have a biased expectancy of what a threshold ought to be and then consciously or unconsciously record and interpret thresholds as a function of these expectancies.

Rintelmann and Harford[3] explained that in Bekesy audiometry, where the listener traces his threshold

---

[1]High, Glorig, and Nixon, "Estimating Reliability of Thresholds," pp. 247-262.

[2]Sataloff, "Pitfalls in Routine Hearing Tests," pp. 717-726.

[3]William F. Rintelmann and Earl R. Harford, "Type V Bekesy Pattern: Interpretation and Clinical Utility," Journal of Speech and Hearing Research, 10 (1967), pp. 733-744.

independent of the tester presenting the stimulus, errors still arise in the classification of such thresholds. These authors pointed out explicitly how Bekesy type classification can be influenced significantly by erroneous interpretations of the separation between continuous and pulsed stimuli. Further, according to Stream and McConnell[1] different thresholds are obtained when slight differences in subject instruction are given for the adjustment of a pure-tone in automatic audiometry.

Even when the aforementioned variables are controlled there is still appreciable fluctuation in the human auditory threshold. Investigations have shown that at a given time a specific stimulus of determined magnitude may elicit a response, and at another moment this same stimulus with increased intensity will not be perceived by the same listener.[2] Menzel[3] has attempted to explain the fluctuation of auditory thresholds and the variability related to threshold measurement. He stated:

---

[1]Richard W. Stream and Freeman McConnell, "A Comparison of Two Methods of Administration in Bekesy-Type Audiometry," Journal of Auditory Research, 4 (1961), pp. 263-271.

[2]Ward, "Auditory Fatigue and Masking," pp. 242-243.

[3]Otto J. Menzel, "Error in Audiometry," Eye Ear Nose and Throat Monthly, 42 (1963), p. 74.

". . .in nearly all instances something other than the intended stimulus has determined the overt responses upon which the auditory measurement is based."

From the literature reviewed, it is quite evident that the auditory threshold must be defined according to measurement technique for each research project in which threshold values are obtained. This is particularly necessary when defining the extent to which the variability of measurement can deviate on repeated measures. In other words, threshold can be defined theoretically as a particular point with a ± 2 decibel of variability, or it can be defined as a specific point with a ± 5 dB margin of variability. Often, the amount of variability is dependent upon the size of the attenuator step (2 dB versus 5 dB) employed in the threshold measurement.

Reliability of Audiometric Tests

Whenever an experimenter or clinician administers a test or a series of tests he is usually interested in the validity and the reliability of the instrument or instruments he uses. The validity of a test is the degree to which the instrument measures what it purports to measure. The reliability of a test is how consistently the instrument measures a specific parameter.

Since the primary purpose of this investigation was to determine how the consistency of various audiometric tests was affected by a number of testers receiving different audiological information, it was necessary to determine the reliability of the tests used in this research project.

Jerger[1] defines reliability as applied to audiological tests in the following manner:

> Reliability is repeatability of test scores. Absolute consistency refers to the absolute variability in performance from test to retest. It is concerned with basic precision of measurement irrespective of the extent to which rank order is preserved from test to retest. It is defined by the standard error of measurement. It may be concretely visualized as the standard deviation of random error of measurement--symbolized S_e.

Following are the test-retest results obtained by various investigators concerning the audiometric stimuli used in this study.

Pure-Tone Air-Conduction Stimuli

In 1939 Witting and Hughson[2] conducted a study to determine the errors in pure-tone air-conduction thresholds of hard-of-hearing and normal hearing individuals. These authors found that on repeated threshold measures the smallest variability occurred at 1000 Hz with

---

[1] James F. Jerger, "Comparative Evaluation of Some Auditory Measures," Journal of Speech and Hearing Research, 5 (1962), pp. 3-17.

[2] E. G. Witting and Walter Hughson, "Inherent Accuracy of a Series of Repeated Clinical Audiograms," Laryngoscope, 50 (1940), pp. 259-269.

125 and 8000 Hz showing the greatest variability. The major findings of Witting and Hughson are enumerated below.

1. Hard-of-hearing individuals exhibit an inherent error of less than 5 dB for pure-tone air-conduction stimuli. The least variability occurred at 1000 Hz. Normal hearers were more consistent and exhibited less variability at each of the frequencies tested.

2. Several large deviations occurred during the testing that were not explained by the authors; however, all of the large deviations were found with the hard-of-hearing subjects.

3. Subjects tended to demonstrate better thresholds on repeated measures even when these repetitions were made two years after the initial testing.

4. Variations greater than reported by these authors should be considered erroneous and related to factors other than the listener's hearing sensitivity.

Carhart and Hayes[1] confirmed Witting and Hughson's findings that the standard error of measurement of air-conduction audiometry is a function of frequency. They found that 1000 Hz was the least variable frequency

---

[1]Raymond Carhart and Claude Hayes, "Clinical Reliability of Bone Conduction Audiometry," Laryngoscope, 50 (1940), pp. 1084-1101.

5.9 dB). These authors reported the following standard errors of measurement: 6.5 dB at 250 Hz, 5.9 dB for 1000 Hz, and 6.8 dB for 4000 Hz.

Jerger[1] found the absolute variability for air-conduction audiometry to be 5 dB at 250 Hz, 6 dB at 1000 Hz, and 4.7 dB at 4000 Hz. These variations are comparable to the findings of Carhart and Hayes.

In an attempt to standardize clinical procedures for obtaining pure-tone air and bone-conduction thresholds, Carhart and Jerger[2] found that when presenting pure-tone stimuli in an ascending manner, the test-retest repeatability for 250 Hz was 4 dB, for 1000 Hz was 4.6 dB, and for 4000 Hz was 7.2 dB.

Harris and Myers[3] found that the inherent stability of pure-tone air-conduction thresholds for normal hearers was less than ± 5 dB. These authors found that when using a 1 dB step attenuator, this variability or inherent stability were reduced from ± 5 dB to ± 2 dB.

---

[1] James Jerger, "Comparative Evaluation of Some Auditory Measures," pp. 3-17.

[2] Raymond Carhart and James Jerger, "Preferred Method of Determination of Thresholds," Journal of Speech and Hearing Disorders, 24 (1959), pp. 330-345.

[3] J. Donald Harris and Cecil Myers, "Experiments on Fluctuations of Auditory Acuity," Medical Research Laboratory U. S. Naval Submarine Base, New London, Connecticut, Volume XI No. 13 Report No. 196, 22 June 1952, Project NM 003 041.21.08.

Pure-Tone Bone-Conduction Stimuli

Carhart and Hayes[1] measured the absolute consistency of pure-tone bone-conduction stimuli and were interested in determining the relationship between air-conduction and bone-conduction threshold stability. They found the test-retest reliability for bone-conduction thresholds to be equally or more reliable than for air-conduction. They concluded that when bone-conduction audiometry is administered in an appropriate manner, its accuracy is as good as that of air-conduction audiometry. As was found for air-conduction stimuli, the least variability occurred at 1000 Hz (5.4 dB). This variability increased as the frequencies became higher. The authors reported absolute variability of 5.4 dB for 250 Hz and 6.6 dB for 4000 Hz.

Jerger[2] also concluded that test-retest reliability for bone-conduction audiometry is comparable to conventional air-conduction audiometry.

Spondaic Word Stimuli

Although audiologists have notions of test-retest consistency based upon empirical clinical evidence, there

---

[1]Carhart and Hayes, "Clinical Reliability of Bone Conduction Audiometry," pp. 1084-1101.

[2]Jerger, "Comparative Evaluation of Some Auditory Measures," pp. 3-17.

is a dearth of laboratory information on the test-retest consistency of speech audiometry.[1]

Research indicates that the average value for the pure-tone air-conduction thresholds of 500, 1000, and 2000 Hz correlate positively with speech reception thresholds obtained by using spondaic words.[2,3,4,5]

Hudgins and his associates[6] developed two lists of spondaic words which were named Auditory Test No. 9. The authors determined that the absolute test-retest

---

[1] Ibid.

[2] Raymond Carhart, "Speech Reception in Relation to Pattern of Pure-Tone Loss," Journal of Speech Disorders, 11 (1946), pp. 97-108.

[3] Raymond Carhart, "Individual Differences in Hearing for Speech," Annals of Otology, Rhinology, and Laryngology, 55 (1946), pp. 223-267.

[4] James F. Jerger, Raymond Carhart, and Tom W. Tillman, "Some Relations Between Normal Hearing for Pure Tones and for Speech," Journal of Speech and Hearing, 2 (1959), pp. 126-140.

[5] Harvey Fletcher, "A Method of Calculating Hearing Loss for Speech from an Audiogram," Journal of the Acoustical Society of America, 22 (1950), pp. 1-5.

[6] C. V. Hudgins, J. F. Hawkins, J. E. Karlin and S. S. Stevens, "The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech," Laryngoscope, 57 (1947), pp. 57-89.

reliability of these words was 2.8 dB. Hirsh and his associates[1] developed the CID Auditory Test W-1 (based on Auditory Test No. 9) to measure speech thresholds. However, these authors did not publish test-retest values for the W-1 test.

In a study comparing 2 and 5 dB methods of determining spondee thresholds (speech reception thresholds), Chaiklin and Ventry[2] found that when using a 2 dB-step attenuator, ninety-three percent of their twenty-seven subjects had test-retest differences from 0 to ± 6 dB. No subject had a test-retest difference greater than 8 dB.

Chaiklin and Ventry[3] reported that Barrett found that ninety-six percent of his normal hearing subjects repeated speech reception thresholds within ± 6 dB and that a joint study conducted by the National Institute of Health in cooperation with the Veterans Administration reported that ninety-three percent of their subjects had test-retest scores of ± 6 dB on spondee words.

---

[1]Ira Hirsh, Hallowell Davis, S. Richard Silverman, Elizabeth G. Reynolds, Elizabeth Eldert, and Robert Benson, "Development of Materials for Speech Audiometry," Journal of Speech and Hearing Disorders, 17 (1952), pp. 321-337.

[2]Joseph B. Chaiklin and Ira M. Ventry, "Spondee Threshold Measurement: A Comparison of 2- and 5- dB Methods," Journal of Speech and Hearing Disorders, 29 (1964), 47-59

[3]Joseph B. Chaiklin and Ira M. Ventry, "Functional Hearing Loss," ed. James Jerger, Modern Developments in Audiology (New York: Academic Press, 1963), pp. 76-125.

s

l

p

Sp
l

Af
Jou
146

Spe
(196

Test
Words
Aeros
Air F

The results of studies conducted by Rhum and Carhart[1] and Tillman and Jerger[2] have shown that test-retest variability should not exceed ± 6 dB for spondee words.

Speech Discrimination Stimuli

For the past twenty-five years monosyllabic words have been used to measure a listener's ability to discriminate speech. The first monosyllabic word lists were the twenty PB-50 word lists constructed at Harvard during World War II. From these original lists several monosyllabic, phonetically balanced word lists have been developed. All of these lists consist of fifty one-syllable words that are relatively familiar to most listeners.[3]

Tillman and Carhart[4] developed four lists of fifty phonetically balanced monosyllabic words, Northwestern

---

[1]Howard B. Rhum and Raymond Carhart, "Objective Speech Audiometry," Journal of Speech and Hearing Research, 1 (1958), pp. 169-178.

[2]Tom W. Tillman and James F. Jerger, "Some Factors Affecting the Spondee Threshold in Normal-Hearing Subjects," Journal of Speech and Hearing Research, 2 (1959), pp. 141-146.

[3]Raymond Carhart, "Problems in the Measurement of Speech Discrimination," Archives of Otolaryngology, 82 (1965), pp. 253-260.

[4]Tom W. Tillman and Raymond Carhart, "An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words Northwestern University Test No. 6, USAF School of Aerospace Medicine Areospace Medical Division (AFSC) Brooks Air Force Base, Texas (1966), pp. 1-12.

5

d

m

th

wh

Ha

Rel
Mon
Uni

of S
Mast

University Auditory Test No. 6, to measure speech discrimination. These lists were found to be equivalent and yielded good test-retest reliability. At a 24 dB sensation level the absolute variability of these words for normal hearers was 2% and it was 3.9% for individuals with sensorineural hearing impairments. In a study conducted by Rintelmann and Jetty[1] at Michigan State University it was found that the standard error of measurement for Northwestern University Auditory Test No. 6 was 3.58% when presented to normal hearers at a 24 dB sensation level.

Haskins[2] developed four lists of fifty phonetically balanced words for testing children. These were named PBK-50's. Haskins stated: ". . .the retest presentation used during the present study appears to be unsuitable as a means of estimating the reliability of the materials and the procedures employed." Nevertheless on PBK-List 1, which was the List used in the present investigation, Haskins reported test-retest values of 2.48%.

---

[1]William F. Rintelmann and Albert J. Jetty, Reliability of Speech Discrimination Testing using CNC Monosyllabic Words" (Unpublished Study, Michigan State University, 1968).

[2]Harriet Haskins, "A Phonetically Balanced Test of Speech Discrimination for Children," (Unpublished Master's Thesis, Northwestern University, 1949).

c

b

a

Di

Carhart[1] stated:

> In general as long as the test items are meaningful monosyllables for the patient and their phonetic distribution is appropriately diversified, one 50-word compilation is relatively equivalent to another.

Therefore, based on Carhart's statement, it is assumed that the Phonetically Balanced Kindergarten Word Lists have approximately the same test-retest repeatability as the Northwestern University Auditory Test No. 6.

Summary

It has been adequately established that for pure-tone air and bone-conduction thresholds, the test-retest repeatability should not exceed ± 5 dB for the octave frequencies between 250 and 8000 Hz. In other words, test-retest variability for conventional pure-tone audiometry should be no greater than ± 5 dB. When employing spondee words to establish speech reception thresholds, it has been found that test-retest variability should not exceed ± 6 dB. Also, there is a positive correlation between pure-tone air-conduction thresholds of 500, 1000, and 2000 Hz and speech reception thresholds.

---

[1]Carhart, "Problems in the Measurement of Speech Discrimination," p. 254.

pa

tr

to

in

res

the

the

Although more sophisticated research is needed concerning test-retest reliability of phonetically balanced monosyllabic words, it has been illustrated that these words have an absolute consistency on the order of ± 4%.

The Phenomena of Experimenter Bias and Experimenter Effect

The placebo effect in medical practice illustrates that humans can be readily influenced. Shapiro gave a detailed description of the placebo effect.

> This is the psychological, physiological, and psychophysiological effect of any medication or procedure given with therapeutic intent, which is independent of or minimally related to the pharmacological effect of the medication or to the specific effects of the procedure, and which operates through the psychological mechanism. Although the pharmacological effect of a drug may have been deleterious or of little consequence to the organism, its effect could have been beneficial nevertheless.[1]

It can be said that the placebo effect is the phenomenon that causes the patient to respond in a positive manner physiologically, if psychologically he expects to respond in a positive manner.

In the late 1920's psychologists became interested in obtaining information about how they were affecting the responses of their patients. They were also concerned with the possibility that a different psychologist could obtain

---

[1] Arthur Shapiro, "A Contribution to a History of the Placebo Effect," Behavioral Science, 5 (1960), p. 109.

I

E

F

i

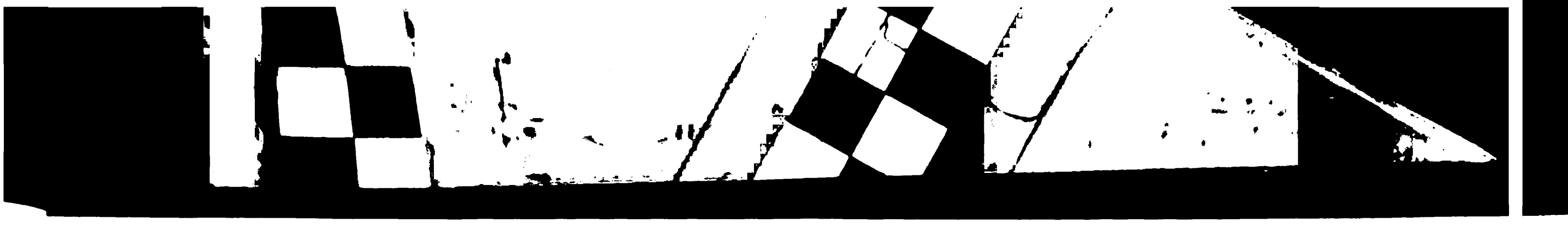
t

an

Det

hav

196



different results from the same patient doing the same task.[1] From the psychologists' interest, research projects were designed that experimentally investigated the phenomena of experimenter bias and experimenter effect. The degree to which the tester is influenced by previous test information has been labeled experimenter bias. The degree to which the tester influences a person's test results has been titled experimenter effect. In other words, an experimenter bias occurs when experimenters obtain scores from their subjects that they expect to obtain. Experimenter effect occurs when different experimenters record different scores from the same subjects performing essentially the same tasks.[2]

Since the first experiments exploring experimenter bias and effect, numerous investigations concerning these phenomena have been reported in several areas. Robert Rosenthal, an experimental psychologist at Harvard University, recently published a book that is a compilation of the research exploring these two phenomena.[3]

---

[1]Rosenthal, "Experimenter Attributes as Determinants of Subjects' Responses," p. 324.

[2]Freidman, Kurland, and Rosenthal, "Unintended Determinant of Experimental Results," pp. 479.

[3]Robert Rosenthal, Experimenter Effects in Behavioral Research (New York: Appleton-Century-Crofts, 1966).

Even though experimenter influences were not formally reported prior to 1929, people were aware that many observations and decisions were directly influenced by the observers' biases or expectancies or both. As early as the eighteenth century it was determined that humans observing the movement of the stars disagreed violently on the velocity at which stars moved. It was later discovered that these differences in observations were not intentional errors but rather were beyond the control of the observer and were classified as "personal equations."[1] In other words, if an astronomer had been schooled to measure the velocity of the stars in a particular manner using a specific point of reference, in all probability his data would differ significantly from another astronomer who was measuring the speed of the same star using a somewhat different technique. The reason for these differences in data might be attributed to differences in training and the use of different points of reference. Although both astronomers were probably unaware of it, their recorded data were being influenced by their own biases and expectancies relating to the speed at which a particular star should travel.

Experimenter biases and effects have been observed as an everyday occurrence in society as well as in the

---

[1] *Ibid.*, p. 3.

E

(C

Ex
He
(1

sciences. It has been illustrated that frequently man is able to gain acceptance in social activities if the group feels he can contribute to their ultimate goals. For example, if the group expects that a certain individual can perform well enough to enhance their status, this expectancy may directly influence how the individual performs. Whyte's study[1] of social groups found that a man's ability to bowl was continuously being evaluated; therefore, the man's ability to bowl was directly responsible for the position he earned in the group. If the group expected the man to bowl adequately or inadequately, they partially influenced the man's actual performance.

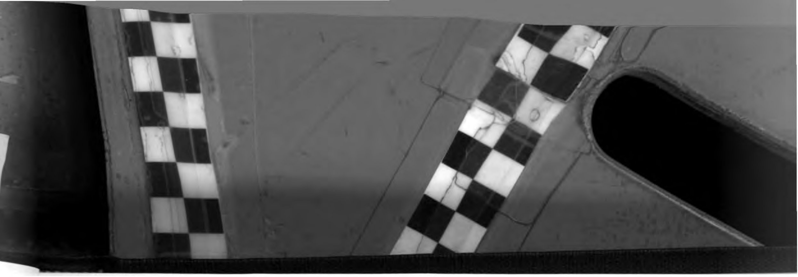
Following are several studies that have been reported illustrating the phenomena of experimenter bias and experimenter effect. These examples were selected from many sources in order to demonstrate the existence of these phenomena in a variety of situations.

In 1936 it was found that personnel highly skilled in agriculture were not capable of making selections of plant life without exhibiting a definite bias.[2]

---

[1]William Foote Whyte, Street Corner Society (Chicago: University of Chicago Press, 1943), pp. 18-23.

[2]William G. Cochran and David J. Watson, "An Experiment on Observer's Bias in the Selection of Shoot-Heights," Empire Journal of Experimental Agriculture, 4 (1936), pp. 69-76.



It has been found that interviewers can alter survey results by reflecting their expectancies during the interview. The following factors have been discovered to be the most influential in affecting survey results: (1) interviewer resistance in stating a question, an inclination to reword the question or assume the answer or both; (2) relatively high ambiguity, subjectivity or complexity in the concept or wording of the inquiry; and (3) supplementing the questions with additional cues that will elicit an expected response from the interviewee.[1] Harvey[2] reported that his experiment in interviewer biases adequately illustrates that interviewers give sufficient cues to evoke expected responses. Blakenship[3] agreed with Harvey; however, he attributed the cues to dishonesty rather than a genuine unconscious bias.

---

[1]Robert H. Hanson and Eli S. Marks, "Influence of the Interviewer on the Accuracy of Survey Results," Journal of the American Statistical Association, 53 (1958), pp. 635-655.

[2]S. M. Harvey, "A Preliminary Investigation of the Interview," British Journal of Psychology, 28 (1938), pp. 263-287.

[3]Albert B. Blakenship, "The Effect of the Interviewer Upon the Response in a Public Opinion Poll," Journal of Consulting Psychology, 4 (1960).

A detailed and controlled study was conducted by Back[1] on the importance of being accepted by the group. The differences among ways in which cohesiveness was achieved enabled the author to observe how communication was employed to "sway" the group. It was found that if cohesiveness of the group was determined by personal attraction, the conversation would evolve around a lengthy and pleasant topic. Each member of the group assumed the role of making the conversation a personal effort. Conversation became succinct and quite efficient when cohesiveness was believed to be based on task performance. All conversation that was seemingly unimportant was terminated and the discussion centered around a theme that was felt to achieve the task. When the group believed that cohesiveness was being determined by prestige, the conversation was altered to the extent that there was little or no risk taken that could possibly hinder cohesiveness. The group conversed cautiously and each member adjusted his own actions to conform to the environment established by the group. It is evident from this investigation that the group's behavior was significantly governed by what each member felt was expected of him.

---

[1] Kurt W. Back, "Influence Through Social Communication," Journal of Abnormal and Social Psychology, 46 (1951), pp. 9-23.

t

a

z

i

to

pa

ma

str

ser

the

Rese

To further illustrate the "real" existence of experimenter bias and experimenter effect in research, several investigations will be described in detail.

Rosenthal and Fode[1] conducted two replications of an experiment concerning a person perception task and found some interesting and revealing results. The authors presented fifty-seven photographs that showed the faces of fifty-seven different individuals to 104 students enrolled in an introductory psychology course. The students were asked to rate the faces in the pictures as to how successful they felt the person was. A rating scale ranging from a -10 to +10 was used to judge the degree of failure or success represented by each of the faces in the pictures. From the original fifty-seven photos, the authors selected twenty photos that had been rated as zero by the 104 students. These twenty photos were used in subsequent experiments. In the original experiment a total of ten students who were psychology majors participated as experimenters. This group consisted of eight males and two females. Two-hundred and six undergraduate students enrolled in an introductory psychology course served as subjects. All of the experimenters were given the same instructions; however, half the experimenters

[1]Rosenthal, Experimenter Effects in Behavioral Research, pp. 143-157.

were given ten of the twenty neutral photos and told that their subjects should average a +5 rating of these photos. The remaining experimenters were given the other half of the twenty neutral photos and told that their subjects should average a -5 rating of the photos. Prior to the subjects' ratings of the photos, each experimenter rated his ten photos using the same twenty point scale. The experimenters were given specific instructions to be read to their subjects. The results showed that difference mean ratings obtained by the experimenter expecting success (+5) scores and those biased toward failure (-5) ratings were significant at the .007 level of confidence. Without exception the experimenters expecting success ratings obtained higher ratings than did their counterparts expecting failure ratings. The mean scores obtained by the female experimenters did not differ significantly from the mean scores recorded by the male experimenters. According to the authors, there was no relationship between mean photo ratings obtained from subjects or the magnitude of the biasing phenomenon and the academic success of the experimenters.

Because the results of the aforementioned experiment were so "clean" and significant, Fode was compelled to replicate it. He employed twelve male students who were enrolled in an industrial psychology course as experimenters. The subjects were eighty-six students

I
s
c

or
in
Th
st
The
sel
per
tial
sign

enrolled in an introductory course in psychology. Basically Fode followed the same procedures as were presented in the original experiment; however, the experimenters did not manipulate the photos. The twenty neutral photos, divided into two groups of ten, were presented on a large poster and labeled so that the subjects could express their ratings of each photo to the experimenter. This was done in order to control for the possibility that the experimenters were relating their biases to their subjects by the way they presented the photos. Again, the results were significant, this time at the 0.0003 level of confidence. As was reported in the original experiment, all experimenters that expected successful scores obtained higher ratings than did any of their counterparts who were biased toward lower scores.

Fode followed his first replication with a second one. In this experiment eight students from an advanced industrial psychology class participated as experimenters. The ninety subjects used in this investigation were also students enrolled in an introductory psychology course. The author used the Taylor Scale of Manifest Anxiety to select his experimenters. It was determined that all experimenters were medium anxious. The results were essentially the same as the two aforementioned studies with significance at the 0.0005 level.

The authors concluded that the subjects observed behavioral cues exhibited by the experimenters. These cues were significant in influencing the response of each subject even though the experimenters read identical instructions to each subject.

In a similar type of experiment, sound motion pictures were filmed of the experimenters during the experiment. This procedure was followed in an attempt to determine how the experimenters revealed their biases to their subjects. The film revealed that experimenters exhibited different behavioral patterns when conducting an experiment. These behavioral patterns were significantly influenced by the experimenter's expectancies. Those experimenters whose conduct revealed a more personal attitude and warmth obtained ratings of the photos as more successful; on the other hand, the experimenters who conducted the experiment in a more professional manner and exhibited more professional competence obtained ratings that were judged to be failures.[1]

There is evidence that experimenter's behavior change whenever they believe that they are to obtain different results. Two experiments illustrate that

---

[1]Freidman, Kurland and Rosenthal, "Unintended Determinant of Experimental Results," pp. 490.

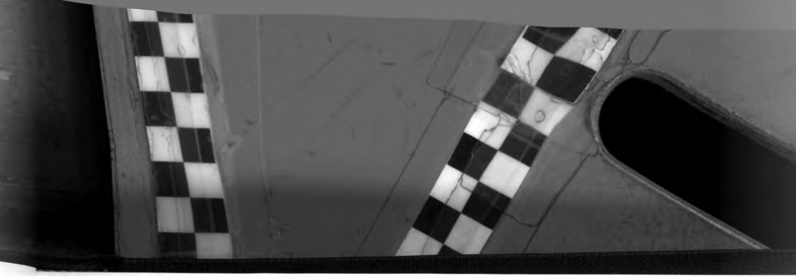
b

u

Pe
Hy
Jo

Gr
mer
sea
(19

atic
Abno



experimenter biases can be altered within an experiment. The results of these two experiments demonstrate that experimenters obtain responses from their students that they expect to obtain. The data also shows that experimenters alter their hypotheses in mid-experiment and then record results that are in direct agreement with their revised expectancies.[1]

There are several experiments reported that investigate experimenter bias and effect utilizing a verbal conditioning task. Rosenthal and associates[2] reported that in their experiment, verbal conditioning occurred as a function of experimenter expectancy. Again, it was reported that experimenters exhibit specific behavioral patterns that they believe will gain responses in accord with their expectancies.

Sarason[3] reported that the more hostility shown by the experimenter, the more the subjects are likely to use hostile verbs when developing sentences. It can

---

[1]Robert Rosenthal, Ray C. Murly, Gordon W. Persinger and Linda Vikan-Kline, "Changes in Experimental Hypotheses as Determinants of Experimental Results," Journal of Projective Techniques, 28 (1964), pp. 465-469.

[2]Robert Rosenthal, Paul Kohn, Patricia M. Greenfield and Noel Carota, "Data Desirability, Experimenter Expectancy, and the Results of Psychological Research," Journal of Personality and Social Psychology, 3 (1966), pp. 20-27.

[3]Irwin G. Sarason, "Individual Differences, Situational Variables, and Personality Research," Journal of Abnormal and Social Psychology, 23 (1959) 336-341.

generally be stated that most research exploring experimenter biases using verbal conditioning has discovered that the experimenter's need for approval (anxiety) does not validly predict experimenter's behavior during the experiment.

In an investigation conducted by Masling,[1] it was reported that examiners tended to be more lenient in scoring answers to various questions selected from the Wechsler Bellevue II intelligence test when the subjects were warm and friendly than when they were cold and indifferent. The examiners also gave the warm subjects more verbal reinforcement and additional chances to respond and document their answers. These examiner characteristics vanished when administering the test to "cold" subjects.

An experiment was conducted by Lord[2] to determine the effects of the experimenter's warm, cool, and natural behavior when administering the Rorschach. She found that subjects who were given the examination by a warm experimenter produced answers that were more detailed and abstract than when they were given the test by a cold

---

[1]Joseph Masling, "The Effects of Warm and Cold Interaction on the Administration and Scoring of an Intelligence Test," Journal of Consulting Psychology, 23 (1959), 336-341.

[2]Edith Lord, "Experimentally Induced Variations in Rorschach Performance," Psychological Monographs, 64 (1961), pp. 321-333.

[illegible]

examiner. Bellak[1] found essentially the same thing when using a story-telling project as the stimulus. Luft[2] also explored the influence of warm and cold behavioral characteristics contact with these types of behavior. He discovered that when the experimenter played the role of a warm examiner, positive responses were obtained. Negative responses were obtained when the experimenter played the role of a cold examiner.

All of the studies reported above employed human subjects. However, investigations employing animal subjects also have been conducted. Two classical experiments are reported in this regard.

In 1963 Rosenthal and Fode[3] reported significant data that revealed experimenter bias and experimenter effect in the conditioning of rats. Twelve students in the field of experimental psychology were employed as experimenters. The experimenters were given identical instructions, except that six of them were made to believe that their rats were "Maze-Bright;" the remaining

---

[1] Leopold Bellak, "The Concept of Projection: An Experimental Investigation and Study of the Concept," Psychiatry, 7 (1944), pp. 353-370.

[2] Joseph Luft "Interaction and Projection," Journal of Projective Techniques, 17 (1953), pp. 489-492.

[3] Robert Rosenthal and Kermit L. Fode, "The Effects of Experimenter Bias on the Performance of the Albino Rat," Behavioral Science, 8 (1963), pp. 183-189.

1

e

a

a

ta

Th

gro

All

pri

rat

expe

ment

subj

in pa

that

who f

experimenters were told that their rats were "Maze-Dull." In reality, all of the rats were essentially the same. The experimenters assigned to the "Maze-Bright" rats were instructed that they should definitely detect some type of learning from their animals during the first day. Also, their rats should exhibit a uniform learning rate during the experiment. The experimenters observing the "Maze-Dull" rats were told that these rats should not be expected to show any learning during the first day and very little, if any, during the entire experiment. The experimenters utilized a T-maze, which had one white arm and one gray arm. The maze was so constructed that the arms could be interchanged.

All of the animals were naive to conditioning tasks and were classified as Sprague-Dawley Albino rats. The rats were arranged into thirteen groups and each group was composed of two male and three female animals. All rats had been deprived of food for twenty-four hours prior to the experiment. Each experimenter was asked to rate his feelings about working with rats prior to the experimenter. This was done because none of the experimenters had previous experience working with animal subjects. The authors assigned the experimenters to work in pairs. Each pair consisted of a person who related that he felt he would enjoy working with rats and a person who felt he would dislike working with rats. One member

T

e

a

da

"Ma

exp

fou

ifi

impr

lear

impr

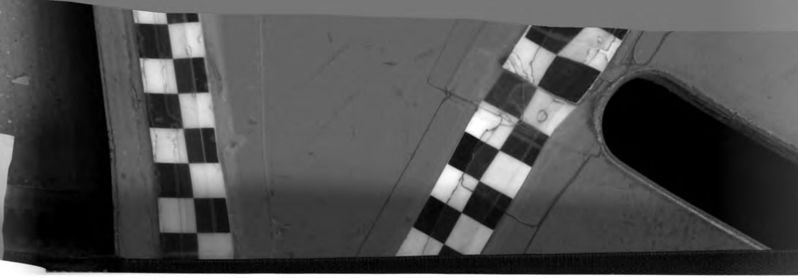
show

of each pair was randomly assigned to a group of rats designated as being "bright," and one member was to observe a group of "dull" rats.

Immediately before each experimenter was to test his rats, he was asked to evaluate on a 20-point scale how he felt his rats would perform. He was to assign a +10 if he felt his rats would perform extremely well and a -10 if he felt his rats would respond poorly.

The experimenters tested each of their rats ten times a day for five days. Two observations were made during each test; timing the animal to complete the run and noting if the run had been made in a correct or incorrect manner. The gray arm always contained the reward. This arm was alternated between the right and left sides equally; however, it was determined randomly when the arms would be changed. Thus, the alterations varied from day to day.

It was found that the rats who were labeled as "Maze-Bright" and who were believed to be brighter by the experimenters exhibited better performances on the first, fourth, and fifth days than did the rats who were classified as "dull." The "bright" rats showed systematic improvement in maze running as would be expected if learning was occurring. The "Maze-Dull" animals exhibited improvement at running the maze up to the third day but showed signs of regression on the fourth day and no change



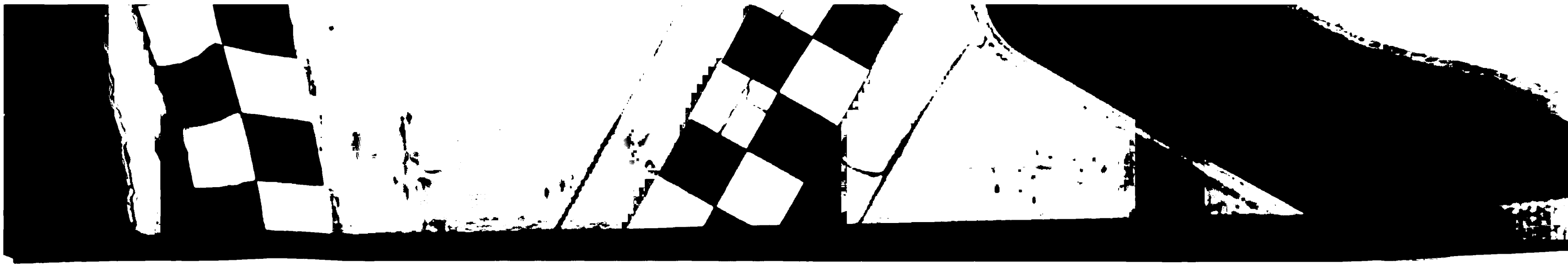
on the fifth day. It was illustrated that the rats being observed by experimenters who felt that they were "dull" had much less chance to learn the conditioning task than did the animals who were thought to be "bright." The "dull" rats were recorded as being much slower in running the maze correctly than were the "bright" rats. The "Maze-Bright" animals exhibited improvement that was more uniform than did the "Maze-Dull" rats.

At the termination of the experiment, the experimenters expressed their feelings about how they reacted during their rats' performance. Three-fourths of the experimenters reported that they felt good when their rats performed correctly and were disappointed when their animals' performance was inferior.

Another classical study conducted by Corado and Ison[1] exploring experimenter bias and effect, employed planaria (flat worms) as subjects. Seven experimenters were assigned to record the head turns and body contractions of planaria. The experimenters were told that half the planaria were hyperactive and half were hypoactive. Actually, all of the worms were essentially the same. This experiment revealed that experimenters reported twice as many head turns and three times as many body

---

[1] Lucion Corado and James R. Ison, "Observer Bias in Conditioning of Planaria," Psychological Reports, 13 (1963), pp. 787-789.



contractions when observing worms that were labeled "hyperactive," as when recording the movements of the "hypoactive"ones.

This investigation was replicated by the same authors using ten different experimenters.[1] Half of the experimenters recorded the head turns and body contractions of the "hypoactive" planaria and the other five observers recorded the head turns and body contractions of the "hyperactive" worms.

The results were complimentary to the findings of the original experiment; however, they were more dramatic. The group observing the "hyperactive" worms reported five times as many head turns and twenty times as many body contractions as did the group who were recording the movements of the hypoactive planaria.

Rosenthal and Lawson[2] conducted an investigation that used rats as subjects to illustrate operant learning. Again, it was determined that experimenters' beliefs and expectations about the performance of the animals were responsible in part for the data obtained.

---

[1] Ibid.

[2] Robert Rosenthal and Reed Lawson, "A Longitudinal Study of the Effects of Experimenter Bias on the Operant Learning of Laboratory Rats," Journal of Psychiatric Research, 2 (1964), pp. 61-72.

:

t

r

su

me

me

are

out

hum

dea

the

por

tes

mea

seen

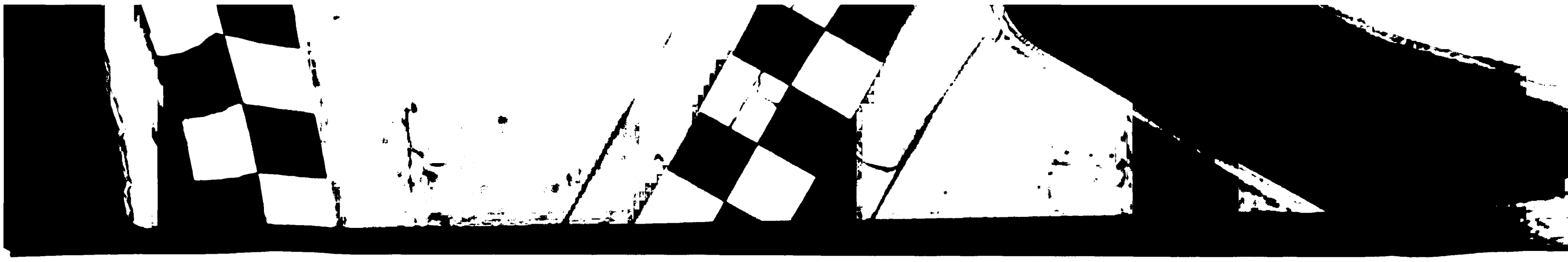
are

assu

---

of

pp.



Although this review of the literature is not exhaustive, it does demonstrate that the phenomena of experimenter bias and experimenter effect do occur in behavioral research. These studies illustrate how an experimenter can evoke or observe responses from his subjects that are in accord with his expectancies. It has been discovered through sophisticated and well-controlled research projects that previous information about the subjects governs significantly the behavior of the experimenters. Thus, it has been generally found that experimenters tend to obtain scores from their subjects that are highly correlated with their expectancies.

Authorities in audiology and otology are aware of outside influences that contribute to the measurement of human auditory thresholds. Nevertheless, there is a dearth of information concerning how the tester influences the listener's thresholds. There is no information reported in the literature expressing the significance of testers having previous audiometric results of threshold measurements. Authorities in audiology and otology have seemingly assumed that tester bias and effect exist and are significant in auditory threshold measurements. This assumption has been made by High, Glorig, Nixon,[1] Sataloff,[2]

---

[1]High, Glorig, and Nixon, "Estimating Reliability of Thresholds," pp. 247-262.

[2]Sataloff, "Pitfalls in Routine Hearing Tests," pp. 717-726.

and Ward;[1] however, none of these authors have documented their statements with research that supports this assumption.

Chaiklin and Vrntry[2] reported that they used a control to eliminate the tester bias effect. They stated: ". . .the measurement process and the resulting thresholds may be biased by the clinician's tendency to estimate spondee threshold from pure-tone averages." These authors believe that if the clinician adheres to stringent methodology, he will eliminate the bias effect. Again, the authors did not cite evidence that reveals the existence of the bias effect in auditory threshold measurements. In a dissertation by Barrett,[3] it was reported that the author controlled for tester bias by employing a Random Variable Attenuator; however, the writer did not report that the bias effect had been shown to contribute significantly to auditory measurements obtained.

If tester bias and tester effect are truly significant variables in auditory threshold measurements, they should be investigated under controlled conditions.

---

[1] Ward, "Auditory Fatigue and Masking," pp. 242-243.

[2] Joseph B. Chaiklin and Ira M. Ventry, "Spondee Threshold Measurement: A Comparison of 2 and 5-dB Methods," pp. 54-55.

[3] Lyman S. Barrett, "Threshold Relationships in Simulated Hearing Loss" (Unpublished Doctor's Thesis, Stanford University, 1959).

re
ha
cl
co
th
me
res
cor
mad
are
bei

publ
ment
attr
rese
will
had g
and i
worki
in te
the p

as Det
jectiv

According to Rosenthal,[1] most of the research reported to date concerning experimenter bias and effect has been conducted in the areas of survey research and clinical psychology and very few investigations have been conducted in a laboratory setting. It was the purpose of the present study to investigate the phenomena of experimenter bias and experimenter effect upon audiometric test results obtained by student testers under a variety of conditions. Since auditory threshold measurements are made using electronic equipment and quantified results are recorded, these measurements can be considered as being conducted in a laboratory setting.

The reasons given by Rosenthal for the lack of published research to determine the phenomena of experimenter bias and effect in a laboratory setting may be attributed to the following: (1) in survey or clinical research it is more feasible to construct conditions that will produce biases and effects; (2) investigators have had greater opportunity to conduct research in the field and in the clinic; and (3) the behavior of the person working in the laboratory can be more highly structured in terms of recording responses than can the behavior of the person performing the experiment in the field or in

---

[1]Robert Rosenthal, "Experimenter Modeling Effects as Determinants of Subjects' Responses," Journal of Projective Techniques, 27 (1963), pp. 467-471.

i

b

la

th

so

Ho

the

res

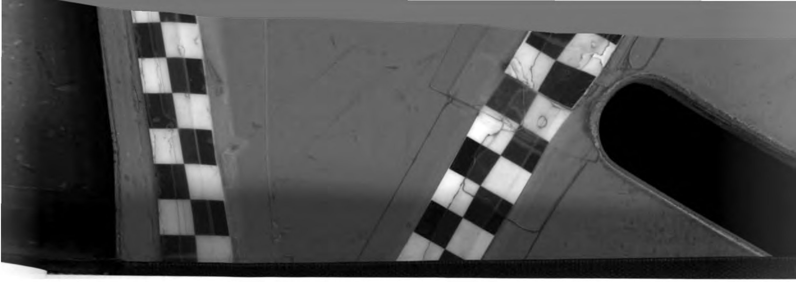
Researc

in the clinic.[1] Thus, it would appear that experimenter biases and effects would be less likely to occur in a laboratory setting. Nevertheless, Rosenthal feels that there is no reason why these phenomena cannot occur in some circumstances utilizing a laboratory setting. However, he stated ". . .for a full understanding of how these effects operate, we must wait for the results of research perhaps not yet begun."[2]

---

[1] Ibid.

[2] Rosenthal, Experimenter Effects in Behavioral Research, p. 119.



CHAPTER III

PROCEDURES

This investigation was composed of four individual experiments that were designed to study the phenomena of experimenter (tester) bias and effect as related to pure-tone and speech reception threshold measurement and speech discrimination. The variables were as follows: age of subjects, audiometric tests employed, articulation of subjects, status of subjects' hearing mechanism, sophistication of testers, and types of audiometric information (conditions).

This chapter is organized so that the experimental conditions, testers, subjects and tests administered are described for each experiment and the equipment and procedures employed are explained for the overall study. However, it is felt that a brief overview of the entire research project is appropriate at this time.

Overview of Experiments

There were two groups of testers with four testers in each group. One group of testers was composed of pre-doctoral students in audiology and the second group of testers consisted of undergraduate audiology and speech science majors who were enrolled in an introductory audiology course. The subjects employed in this study consisted of both adults and children. The adults were selected on the basis of their hearing sensitivity and audiometric configurations. The children were selected on the basis of their hearing sensitivity and their ability to articulate English phonemes.

The conditions, the testers, and the subjects are described immediately below for the overall study.

The following paragraphs briefly describe the four experiments in this study.

Experiment I

In this experiment four sophisticated clinical audiologists were employed as testers to study experimenter bias and effect upon pure-tone air-conduction thresholds and speech reception thresholds of eight normal hearing adults.

Experiment II

In this experiment the same four testers were utilized to explore the bias and effect phenomena as related to pure-tone air and bone-conduction thresholds and speech audiometry scores, speech reception thresholds and speech discrimination scores of eight adults with sensorineural hearing disorders and eight adults with conductive hearing impairments.

Experiment III

In this experiment the same four testers were used to investigate the bias and effect phenomena as related to pure-tone air-conduction thresholds and speech audiometry scores (speech reception thresholds and speech discrimination scores) of sixteen normal hearing children. Half of the children had defective articulation and the remainder had normal articulation.

Experiment IV

In this experiment four unsophisticated testers were used to study experimenter bias and effect as related to pure-tone air and bone-conduction thresholds of sixteen hard-of-hearing adults. Half of these subjects had a sensorineural hearing loss and the remaining eight exhibited a conductive hearing disorder.

The variables listed above and in Chapter I were investigated by studying the outcome of the four experiments and the relations between these experiments.

Experimental Conditions

To determine the extent to which the testers influenced subjects' pure-tone air and bone-conduction thresholds and speech audiometry scores, four types of audiological conditions were constructed concerning the type of information given to the testers at the beginning

of the test session as related to the subjects' hearing sensitivity and speech discrimination. The conditions were as follows: actual test results obtained previously, better than actual test results, poorer than actual test results, and no test results.

Actual test results were the subjects' pure-tone thresholds and speech audiometric scores recorded from the original audiograms obtained by the investigator of the present study.

Better test results were the subjects' pure-tone and speech audiometric scores recorded by improving the thresholds by 10 dB with respect to their original audiograms. In other words, the pure-tone audiograms were recorded as 10 dB better at each frequency tested by air and bone-conduction. Also, speech audiometric scores were recorded by improving the speech reception thresholds by 10 dB and increasing the speech discrimination scores by ten percent. However, when a subject had a speech discrimination score of 92% or better, the experimental condition of better than actual test results could not be fully employed, because it was not possible to make this score better by 10 percent. Therefore, when a subject had a speech discrimination score as just described, it was made better by the maximum percentage. In other words, speech discrimination scores ranging from 98% to 90% were changed to 100% for the condition of better than actual test results.

Poorer test results were the subjects' pure-tone and speech reception thresholds that were recorded poorer by 10 dB with respect to their original audiograms. Also, speech discrimination scores were recorded ten percent poorer than those originally obtained.

No test results mean that the subjects' original pure-tone and speech audiometric scores were not shown to the testers. In other words, in this condition the testers were not provided with any previous audiometric data.

The following figure and table illustrate how each of the conditions were recorded. All conditions were recorded with respect to the subjtects' original audiograms. These illustrations are representative of the test results the testers received prior to testing each subject.

Figure 1 shows the right ear air-conduction thresholds for one subject according to the conditions of actual, better, and poorer test results. The same procedure was used for recording bone-conduction thresholds. The audiograms given to the testers at the beginning of each test session showed right and left ear air and bone-conduction thresholds plotted according to one of the three conditions: actual, better, or poorer.

Table I is an example of the three possible conditions that could be given to the tester at the beginning of the test session.

FIGURE 1.--Illustrative Audiogram containing the conditions of (A) actual test results, (B) better test results, and (P) poorer test results for the right ear air-conduction thresholds for one subject.

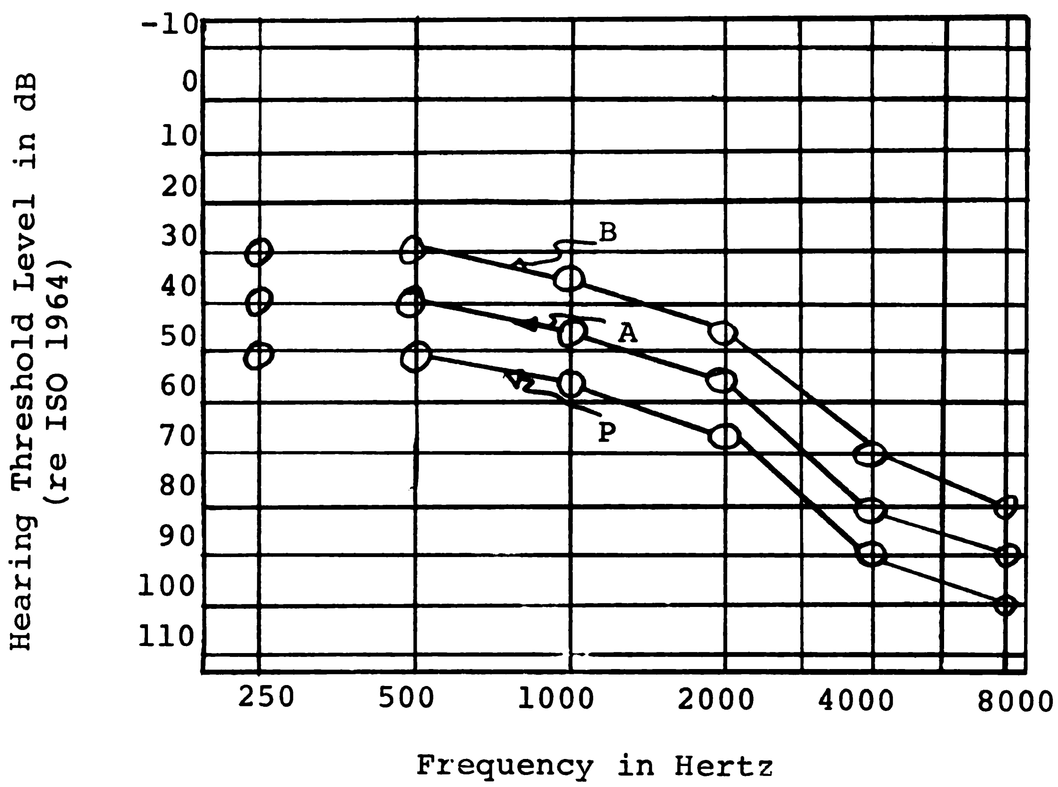


TABLE I.--Speech audiometric results with respect to the conditions of actual, better, and poorer test results.

Conditions	Actual	Better	Poorer
Speech Reception Threshold	30 dB	20 dB	40 dB
PB Score	80%	90%	70%

Testers

The testers were divided into two groups according to their level of sophistication or amount of training

and experience in audiology. One of the groups consisted of four pre-doctoral students majoring in audiology at Michigan State University. They were all males ranging in age from twenty-three to thirty-five years with a mean age of twenty-eight years. Each of these students had successfully completed the Master's Degree in the area of audiology and was judged to be a sophisticated clinical audiologist by three professors of audiology at Michigan State University.

The second group was composed of four undergraduate female students who were majoring in audiology and speech sciences at Michigan State University. Each of these students was enrolled in an introductory audiology course. They ranged in age from nineteen to twenty-one years with a mean age of 20.2 years. Each of them had administered only four pure-tone air and bone-conduction audiometric tests prior to participating in this investigation. Each was observed by the investigator of this study while administering four practice audiometric tests prior to participating in this study and were judged to be sufficiently competent in basic skills necessary for participating in the experiment. Thus, these testers had minimal training as opposed to the four graduate students who were experienced audiologists.

Subjects

All subjects were contacted by the investigator, first by mail and then by telephone, prior to participating in the present investigation. A copy of the letter sent to the subjects is in Appendix I. Each subject volunteered to participate and received no monetary reward for his participation.

Both adults and children were employed as subjects. The adult subjects consisted of eight normal hearing individuals, sixteen individuals with sensorineural hearing disorders, and sixteen persons with conductive hearing disorders. The sample of children consisted of eight normal hearing individuals with socially acceptable articulation and eight normal hearing individuals with faulty articulation. The subjects are described in greater detail below under each of the four experiments.

Experimental Design

The design of each of the four experiments is given below. The details of the procedures employed for all four experiments, however, are reported later in the chapter.

Experiment I

Conditions.--The four audiological conditions were as follows: actual test results, better than actual a

w

y

h

te

me

40

rec

Exp

in

in E

of-h

hear

and c

subje

not e

test results, poorer than actual test results, and no test results.

Testers.--Four sophisticated clinical audiologists were employed as testers.

Subjects.--The subjects were eight normal hearing adults ranging in age from twenty to twenty-five years with a mean age of 21.4 years and a median age of 20.5 years. All but one of the subjects were female, and all had socially acceptable articulation.

Tests Administered.--One ear of each subject was tested by each tester by pure-tone air-conduction audiometry at the following frequencies: 250, 500, 1000, 2000, 4000, and 8000 Hz. This was followed by obtaining a speech reception threshold on the same ear.

Experiment II

Conditions.--The same four conditions were employed in this experiment as those in Experiment I.

Testers.--The testers were the same as those used in Experiment I.

Subjects.--The subjects consisted of sixteen hard-of-hearing adults. Eight subjects had a sensorineural hearing loss that had been diagnosed both audiologically and otologically. A criterion for selection of these subjects was that their speech discrimination scores did not exceed 88% at a 26 dB sensation level re their speech

]

4

ma

be

Th

yea

32.

dis

pers

clos

word

exper

the f

teste

audior

crimin

tested

Experim

were th

II.

reception threshold. Six of the subjects were male and two were female ranging in age from forty to sixty-three years with a mean age of 47.0 years and a median age of 43.5 years.

The remaining eight subjects, six females and two males, exhibited a conductive hearing impairment that had been diagnosed both audiologically and otologically. These subjects ranged in age from eighteen to fifty-two years with a mean age of 35.6 years and a median age of 32.0 years. No limitations were placed on these subjects' discrimination scores for inclusion in the study, since persons with conductive hearing losses generally score close to 100% when phonetically balanced monosyllabic words are presented at a 26 dB sensation level.

Tests Administered.--Each of the subjects in this experiment was tested bilaterally by air conduction at the frequencies of 500, 1000, and 2000 Hz and one ear was tested by bone conduction at the same frequencies. Speech audiometry (speech reception thresholds and speech discrimination) was administered to the same ear as was tested by bone conduction.

Experiment III

Conditions.--The conditions in this experiment were the same four as those utilized in Experiments I and II.

Testers.--The same four sophisticated testers participated in this experiment as in Experiments I and II.

Subjects.--The subjects were sixteen normal hearing children. Eight of these children, five females and three males, ranged in age from five to seven years with a mean age of 5.8 years and a median age of 5.0 years. These subjects were judged to have normal speech articulation. The remaining eight, six males and two females, ranged in age from five to nine years with a mean age of 6.3 years and a median age of 6.0 years. These children had defective articulation characterized by substitutions, omissions, and/or distortions. The diagnosis of defective articulation was made by speech pathologists at the Michigan State University Speech and Hearing Clinic.

Tests Administered.--Each of these subjects was given a pure-tone air-conduction test at 500, 1000, and 2000 Hz for one ear and speech audiometry (speech reception threshold and speech discrimination) on the same ear as the pure-tone thresholds.

Experiment IV

Conditions.--The conditions used in this experiment were the same four as those employed in Experiments I, II, and III.

Testers.--The testers were four undergraduate female students who were majoring in audiology and speech

science at Michigan State University and who were taking their first couse in audiology. Thus, they were unsophisticated testers in comparison to the pre-doctoral students.

Subjects.--Sixteen hearing impaired adults served as subjects. Although these subjects were different persons from those who participated in Experiment II, the criteria for their selection were the same. However, no limitations were placed on these subjects' speech discrimination scores since this audiometric test was not given to these individuals. Eight of the subjects, two females and six males, were diagnosed as having a sensorineural hearing loss. They ranged in age from thirty-seven to sixty-five years with a mean age of 50.3 years and a median age of 49.5 years. Eight of the subjects exhibited a conductive hearing impairment. All of these subjects were females with the exception of one, and they ranged in age from sixteen to thirty-five years with a mean age of 31.4 years and a median age of 30.0 years.

Tests Administered.--Each subject was tested by pure-tone air-conduction audiometry at 250, 500, 1000, 2000, and 4000 Hz bilaterally and by pure-tone bone-conduction audiometry bilaterally at the same frequencies.

Equipment

The audiometric test equipment and calibration equipment used in this study are listed below.

Test Equipment

Pure Tone Clinical Audiometer (Beltone, model 15-C)
20 decibel Attenuation Pad (Constructed at Michigan State University)
Earphones (Telephonics, model TDH-39)
Earphone Cushions (model MX 41/AR)
Bone Vibrator (Radioear, model B70-A White Dot)
Narrow Band Masking Unit (Beltone, model NB-102)
Speech Audiometer (Grason-Stadler, model 162)
Commercial Test Room (Industrial Acoustic Company, Inc. 1200 Series)
Tape Recorder (Ampex, model 601)

Calibration Equipment

Sound Level Meter (Bruel & Kjaer, type 4152)
Octave Band Filter Network (Bruel & Kjaer, type 1613)
Artificial Ear (Bruel & Kjaer, type 4152)
Condenser Microphone (Bruel & Kjaer, type 4132, used in conjunction with the artificial ear)
Artificial Mastoid (Beltone, model M5A)
Voltmeter placed in the Audio Frequency Spectometer (Bruel & Kjaer, type 2112)

The equipment was located in the Michigan State University Speech and Hearing Clinic. All audiometric assessments were obtained with the subject seated in the IAC sound-treated room. The ambient noise level in this room was 42 dB SPL as measured on the C scale of the sound level meter. An octave band analysis of the ambient noise in this room was conducted, and it was found that the greatest amount of ambient noise (40 dB average) was in the octave below 125 Hz. The average ambient noise for octave bands between 125 and 8000 was 14 dB SPL.[1]

---

[1]F. Harry Tokay, "Validity and Reliability of Berkesy Audiometry with Preschool Age Children," (Unpublished Doctoral Thesis, Michigan State University, 1967).

Adjacent to the sound-treated IAC room was a control room that contained the audiometric equipment used in all testing. This room was a separate structure and was not pre-fabricated. A window and a talk-back system were located between the two rooms that enabled visual and auditory communication between the tester and the subject during test sessions. However, in the present study the window was covered to eliminate the possibility of the testers obtaining visual clues from the subjects. This procedure was followed to control for the possibility that the tester may have had some acquaintance with the subject. This acquaintance could influence significantly the behavior of the subject or tester or both.[1]

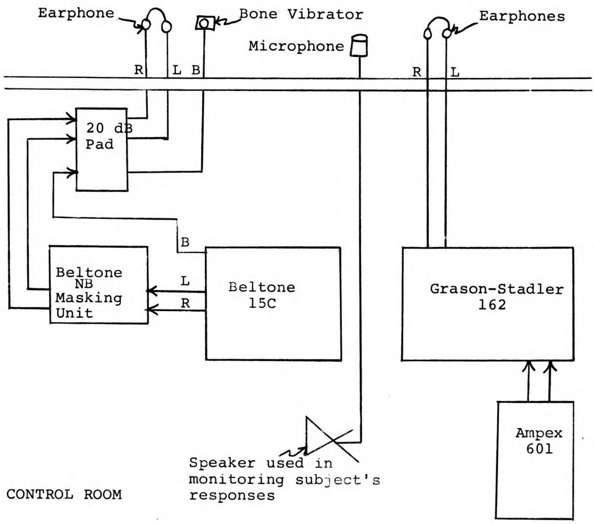
On the following page is a block diagram of the test equipment used in this study.

A pure-tone audiometer that drove a matched pair of TDH-39-10Z earphones housed in MX 41/AR biscuit-type cushions or a hearing aid-type bone vibrator was used for all pure-tone testing. An auxillary attenuator was constructed, and it was found that employment of this pad reduced the minimum output (maximum attenuation) from 0 to -20 dB hearing level re ISO 1964. The response of this pad was applicable for both air and bone-conducted stimuli and was accurate within plus or minus one dB.

---

[1]Rosenthal, Experimenter Effects in Behavioral Research, p. 87.

FIGURE 2.--Block diagram of the audiometric equipment used to make audiometric assessments in this study.



The air-conduction system of the pure-tone audiometer was checked with the artificial ear assembly employed in conjunction with the sound level meter and the associated octave band filter network. These physical measurements were made according to the procedures outlined by

the American Standards Association.[1] The linearity of this system was determined by using the same calibrating equipment.

The stability of the bone-conduction system of the pure-tone audiometer was periodically checked by following the calibration procedures described in the Operating Manual for the Beltone Artificial Mastoid.[2] A voltmeter and the artificial mastoid assembly were employed when making the physical measurements. The output voltage was converted to dB re 1 dyne and was compared to the "HAIC Interim Bone Calibration."[3]

All speech audiometry scores were obtained by using a speech audiometer that was capable of amplifying and attenuating the output of the tape recorder used to present the tape recorded speech stimuli. For this study, the speech audiometer drove a matched pair of TDH-39-10Z earphones housed in MX 41/AR cushions. This audiometer was calibrated so that audiometric zero was equivalent to 20 dB re 0.0002 dyne/cm^2.

---

[1] ________. "American Standard Specifications for Audiometers for General Diagnostic Purposes," American Standard Association, Incorporated (1960), p. 9.

[2] ________. Operating Manual for the Beltone Artificial Mastoid, M5 Series, Chicago, Illinois, Beltone Electronics Corporation.

[3] S. F. Lybarger, "Interim Bond Conduction Thresholds for Audiometry," Journal of Speech and Hearing Research, 9 (1966), pp. 483-487.

Calibration of this system was conducted by using "Speech Noise" instead of a 1000 Hz pure tone as described by the American Standards Association. Tillman, Johnson, and Olsen[1] reported that the spectral configuration of "Speech Noise" is closer to the spectrum of speech produced by male speakers than is the 1000 Hz pure tone. The calibration equipment used to check this system was the same as that employed to calibrate the air-conduction system of the pure-tone audiometer. The "Speech Noise" was channeled into the sound level meter assembly at a 60 dB attenuator setting and with a deflection of zero on the speech audiometer VU meter. The acoustic output of the system was measured and this was considered to be equivalent to the intensity of the spondee words at the same attenuator setting when the peaks of the words produced a deflection of zero on the VU meter.

Recorded Speech Materials

All speech stimuli were recorded by an experienced male speaker. The spondee words were taken from the CID Auditory Test W-1 word lists and were recorded on high quality magnetic tape by a local male talker having

---

[1]Tom W. Tillman, Robert M. Johnson, and Wayne O. Olsen, "Earphone versus Sound Field Threshold Sound-Pressure Levels for Spondee Words," Journal of the Acoustical Society of America, 39 (1966), pp. 125-133.

General American speech. The tape consisted of six randomizations of thirty-six spondee words, namely, CID Auditory Test W-1 Lists A, B, C, D, E, and F.[1] Each list was produced by re-recording from a master list and cutting and splicing the words into the appropriate order to construct the six randomizations. Each word on the tape was monitored by the speaker on a VU meter so that each syllable of each word peaked at the same intensity level as the 1000 Hz calibration tone that had been recorded at the beginning of the tape. Both syllables of each word were monitored by the speaker to a level of plus or minus one decible re the calibration tone.

The speech discrimination materials were taken from two sources and recorded by the same male speaker who recorded the spondee words.

Speech discrimination stimuli for the adult subjects were taken from the NU Auditory Test No. 6.[2] The speech discrimination stimuli for the children were taken from the Phonetically Balanced Kindergarten lists.[3]

---

[1]Hirsh, <u>et al</u>., "Development of Materials for Speech Audiometry," pp. 321-337.

[2]Tillman and Carhart, "An Expanded Test for Speech Discrimination," pp. 1-12.

[3]Haskins, "Speech Discrimination for Children."

The NU Auditory Test No. 6 consists of four lists of fifty monosyllabic words patterned after the CNC lists developed and revised by Lehiste and Peterson.[1, 2] These lists were recorded on magnetic tape. The speaker used the carrier phrase, "You will say. . ." before enunciating one of the monosyllabic words. The speaker monitored his vocal production for these words by monitoring his vocal output for the word "SAY" of the carrier phrase to zero dB on the VU meter. Again, all monitoring was done by means of observing the deflection of the needle of the VU meter re 1000 Hz calibration tone that had been recorded at the beginning of the tape. Four additional lists, necessary for the experimental conditions in the present investigation, were made from scramblings of the original four lists. The scramblings were achieved in the same manner as explained for the randomizations of the spondee words.

The four CNC lists were standardized earlier by Rintelmann and Jetty[3] on ten young adult subjects with normal hearing following the procedures outlined by

---

[1]I Lehiste and G. E. Peterson, "Linguistic Considerations in the Study of Speech Intelligibility," Journal of the Acoustical Society of America, 31 (1959), pp. 280-286.

[2]G. E. Peterson and I. Lehiste, "Revised CNC Lists for Auditory Tests," Journal of Speech and Hearing Disorders, 27 (1962), pp. 62-70.

[3]Rintelmann and Jetty, "Reliability of Speech Discrimination Testing."

Th

wh

st

ch

ga

to

pur

Disc

Tillman and Carhart.[1] The results were essentially the same as those obtained by Tillman and Carhart, and the lists were equivalent plus or minus four percent at a 24 dB sensation level.

The speech discrimination materials for the children consisted of four randomizations of List 1 of the PBK Word Lists. These words were recorded and randomized in the same manner as were the monosyllabic words used for speech discrimination testing of adults.

Procedures

Each subject was tested audiometrically five times. The first examination was conducted by the investigator, whereas the remaining sessions were conducted by four student testers. The subject was seated in a comfortable chair located in the IAC sound-treated room. The investigator gave each subject the following instructions prior to testing him by pure-tone air-conduction audiometry:

> You are now going to hear a tone. First, all of the tones will be in your right ear and then all of the tones will be in your left ear. When you hear the tone please push this button. Some of the tones will be very faint, but push the button whenever you think you hear the tone. Are there any questions?

Each of the fifty-six subjects was first given a pure-tone air-conduction test bilaterally by the

---

[1]Tillman and Carhart, "An Expanded Test for Speech Discrimination," pp. 1-12.

r

i

a

b

c

t

a

te

di

ma

Th

fo

the

gen

min

investigator. This was accomplished by using the Hughson-Westlake ascending technique.[1] Testing started at 1000 Hz and the subject's threshold was determined for that frequency. Threshold was established when the subject responded at the lowest hearing level three times. The investigator followed the same procedures for 2000, 4000, and 8000 Hz. The threshold at 1000 Hz was rechecked before establishing thresholds at 250 and 500 Hz. Upon completion of the air-conduction testing, bone-conduction thresholds were obtained at octave intervals between 250 and 4000 Hz. The investigator followed the same testing techniques for measuring bone-conduction thresholds as he did for establishing air-conduction thresholds; however, masking was used routinely during bone-conduction testing. The following instructions were given to each subject before the pure-tone bone-conduction test was administered.

> You are now going to hear the same kind of tone through this apparatus; however, you will hear a noise in this ear (pointing to the subject's ear that would receive the masking agent). Try not to pay attention to the noise, but rather listen very carefully for the tone. When you hear the tone please push the button just as you have been doing. Are there any questions?

A narrow band masking unit was utilized to produce the masking stimulus. Previous measurements of this noise generator indicated that the band widths, determined at a

---

[1]Carhart and Jerger, "Preferred Method for Determination of Thresholds," pp. 330-345.

level 3 dB down from the peak intensity, were greater than the critical band widths defined by Fletcher.[1, 2] Effective masking for zero dB hearing level was determined at each band by applying the critical band data of Fletcher in a manner outlined by Sanders and Rintelmann.[3]

The Hood Technique[4] was employed to establish masked thresholds. The procedures for this technique are as follows:

> Find the bone-conduction threshold with the bone-conductor applied to the mastoid of the tested ear without masking of the untest ear. Apply the masking noise of the appropriate band to the untest ear by means of air conduction and find a bone-conduction threshold reading. Increase the level of masking noise by 10 dB above threshold and retest the bone-conduction threshold. If the bone conduction threshold is raised by 10 dB increase the masking intensity by another 10 dB and repeat. Continue this procedure until the point is reached at which the bone conduction remains constant with further additional incremental steps of 10 dB of masking noise. This is the "change-over point" and gives the true bone-conduction threshold of the tested ear.

---

[1]Harvey Fletcher and W. A. Munson, "Relation Between Loudness and Masking," Journal of the Acoustical Society of America, 9 (1937), pp. 1-10.

[2]Harvey Fletcher, "Auditory Patterns," Review of Modern Physics, 12 (1940), pp. 47-65.

[3]Jay W. Sanders and William F. Rintelmann, "Masking in Audiometry," Archives of Otolaryngology, 80 (1964), pp. 541-556.

[4]J. D. Hood, "Principles and Practice of Bone Conduction Audiometry," Laryngoscope, 70 (1960), p. 1215.

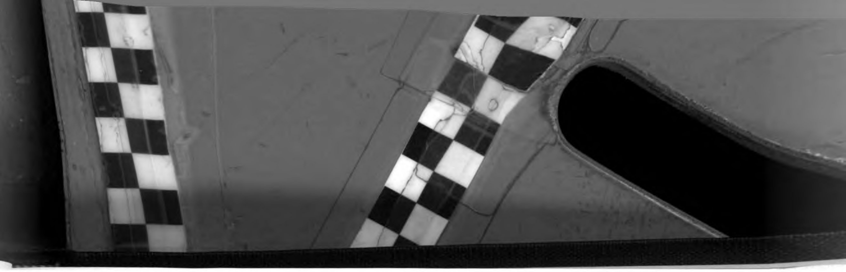
Immediately following the pure-tone air and bone-conduction testing, forty of the subjects (Experiments I, II, and III) were given speech reception threshold tests and thirty-two subjects (Experiments II and III) were given speech discrimination tests. All speech audiometric scores were obtained monaurally, but both ears were tested for all subjects. Each subject was given the following instructions before the administration of the spondee words used to establish the subjects' speech reception thresholds.

> You are now going to listen to a number of two-syllable words, for example, horseshoe, hothouse, northwest, etc. You are to repeat the words that you hear. You will hear thirty-six of these words at a comfortable level and you are to repeat them. Following this you will again listen to the same words; however, they will be in a different order than the first list you heard. Again you are to repeat these words which will get fainter and fainter. If you are not sure of the word please guess. You will hear the words in your right ear first and then in your left ear. Are there any questions?

The subject then heard and responded to an entire list of spondee words presented at a suprathreshold level that was comfortable for him. This procedure was based on the findings of Tillman and Jerger.[1] These authors reported that the accuracy of the speech reception threshold is a function of word familiarity; therefore,

---

[1]Tillman and Jerger, "Some Factors Affecting Spondee Threshold," pp. 141-146.



if the listener has the opportunity of hearing all the words before his speech reception threshold is recorded, he will tend to improve the preciseness of his scores for this test. Following this procedure, the subject was then presented one of the taped spondee lists. An ascending technique was employed to establish threshold; however, spondees were presented in 2-dB steps instead of 5-dB steps as were the pure tones. A subject's speech reception threshold was established when he correctly responded to four out of eight spondee words at the lowest intensity level.

Following speech reception threshold measurements, specific subjects (the thirty-two participating in Experiments II and III) were given speech discrimination tests. Prior to obtaining the speech discrimination scores the following instructions were given to each subject.

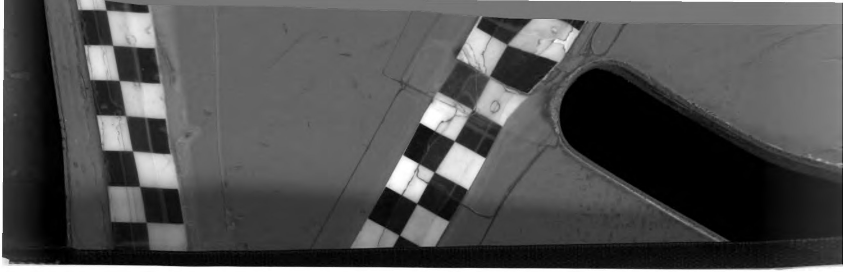
> You are not going to listen to some one-syllable words. These words will be very short and you will have to listen very carefully. Again, please guess if you are not sure of the word; however, only say the word you hear or think you hear. If you talk or explain that you cannot understand the words, it is likely to cause you to miss the following word. Are there any questions?

The adult subjects were administered fifty tape recorded one-syllable phonetically balanced words from the NU Auditory Test No. 6. Children were given fifty monosyllabic words taken from the PBK Word lists. Both were presented at a 26 dB sensation level re the subjects' speech reception thresholds.

The number of audiometric tests a subject received was determined by the experiment in which he participated. When the subject had been given the necessary number of audiometric tests by the investigator, he was instructed when to return for additional testing. However, in each experiment except Experiment III, all subjects were seen by the testers immediately following the investigator's audiometric evaluation. In experiment III where children were used as subjects, the investigator conducted the initial evaluation on a different day than did the testers. However, each of these subjects was seen by all four testers within forty-eight hours after the initial evaluation.

Eight testers were selected to participate in four separate experiments in this study. The testers were divided into two groups. These categories were specifically defined earlier in this chapter, but to briefly reiterate, one group of testers consisted of four pre-doctoral students and the other group was composed of four undergraduate students enrolled in an introductory audiology course.

Prior to their examination of the subjects, the testers were given specific instructions. These instructions were read to the group of testers who were going to participate in a particular experiment. This procedure was followed in order that each tester head the same

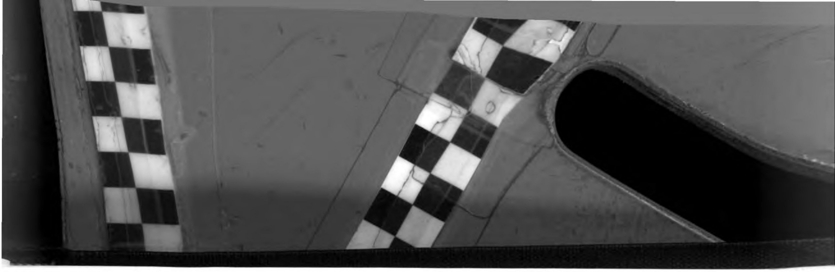


instructions at the same time; therefore, each one heard the same intonations, pauses, repetitions, and verbal clues that were unintentionally read by the investigator. Following are the instructions given to the testers. These instructions are stated in their entirety. However, depending on the experiment, some of the instructions were deleted to meet the specific needs of each experiment.

> You are going to administer a pure-tone air or bone-conduction test(s) or both to eight (or sixteen) subjects. You will use the Carhart-Jerger modified method of the original Hughson-Westlake ascending technique to establish thresholds. You will test one or both ears of each subject. This will be specified to you just prior to testing the subject. In obtaining pure-tone thresholds, you will start at 1000 Hz at suprathreshold and attenuate in 10 dB steps until the subject fails to respond, at this point you will increase the intensity in 5 dB steps until the subject again responds. Following this procedure you will record the subject's threshold as the lowest intensity that he responds to three times. You will follow this same procedure for the following frequencies: 250, 500, recheck 1000, 2000, 4000, and 8000 Hz.
>
> You are to obtain bone-conduction thresholds in the same manner for the same frequencies as by air-conduction except 8000 Hz. If masking is deemed necessary according to Studebaker's criteria, you will use narrow band white noise and the Hood Technique to establish thresholds. You will increase the masking stimulus 15 dB re effective masking beyond the point where the subject's masked threshold fails to shift.
>
> You will also administer speech audiometry to the subjects. You will obtain a subject's speech reception threshold or speech discrimination score or both on one ear. When administering the spondees you will present an entire list of CID W-1 words at suprethreshold (comfortable to the subject). You will begin the test following the same procedures as you did in establishing pure-tone thresholds except when ascending you will ascend in 2-dB steps. Threshold will be established when a subject correctly responds to four out of eight words at the lowest intensity level.

d
t
ar
gr
so
su
in
vi
thr
rec
mak
in t
ity
tion
this

t
s



> When determining a subject's speech discrimination score, you will present a list of fifty monosyllabic words at a 26 dB sensation level re the subject's speech reception threshold.
>
> As was previously stated, you will be testing eight (or sixteen) subjects. The subjects may or may not have an accompanying audiogram. When the subject does have an audiogram you are to study it carefully. You are to replicate the tests that are reported on the audiogram. If you do not receive an audiogram you will be told which tests to administer to which ear.
>
> You are not to discuss your results among yourselves. I will be in the testing area to give the subjects instructions, change transducers, and prepare tapes. Are there any questions?

Each tester was given one of the previously described audiograms (actual, better, poorer, or no information) just prior to testing each subject. Each tester had ample time to review the information contained on the audiograms while the investigator seated the subject in the soundtreated room IAC room. The investigator gave each subject the same instructions he had given him during the initial test session and also placed the earphones or bone vibrator on the subject. This procedure was followed for three basic reasons: (1) to make sure that each subject received the same instructions and information, (2) to make sure the transducers were correctly placed and always in the same position, and (3) to control for the possibility of the subject's giving the tester additional information causing him to alter his bias. Rosenthal classified this type of design as a "blind experiment." He wrote:

> . . .if the experimenter did not know the hypothesis being tested but did know to which group each subject belonged, the results of the study are more

a

b

o

(s

me

co

of

inf

sen

of

fol

bone

disc

Resea

> likely to be biased in the direction of the hypothesis (or opposite to it) rather than biased irrelevantly with respect to the hypothesis.[1]

The investigator remained in the control room while the tester conducted the evaluation; however, his only function was to determine when the transducers needed changing, when different instructions were necessary for the subject, and when the tapes containing the speech stimuli needed to be changed.

By following these procedures, each subject in all four experiments was tested five times. He was tested by the investigator first and then by each tester under one of four conditions. Therefore, in each experiment (sub-groups of eight subjects) each tester obtained audiometric scores on two subjects under each of the four conditions. In each experiment a counter-balanced order of condition presentation (actual, better, poorer, and no information) was employed so that each condition was presented first, second, third, and fourth an equal number of times. All audiometric tests were presented in the following order: pure-tone air-conduction, pure-tone bone-conduction, speech reception threshold, and speech discrimination.

---

[1]Rosenthal, Experimenter Effects in Behavioral Research, p. 370.

E

us

as

he

dis

tes

fred

bone

metr

tion,

by bo

Summary of Experiments

Experiment I

Testers.--Four sophisticated pre-doctoral students participated as testers.

Subjects.--Eight normal hearing adults were employed as subjects.

Tests Administered.--Each subject was tested unilaterally by air conduction at the following frequencies: 250, 500, 1000, 2000, 4000, and 8000 Hz. This was followed by obtaining a speech reception threshold on the same ear.

Experiment II

Testers.--The testers were the same four as those used in Experiment I.

Subjects.--Sixteen hearing impaired adults served as subjects. Eight of these subjects had a conductive hearing impairment and eight had a sensorineural hearing disorder.

Tests Administered.--Each of these subjects was tested monaurally in both ears by air conduction at the frequencies of 500, 1000, and 2000 Hz and unilaterally by bone conduction at the same frequencies. Speech audiometry (speech reception thresholds and speech discrimination) was administered to the same ear as the ear tested by bone conduction.

p

o

th

th

E

s

au

as

he

pu

20

ai

Experiment III

<u>Testers</u>.--The same four sophisticated testers participated in this experiment as in Experiments I and II.

<u>Subjects</u>.--Sixteen normal hearing children participated as subjects. Eight of these children had socially acceptable articulation and eight had defective articulation.

<u>Tests Administered</u>.--Each subject was given a pure-tone air-conduction test at 500, 1000, and 2000 Hz on a single ear and speech audiometry, (speech reception threshold and speech discrimination) was administered on the same ear as the pure-tone thresholds.

Experiment IV

<u>Testers</u>.--Four undergraduate students majoring in speech and hearing therapy and enrolled in an introductory audiology course were employed as testers.

<u>Subjects</u>.--Sixteen hard-of-hearing adults served as subjects. Eight of these individuals had a conductive hearing loss and eight had a sensorineural hearing disorder.

<u>Tests Administered</u>.--Each subject was tested by pure-tone air-conduction audiometry at 250, 500, 1000, 2000, and 4000 Hz monaurally in both ears by pure-tone air and bone-conduction.

CHAPTER IV

RESULTS AND DISCUSSION

This chapter is divided into two basic portions. The first part presents the results of the clinical significance between pure-tone and speech audiometry scores. The second part gives the statistical significance between mean pure-tone and speech audiometry scores. Each of these major divisions is sub-divided into four sections--the four experiments described in Chapter III of this investigation. The primary results obtained for each experiment are reported followed by a discussion of these results.

One method of showing the relationship between testers, conditions, and stimuli employed in this study is by examining the means, standard deviations, standard errors of measurement, and ranges of the audiometric scores obtained in each experiment.

Another method of observing this relationship is by constructing composite audiograms showing the mean pure-tone air and bone-conduction thresholds.

The tables in this chapter present the means, standard deviations, standard errors of measurement and

ranges of the stimuli used in this investigation, while the figures show composite audiograms of the mean pure-tone air and bone-conduction thresholds obtained in the present study.

Experiment I

In this experiment four pre-doctoral students in audiology participated as testers. Each of them tested eight normal hearing adults under one of four conditions. The conditions were as follows: actual test results (actual), better than actual test results (better), poorer than actual test results (poorer), and no test results (no info.). Each tester obtained pure-tone air-conduction thresholds on a single ear for the following frequencies: 250, 500, 1000, 2000, 4000, and 8000 Hz. They also measured speech reception thresholds on the same ear that pure-tone thresholds were obtained.

Clinical Significance

Pure-Tone Air-Conduction Thresholds.--Table II shows the means, standard deviations, standard errors of measurement, and ranges of the pure-tone air-conduction thresholds recorded for eight normal hearing adults by four pre-doctoral students in audiology under four conditions. This table also shows the results (original) obtained for these subjects by the investigator of this study.

TABLE II.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Air-Conduction Thresholds* Obtained for Eight Normal Hearing Adults by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

Condition	Frequency in Hertz					
	250	500	1000	2000	4000	8000
Original						
M	3.75	5.0	-3.75	-2.50	1.87	8.12
SD	4.84	5.0	3.3	3.5	3.47	8.99
Range	15.00	15.00	10.00	10.00	5.00	25.00
Actual						
M	2.50	1.87	-5.0	-5.0	3.75	2.50
SD	4.33	4.28	3.53	6.12	4.14	7.50
SE_m	2.50	3.95	3.53	4.33	3.95	9.18
Range	15.00	15.00	10.00	15.00	15.00	25.00
Better						
M	3.75	1.87	-5.62	-5.00	2.50	4.37
SD	4.14	4.96	4.63	6.12	5.00	6.34
SE_m	2.50	3.95	3.95	5.00	4.63	7.07
Range	10.00	15.00	15.00	20.00	15.00	20.00
Poorer						
M	3.75	3.12	-3.75	-3.12	4.37	4.37
SD	4.14	4.96	4.14	5.55	3.90	10.39
SE_m	3.53	3.95	3.53	3.06	4.33	6.12
Range	10.00	15.00	15.00	15.00	10.00	30.00
No Info.						
M	0.62	1.87	-5.62	-6.25	3.75	5.00
SD	4.63	4.28	3.90	4.14	4.84	8.66
SE_m	3.95	3.95	3.95	5.00	3.95	7.28
Range	15.00	15.00	10.00	10.00	15.00	25.00

*In dB (re ISO 1964 standards)

Examination of Table II reveals that there is very little difference between means for the pure-tone air-conduction thresholds obtained under the four conditions

at any frequency tested. Further, for this study the standard errors of measurement (SE_m) provides the most meaningful information. The SE_m is an estimate of the standard deviation that would be obtained for a series of measurements on the same individual.[1]

Table II shows that the SE_m for all frequencies under all conditions is in good agreement with the data that has been published on the reliability of pure-tone air-conduction audiometry.[2, 3] Test re-test thresholds that are within plus or minus 5 dB are considered to show accaptable clinical reliability. Thus, first, in comparing conditions (actual, better, poorer, and no info.) the reliability of all testers was good from condition to condition at all frequencies tested. Second, the small SE_m at each condition for all frequencies except 8000 Hz shows that the reliability between testers was good for each condition. Therefore, the high repeatability, or lack of variability shown in Table II at all frequencies except 8000 Hz indicates that previous audiometric information did did not significantly cause sophisticated audiologists to influence normal hearing adults' pure-tone air-conduction

---

[1]Jerger, "Evaluation of Some Auditory Measures," p. 6.

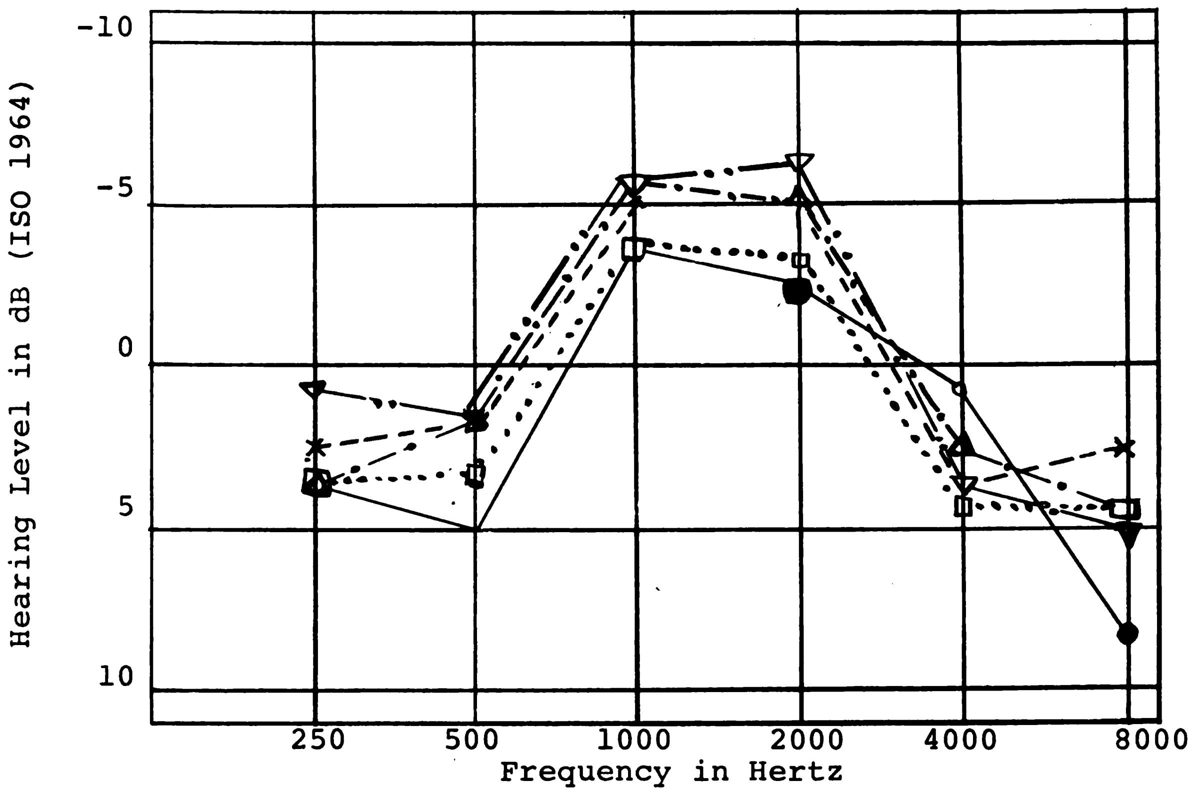
[2]Ibid.

[3]Witting and Hughson, "Inherent Accuracy of Audiograms," pp. 259-269.

thresholds. In other words, according to these results, it seems that previous audiometric information did not cause sophisticated audiologists to exhibit information did not cause sophisticated audiologists to exhibit a bias that clinically affected the reliability or repeatability of pure-tone air-conduction audiometry.

Figure 3 shows a composite audiogram of the mean air-conduction thresholds of the aforementioned eight normal hearing adults obtained by the four testers under the four conditions and the original mean thresholds obtained by the investigator.

FIGURE 3.--Mean Air-Conduction Thresholds for Eight Normal Hearing Adults Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.



Original •——• Actual x---x Better Δ-•--•-Δ Poorer □····□
No Info. ∇-••-∇

The above figure illustrates the high degree of repeatability obtained for the mean air-conduction thresholds for six frequencies under four test conditions. Again, according to reports in the literature on clinical reliability, none of the above means exceed the acceptable amount of error, 5 dB, for pure-tone audiometry.

Speech Reception Thresholds.--The means, standard deviations, standard errors of measurement, and ranges of the speech reception thresholds obtained for eight normal hearing adults by four sophisticated testers under four conditions are presented in Table III. The original SRTs obtained for these subjects by the investigator are also reported in this table.

TABLE III.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Rangés of Speech Reception Thresholds* Obtained for Eight Normal Hearing Adults by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

SRTs	Condition				
	Original	Actual	Better	Poorer	No Info.
M	-2.75	-3.25	-4.00	-3.50	-4.25
SD	1.71	1.71	2.44	2.22	1.19
SE_m		1.73	2.12	2.00	2.00
Range	4.00	4.00	8.00	2.00	4.00

*In dB re speech audiometric zero

The data contained in Table III show that the variability between mean speech reception thresholds under all conditions is minimal. The SE_m between the conditions indicates that the previous audiometric information given to each of the testers prior to their testing each subject did not cause them to influence the subjects' speech reception thresholds significantly. The standard errors of measurement reveal impressively consistent results across all conditions. The average SE_m for all four conditions was approximately 2 dB, which means that even when the audiologist is provided either true or erroneous test results, repeated speech reception thresholds would be no greater than plus or minus 2 dB sixty-eight percent of the time. From the data in Table III, it would be reasonable to expect that a sophisticated audiologist would record speech reception thresholds as clinically accurate no matter what type of previous audiological information he had about the subject. According to the literature on speech reception theshold reliability, the acceptable SE_m for this threshold is 6 dB.[1]

Therefore, the results obtained by the four predoctoral students on these subjects under these conditions

---

[1]Joseph B. Chaiklin and Ira M. Ventry, ". . .A Comparison of 2- and 5- dB Methods," pp. 47-59.

show an acceptable amount of error for speech reception threshold testing. As was evident in pure-tone air-conduction audiometry, the sophisticated testers were not biased by the previous audiometric information, and they did not influence the subjects' speech reception thresholds significantly.

In summary, the descriptive results of this experiment show that all of the differences obtained between means were within the limits of clinical reliability. The average standard error of measurement for pure-tone air-conduction thresholds was 4.46 dB, which is basically the same as the 5 dB test-retest reliability acceptable for pure-tone air-conduction audiometry. The average standard error of measurement for speech reception thresholds was 1.96 dB, which is 4 dB better than the test-retest reliability acceptable for spondee thresholds.

These results indicate that skilled audiologists are not biased by previous audiometric information and, therefore, do not clinically influence air-conduction thresholds or speech reception thresholds obtained from normal hearing adults.

Statistical Significance

A number of null hypotheses were postulated to test the differences among mean pure-tone and speech audiometry scores obtained in this investigation. The

exact number of null hypotheses tested varied from one experiment to another according to the specific auditory stimuli that were presented. Therefore, the null hypotheses postulated for each experiment are enumerated individually. The null hypotheses were tested statistically. A presentation of the statistical technique used and the results are reported followed by a discussion of the significance of the results.

Pure-Tone Air-Conduction Thresholds.--In order to test differences among mean pure-tone air-conduction thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among mean pure-tone air-conduction thresholds obtained on normal hearing adults as a function of the following variables: testers, conditions, and frequencies.

A three-factor analysis of variance with repeated measures on one factor was employed to determine the significance of differences among the variables in the principle comparisons for pure-tone air-conduction audiometry.[1] A three-dimensional table was constructed to organize the data. In this experiment there were four rows representing testers, six columns representing frequency, and four slices representing conditions. The

---

[1]B. J. Winer, Statistical Principles in Experimental Design (New York: McGraw-Hill Book Co., 1962), pp. 338-352.

peated measures were on the frequencies. Thus, an A x x C three-entry table was used to record the data. e analysis of variance was computed with a Control Data orporation 3600 Digital Computer.[1] The pure-tone air-conduction thresholds were the criterion measures entered in each cell. The significance of the variations was determined by the F-ratio.

The means used in comparing differences between testers, conditions, and frequencies are shown in Table II, page 77, and a summary of the analysis is given in Appendix A.

The only variable that was statistically significant was frequency. Levels of confidence within testers, within conditions, and interactions were not statistically significant. Thus, the aforementioned null hypothesis was rejected. However, it was expected that frequency would show significance since research has demonstrated that the inherent stability of human auditory thresholds varies as a function of frequency. In other words, a different amount of variability occurs on repeated pure-tone air-conduction thresholds, depending upon which frequency is being presented to the listener. Thus, the

---

[1]Michigan State University, Agricultural Experiment Station, Analysis of Variance with Equal Frequencies in Each Cell," STAT Series Description #14 (March, 1966).

statistical significance found for frequency can be attributed to variations as a function of frequency for the subjects' pure-tone air-conduction thresholds rather than the bias of the testers.

Speech Reception Thresholds.--In order to test the differences among speech reception thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among the mean speech reception thresholds obtained on normal hearing adults by four pre-doctoral students under four test conditions.

The significance of differences among the variables in this comparison was determined by a two-factor analysis of variance.[1] The two factors were arranged in a two-dimensional table of a 4 x 4 design. The four columns represented the test conditions, while the testers were represented by four rows. The speech reception thresholds obtained were the criterion measures entered in each cell. The F-ratio was used in testing the statistical significance of the variance attributable to the two main effects and the two-way interaction.

The mean speech reception thresholds used in the comparison are shown in Table III, page 80, and a summary of the analysis is given in Appendix A.

---

[1] Winer, Statistical Principles in Experimental Design, pp. 298-318.

The F-statistic was not significant for the two main factors or the interaction between the two factors. Thus, the above null hypothesis could not be rejected.

The three-way analysis of variance on the pure-tone air-conduction thresholds and the two-way analysis of variance on the speech reception thresholds computed for this experiment indicate that sophisticated audiologists were not influenced significantly by previous audiometric results. These analyses illustrate that well-trained testers obtained mean air-conduction thresholds and speech reception thresholds that were quite comparable across conditions with very little variance.

The two-way analysis of variance, which did not employ frequency as one of the variables, showed no significance between testers and test conditions or the interaction between these. These findings give further support to the rationale that in the three-way analysis of variance, where frequency was the only significant factor, the significant finding was not due to tester bias but rather to the inherent variations between auditory thresholds as a function of frequency.

Based on the statistical results of this experiment, it can be seen that experimenter (tester) bias was not exhibited significantly by sophisticated audiologists when obtaining pure-tone air-conduction and speech recep-tion thresholds on normal hearing adults.

Experiment II

In this experiment the same four pre-doctoral students were employed as testers. Each of them tested sixteen hard-of-hearing adults. The subjects were divided into two pathological groups. Eight exhibited a sensorineural hearing loss and eight had a conductive hearing disorder. Each tester obtained pure-tone air-conduction thresholds bilaterally and bone-conduction thresholds on a single ear under one of four conditions for the following frequencies; 500, 1000, and 2000 Hz. The conditions were the same as those described in Experiment I. The testers also measured speech reception thresholds and speech discrimination scores on the same ear that pure-tone bone-conduction thresholds were obtained.

Clinical Significance

<u>Pure-Tone Air-Conduction Thresholds</u>.--Table IV gives the means, standard deviations, standard errors of measurement, and ranges of the air-conduction thresholds obtained by four sophisticated audiologsits under four test conditions for eight adults with a sensorineural hearing loss and eight adults with a conductive hearing impairment. The table also shows the results (original) obtained by the investigator for these subjects.

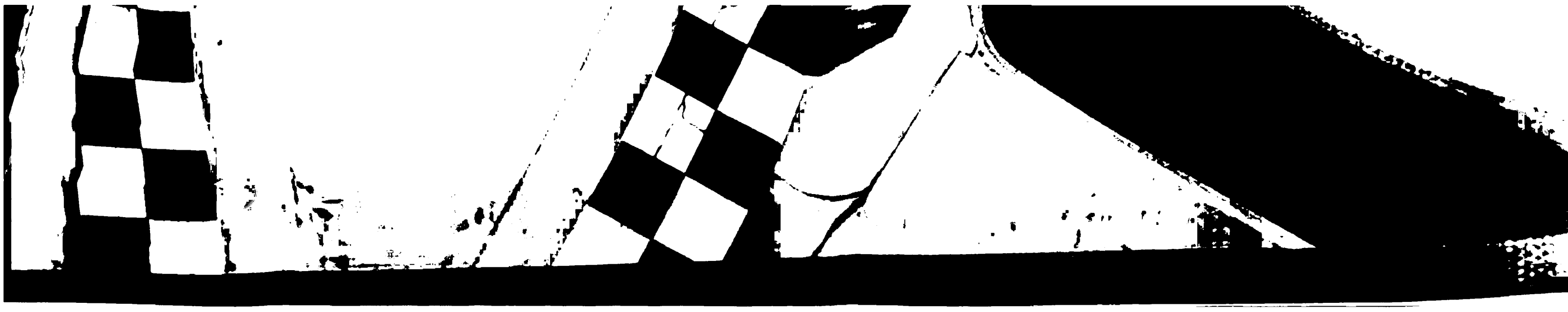


TABLE IV.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Air-Conduction Thresholds* Obtained for Eight Sensorineural (SN) Adult Subjects and Eight Conductive (Cond.) Adult Subjects by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

	Frequency in Hertz					
	500		1000		2000	
Condition	SN	Cond.	SN	Cond.	SN	Cond.
Original						
M	36.25	36.87	45.00	31.25	56.87	30.62
SD	18.15	11.43	11.98	11.65	12.73	9.16
Range	45.00	35.00	40.00	45.00	45.00	35.00
Actual						
M	35.62	35.00	45.62	31.87	56.87	29.37
SD	19.11	11.18	9.49	15.19	12.97	9.82
SE_m	3.06	4.67	6.08	5.86	5.59	3.53
Range	50.00	35.00	35.00	55.00	45.00	35.00
Better						
M	36.25	34.37	46.87	32.50	52.50	27.50
SD	20.11	11.16	11.97	12.23	12.50	8.29
SE_m	4.33	6.08	3.95	3.95	7.28	5.86
Range	50.00	40.00	40.00	45.00	40.00	30.00
Poorer						
M	36.25	34.37	46.87	32.50	55.00	27.50
SD	19.84	7.68	11.97	13.56	12.50	9.68
SE_m	3.53	5.30	3.06	3.95	6.87	5.30
Range	55.00	25.00	40.00	50.00	50.00	35.00
No Info.						
M	35.00	36.25	46.25	31.87	56.25	29.37
SD	17.85	7.68	11.38	13.56	14.94	9.68
SE_m	3.06	5.30	3.53	4.67	3.95	2.50
Range	50.00	40.00	40.00	55.00	55.00	35.00

*In dB re ISO 1964

Table IV shows that the mean air-conduction thresholds are essentially the same under all conditions for the sensorineural subjects. Examination of this

table also reveals that all SE_m for all frequencies under *all* conditions are within the limits of reliability for air-conduction audiometry. The minimal variability for all frequencies under each condition indicates clinically that previous audiometric test result information did not significantly influence sophisticated audiologists when obtaining air-conduction thresholds for subjects with sensorineural hearing disorders. However, it should be noted that the least amount of variability occurred when the testers were provided no previous audiometric information.

The data in Table IV reveal that the mean air-conduction thresholds and the SE_m for the adults with conductive hearing impairments are approximately the same for all frequencies under all conditions. The repeatability for the conductive subjects is essentially the same as it is for the sensorineural subjects. Therefore, according to these results, previous audiometric information did not cause skilled testers to influence significantly hard-of-hearing adults' air-conduction thresholds regardless of the type of hearing disorder. In other words, the data indicate that sophisticated audiologists obtained air-conduction thresholds that were clinically accurate and they were not influenced by past test results.

Figures 4 and 5 show composite audiograms of the mean air-conduction thresholds of the sensorineural and

conductive adult subjects obtained respectively by the four testers under four conditions and the original mean thresholds measured by the investigator. Both figures show that the degree of repeatability between the mean air-conduction thresholds was within 5 dB or less for all conditions.

FIGURE 4.--Mean Air-Conduction Thresholds for Eight Adults with a Sensorineural Hearing Disorder Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

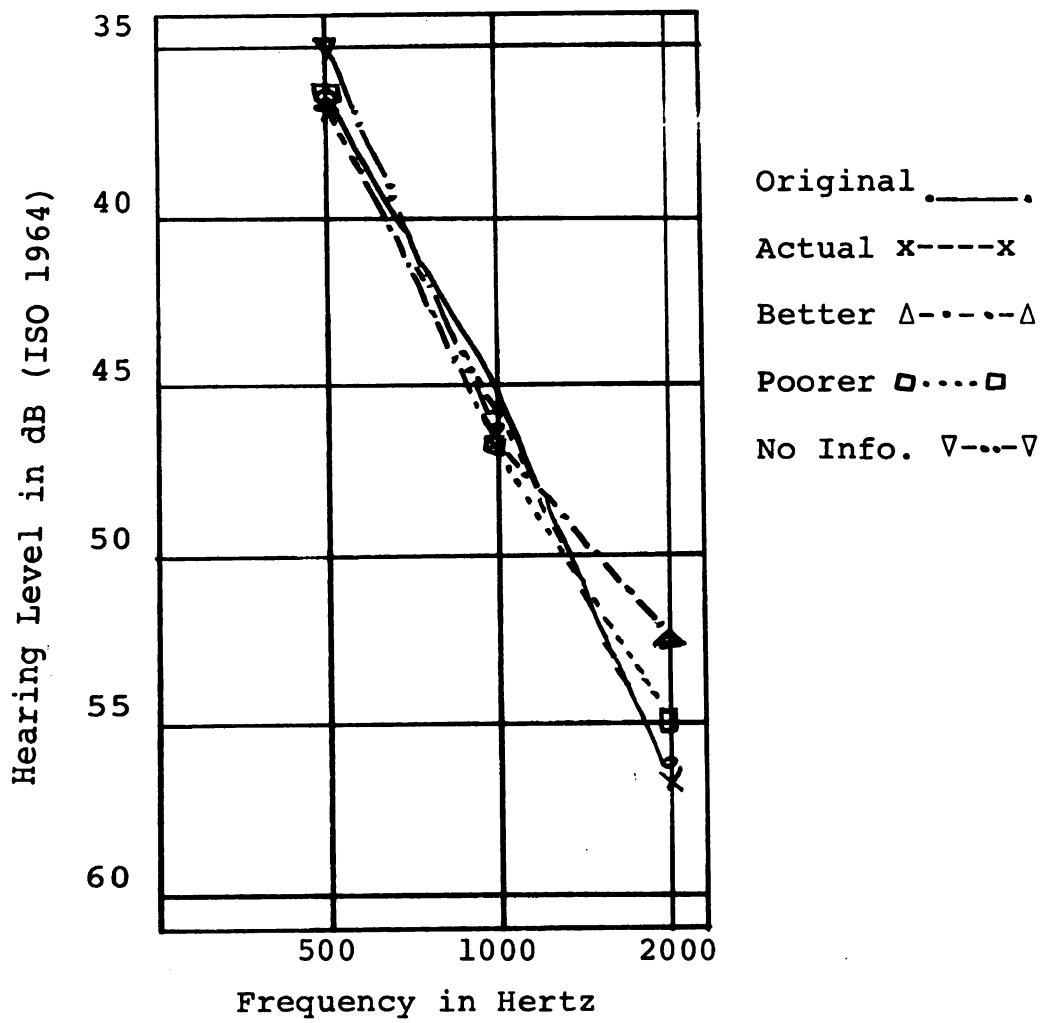
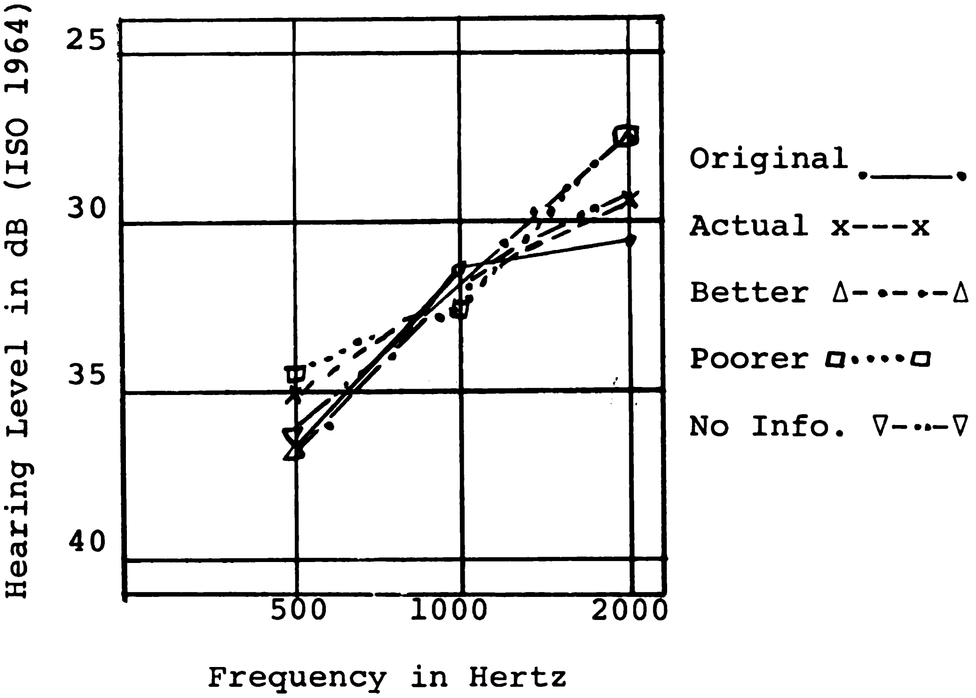


FIGURE 5.--Mean Air-Conduction Thresholds for Eight Adults with a Conductive Hearing Impairment Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.



Pure-Tone Bone-Conduction Thresholds.--The means, standard deviations, standard errors of measurement, and ranges of the pure-tone bone-conduction thresholds obtained for eight adults with a sensorineural hearing loss and eight adults with a conductive hearing impairment by four skilled testers under four test conditions are listed in Table V. The table also shows the results (original) obtained for these subjects by the investigator of this study.

TABLE V.--Means (M), Standard Deviations (SD), Standard Errors of Measurement, (SE_m), and Ranges of Pure-Tone Bone-Conduction Thresholds* Obtained for Eight Sensorineural (SN) Adult Subjects and Eight Conductive (Cond.) Adult Subjects by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

	Frequency in Hertz					
	500		1000		2000	
Condition	SN	Cond.	SN	Cond.	SN	Cond.
Original						
M	38.12	15.62	44.37	10.62	54.37	8.12
SD	12.48	5.82	11.57	6.34	9.82	8.63
Range	35.00	15.00	40.00	20.00	30.00	20.00
Actual						
M	36.87	16.87	45.62	11.25	53.75	8.12
SD	13.44	6.58	10.73	5.99	11.38	6.58
SE_m	4.33	5.00	5.59	3.95	4.67	5.00
Range	35.00	20.00	35.00	20.00	30.00	20.00
Better						
M	36.87	15.62	45.62	13.12	51.25	8.12
SD	10.86	3.90	10.43	4.96	13.16	8.99
SE_m	3.53	5.59	3.53	3.06	5.86	2.50
Range	30.00	10.00	35.00	15.00	40.00	25.00
Poorer						
M	36.87	17.50	45.00	12.50	51.25	8.75
SD	13.44	6.61	11.45	6.12	13.86	8.19
SE_m	5.00	3.95	1.76	3.53	5.86	3.06
Range	40.00	20.00	40.00	20.00	45.00	20.00
No Info.						
M	36.87	16.87	45.00	15.62	51.87	10.62
SD	13.44	5.55	11.16	8.45	14.34	8.07
SE_m	5.30	5.59	4.33	7.07	5.59	5.59
Range	35.00	15.00	40.00	30.00	45.00	25.00

* In dB re ISO 1964

Examination of Table V shows that the mean bone-conduction thresholds are in close agreement under all test conditions for the adults with a sensorineural

hearing disorder. The SE_m for all frequencies under all conditions do not deviate significantly from the amount of error acceptable in bone-conduction audiometry. The largest SE_m occurred when the testers were given no previous test result information. On the other hand, for air-conduction thresholds (see Table IV), the least variability occurred when the testers were not provided with previous test results while testing sensorineural adults. From these results it can be seen that previous audiometric information did not cause sophisticated audio-lOgists to exhibit a bias that affected the reliability of bone-conduction audiometry when obtaining thresholds on subjects with a sensorineural hearing loss.

Table V also shows that the skilled testers were equally as accurate in obtaining thresholds on conductive adult subjects. The mean bone-conduction thresholds were approximately the same for all frequencies under all conditions. The SE_m reveal a high degree of repeatability, indicating that the sophisticated testers were not influenced significantly by the previous test results. Therefore, according to these results, it seems that skilled testers obtain bone-conduction thresholds based on the subjects' hearing sensitivity and that they are not influenced by previous audiometric information.

Figure 6 shows composite audiograms of the mean bone-conduction thresholds of the eight subjects with a sensorineural hearing disorder obtained by the four testers under four conditions and the original mean thresholds measured by the investigator. Figure 7 shows these same data for the eight conductive hearing impaired subjects.

FIGURE 6.--Mean Bone-Conduction Thresholds for Eight Adults with a Sensorineural Hearing Disorder Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

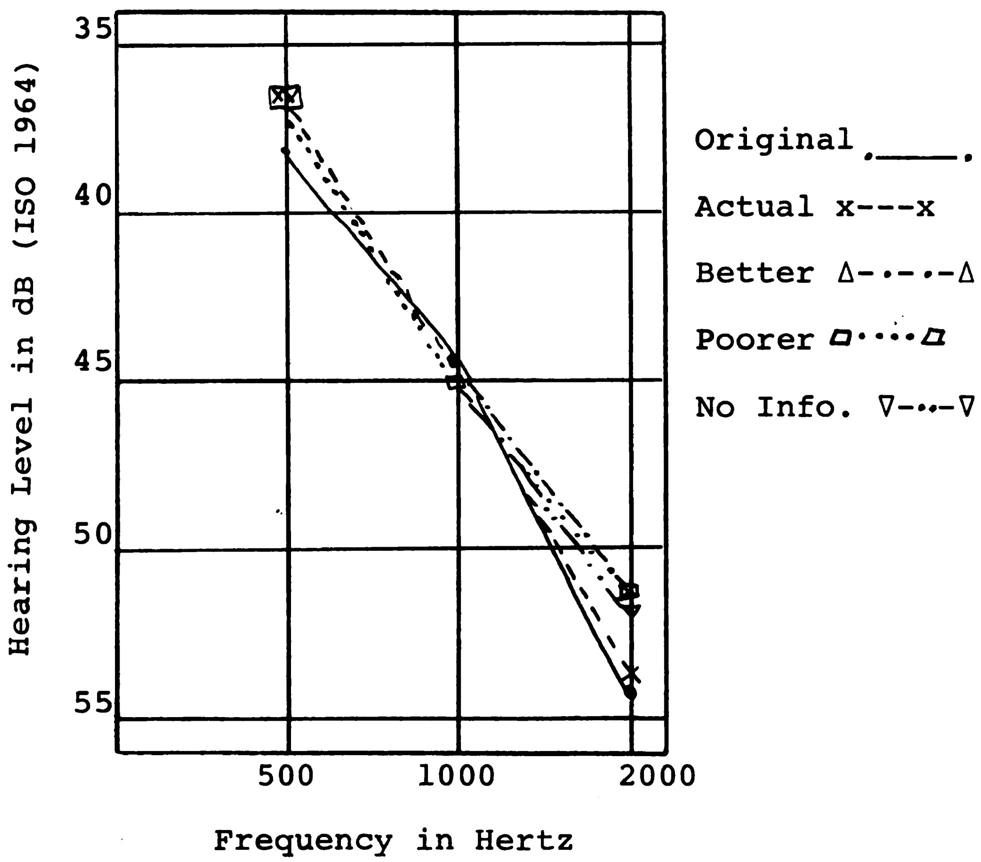
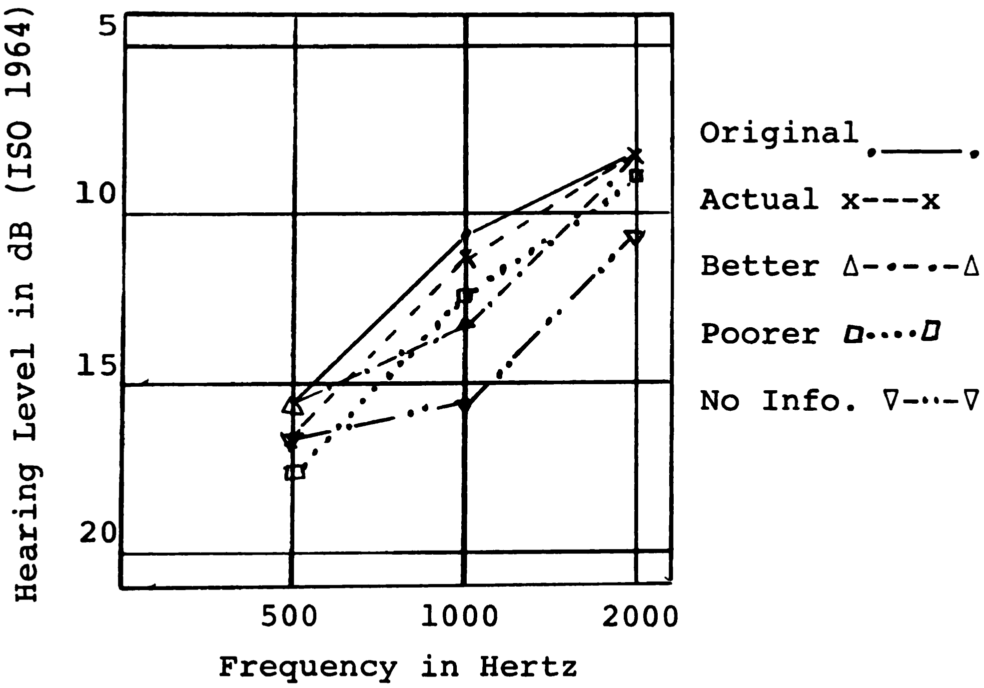


FIGURE 7.--Mean Bone-Conduction Thresholds for Eight Adults with a Conductive Hearing Impairment Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.



These figures show that although the variability between conditions is less than 5 dB at all frequencies for both groups of subjects, the variability was somewhat greater for the bone-conduction thresholds of the conductive loss group than for the sensorineural loss group.

Speech Reception Thresholds.--Table VI reveals the means, standard deviations, standard errors of measurement, and ranges of the speech reception thresholds recorded for eight sensorineural adult subjects and eight conductive adult subjects. The table also shows the original thresholds measured by the investigator of this study.

TABLE VI.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Speech Reception Thresholds* Obtained for Eight Sensorineural (SN) Adult Subjects and Eight Conductive (Cond.) Adult Subjects by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

Subjects	Condition				
	Original	Actual	Better	Poorer	No Info.
SN					
M	38.00	34.75	36.50	35.25	35.50
SD	19.72	19.18	19.08	20.46	17.74
SE_m		4.84	4.47	3.80	3.87
Range	50.00	50.00	48.00	56.00	48.00
Cond.					
M	27.25	28.25	27.50	28.25	26.50
SD	9.99	10.79	12.03	11.19	10.80
SE_m		2.23	2.91	2.23	1.87
Range	38.00	40.00	46.00	42.00	40.00

*In dB re Audiometric Zero

Tha table also shows that the means are approximately equivalent for each pathological group under all conditions. The SE_m are in extremely close agreement and are all below the amount of error acceptable for spondee threshold clinical reliability. Although the SE_m are essentially the same under all conditions for the sensorineural subjects, they are approximately 2 dB larger than the SE_m for all conditions for the adults with a conductive hearing loss.

These results illustrate that even though the variability was slightly greater for the sensorineural

group, trained audiologists were not significantly biased by previous audiometric information when measuring speech reception thresholds of hard-of-hearing subjects regardless of the type of loss.

Speech Discrimination Scores.--The means, standard deviations, standard errors of measurement, and ranges of the speech discrimination scores obtained for eight sensorineural adult subjects and eight conductive adult subjects by four pre-doctoral students under four test conditions are contained in Table VII. The original speech discrimination scores measured by the investigator of this study are also listed in the table.

TABLE VII.-- Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Speech Discrimination Scores* Obtained for Eight Sensorineural (SN) Adult Subjects and Eight Conductive (Cond.) Adult Subjects by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

Subjects	Condition				
	Original	Actual	Better	Poorer	No Info.
SN					
M	72.50	73.75	74.25	73.00	79.75
SD	14.20	16.10	20.89	22.95	15.69
SE_m		5.95	8.51	9.89	9.19
Range	38.00	48.00	62.00	68.00	50.00
Cond.					
M	94.75	92.75	93.50	95.25	93.75
SD	2.22	4.46	4.87	3.59	4.52
SE_m		4.12	3.80	2.64	3.31
Range	6.00	8.00	10.00	10.00	14.00

*In percent correct

Table VII shows that the mean speech discrimination scores for the sensorineural subjects are approximately the same under all conditions except for the condition of "no information." However, all of the mean discrimination scores are higher than the mean score obtained by the investigator. Therefore, the data do not reveal that the testers were biased in the direction of the previous test results. The SE_m between conditions indicate close agreement except for the condition of actual test results, which show the least amount of variability. However, the SE_m for all conditions exceed the acceptable amount of error, four percent, for speech discrimination. Clinically, it is known that subjects with a sensorineural hearing loss usually have relatively poor speech discrimination, and further there is considerable variability in speech discrimination within a population of sensorineural hearing loss subjects. This finding is evident in the present study as shown by the large standard deviations and large ranges for all test conditions. This is dramatically evident when compared to these same measures of variability among the conductive loss group. Nevertheless, based on the data in Table VII, it seems that the variability of the speech discrimination scores was a function of the subjects' responses, rather than a significant influence of the testers' biases. This seems evident because all mean discrimination scores were

recorded in one direction, better than the scores measured by the investigator.

The mean speech discrimination scores for the conductive subjects in Table VII are essentially the same under all conditions. Further, the variability between the scores under all conditions is minimal. The SE_m for all conditions are quite small and well within the limits of speech discrimination reliability. These results indicate that skilled testers were not influenced by previous audiometric information when obtaining speech discrimination scores on adults with a conductive hearing impairment. This finding shows considerable contrast with the sensorineural group.

Table VII shows that sophisticated audiologists obtained speech audiometric scores that were in close agreement with each other when testing the speech discrimination of conductive hearing impaired persons. However, the degree of repeatability measured was poorer when testing speech discrimination of subjects with a sensorineural hearing loss.

In summary the descriptive results of this experiment show that the differences obtained between means for pure-tone air and bone-conduction thresholds, speech reception thresholds, and speech discrimination scores were within the limits of clinical reliability for both pathological groups except for speech discrimination of the sensorineural loss group. The average standard error of measurement for pure-tone air-conduction thresholds was 4.5 dB for the sensorineural group and 4.7 dB for the conductive group thus closely approaching the 5 dB test-retest reliability acceptable for pure-tone air-conduction audiometry.

For bone-conduction audiometry, the SE_m were 4.6 dB for the sensorineural adults and 4.5 dB for the conductive adults. Again, these results show accurate test-retest reliability for bone-conduction thresholds.

The average standard errors of measurement for speech reception thresholds for the sensorineural subjects and conductive subjects were 4.3 dB and 2.3 dB respectively. Both of these SE_m are less than the amount of error permissible for clinically acceptable speech reception thresholds.

The average SE_m for speech discrimination scores were 8.3% for the subjects with a sensorineural hearing loss and 3.5% for the subjects with a conductive hearing disorder. The SE_m for the sensorineural subjects is not within the four percent limit of test-retest reliability for speech discrimination testing. However, since the SE_m under all conditions were approximately the same, these errors were attributed to the subjects' inabilities to discriminate consistently on repeated measures rather than being influenced significantly by testers'

expectancies. On the other hand, the SE_m for the conductive subjects are within the amount of error acceptable for speech discrimination testing. These results indicate that when subjects responded consistently to monosyllable words, there was very little difference in the scores obtained on repeated measures.

According to these results, it is evident that sophisticated audiologists are not influenced significantly by previous audiometric results when obtaining pure-tone air and bone-conduction thresholds, speech reception thresholds, and speech discrimination scores from hard-of-hearing adults regardless of the type of hearing loss.

Statistical Significance

Pure-Tone Air-Conduction Thresholds.--In order to test the differences among pure-tone air-conduction thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among pure-tone air-conduction thresholds obtained for hard-of-hearing adults as a function of the following variables: testers, conditions, pathologies, and frequencies.

A four-factor analysis of variance with repeated measures on one factor was employed to determine the significance of differences among the variables in the principle comparisons for pure-tone air-conduction

audiometry.[1] A four-dimensional table was used to record the data. In this experiment there were four rows representing testers, three columns representing frequency, four slices representing conditions, and two slices representing pathology. Therefore, an A x B x C x D four-entry table was employed to arrange the data. The pure-tone air-conduction thresholds were the criterion measures entered in each cell.

The means used in comparing differences between testers, conditions, frequencies, and pathologies are shown in Table IV on page 88 of this chapter. A summary of the analysis is shown in Table XVII of Appendix B.

The factors of pathology and frequency and the interactions between these were statistically significant. Levels within testers and within conditions were not significant. This indicates that although the testers did obtain thresholds from the sensorineural subjects that were significantly different from the thresholds obtained from the conductive subjects, they did not obtain different thresholds as a function of the conditions. Thus, according to these results, the above null hypothesis was rejected. However, this does not mean that the rejection of the null hypothesis indicates that the testers were

---

[1]Winer, Statistical Principles in Experimental Design, pp. 338-352.

influenced significantly by the previous test results because, as previously mentioned, the means obtained by the four testers under the four conditions were not statistically significant within each pathology group.

Pure-Tone Bone-Conduction Thresholds.--In order to test the differences among pure-tone bone-conduction thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among pure-tone bone-conduction thresholds obtained on hard-of-hearing adults as a function of the following variables: testers, conditions, pathologies, and frequencies.

A four-factor analysis of variance that was the same design as the one described for the air-conduction thresholds in this experiment was used to test the significance of differences among the variables in the principle comparisons for pure-tone bone-conduction audiometry. The pure-tone bone-conduction thresholds were the criterion measures entered in each cell.

The means used in comparing differences between testers, conditions, frequencies, and pathologies are contained in Table V on page 92. A summary of the analysis is shown in Table XVIII of Appendix B.

The factors of pathology and frequency and the interactions between these were statistically significant. Levels within testers and within conditions, however, were not significant. A statistically significant interaction occurred between testers, conditions, and frequencies.

Therefore, its implications need not be explored further. The important finding here was that the interaction within testers and within conditions was not significant.

Based on these findings the above null hypothesis was rejected. However, as stated above, this rejection does not indicate that the testers influenced significantly the subjects' bone-conduction thresholds because, as can be seen, the means obtained by the four testers under the four conditions were not statistically significant within each pathological group.

Speech Reception Thresholds.--In order to test the differences among speech reception thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among mean speech reception thresholds obtained on hard-of-hearing adults as a function of the following variables: testers, conditions, and pathologies.

A three-factor analysis was used to determine the significance of differences among the variables in the principle comparisons for speech reception thresholds. A three-entry table was designed to arrange the data. In this experiment there were four rows representing testers, four columns representing conditions, and two slices representing pathologies. Thus, an A x B x C table was employed to record the data. The speech reception thresholds were the criterion measures entered in each cell.

The means used in comparing differences between testers, conditions, and pathologies are shown in Table VI, page 96, and a summary of the analysis is contained in Table XIX, Appendix B.

The F-statistic was not significant for the three main factors or the interaction among the three factors. Therefore, the above null hypothesis could not be rejected. Thus, the testers did not significantly influence the speech reception thresholds obtained under any of the four conditions for either of the two pathological groups.

Speech Discrimination.--In order to test the differences among speech discrimination scores in this experiment the following null hypothesis was postulated:

> There are no significant differences among the mean speech discrimination scores obtained for hard-of-hearing adults as a function of the following variables: testers, conditions, and pathologies.

The significance of differences among the variables were tested in the same manner as the variables for speech reception threshold in this experiment. The analysis of variance for the mean speech discrimination scores was the same design as for the speech reception thresholds. The obtained speech discrimination scores (means) were the criterion measures entered in each cell, and the F-ratio was used in testing the statistical significance of the variance attributable to the three main effects and the three-way interaction.

The mean speech discrimination scores used in the comparison are shown in Table VII, page 97, and a summary of the analysis is given in Table XX of Appendix B.

The only factor that was statistically significant was pathology. Levels within testers, within conditions, and interactions were not statistically significant. Thus, the aforementioned null hypothesis was rejected. It was not unexpected that pathology would show significance since clinically it is known that hearing acuity is a function of pathology. In other words, subjects with a sensorineural hearing loss have discrimination scores that are poorer than the subjects with a conductive hearing impairment. Therefore, the statistical significance found for pathology can be attributed to variations as a function of the subjects' ability to discriminate monosyllabic words rather than the bias of the testers supported by the fact that the standard deviations, standard errors of measurement, and ranges for these subjects were relatively large. However, each tester under each condition obtained essentially the same scores.

The four-way analyses of variance on the pure-tone air and bone-conduction thresholds and the three-way analyses of variance on the speech audiometry scores conducted for this experiment indicate that sophisticated audiologists were not influenced significantly by previous audiometric results. According to the statistical results

of this experiment, it is evident that skilled testers obtained pure-tone air and bone-conduction thresholds, speech reception thresholds, and speech discrimination scores that are clinically and statistically accurate for hard-of-hearing adults regardless of the type of hearing loss, and they did not influence subjects' audiometric scores based on previous information. The only exception to this finding was with the speech discrimination scores of the sensorineural loss subjects. Although statistically they were not significantly different, the scores for three of the four conditions showed slightly greater variability according to stringent measures of clinical acceptability.

Experiment III

In this experiment the same four testers were employed. They each tested sixteen normal hearing children. Eight of the children had normal articulation and eight had defective articulation. Each tester obtained pure-tone air-conduction thresholds, speech reception thresholds, and speech discrimination scores on a single ear under one of four test conditions. The conditions were the same as those described in Experiment I.

Clinical Significance

<u>Pure-Tone Air-Conduction Thresholds</u>.--Table VIII contains the means, standard deviations, standard errors

of measurement, and ranges of the pure-tone air-conduction thresholds recorded for eight normal hearing children with normal articulation and eight normal hearing children with defective articulation by four pre-doctoral students under four conditions. The frequencies tested were: 500, 1000, and 2000 Hz. The column titled "original" presents the results obtained for the subjects by the investigator of this study.

Examination of Table VIII shows that there is very little difference between the mean pure-tone air-conduction thresholds obtained under the four conditions. This table reveals that the SE_m for all frequencies under all conditions are in good agreement with pure-tone air-conduction test-retest reliability. However, there is one exception. This exception occurs at 500 Hz under the condition of better test results for the defective articulating children. The reason for this exception cannot be fully explained or understood; however, at 500 Hz it appears that previous information did have an effect on one condition (better).

Although the standard errors of measurement for these children were somewhat larger than the standard errors of measurement for the normal hearing adults in Experiment I, they were approximately the same under all conditions and were basically within the limits of repeatability for pure-tone audiometry. Clinically this high

TABLE VIII.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Air-Conduction Thresholds* Obtained for Sixteen Normal Hearing Children--Eight with Normal Articulation (N.A.) and Eight with Defective Articulation (D.A.)--by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

	Frequency in Hertz					
	500		1000		2000	
Condition	N.A.	D.A.	N.A.	D.A.	N.A.	D.A.
Original						
M	16.25	10.62	6.87	5.00	6.25	4.37
SD	4.84	3.70	4.96	2.82	4.84	3.12
Range	15.00	10.00	15.00	10.00	15.00	10.00
Actual						
M	10.62	9.37	1.25	3.12	1.87	3.12
SD	5.82	3.70	2.16	3.42	3.47	2.78
SE_m	7.70	5.00	6.84	4.67	5.86	5.00
Range	15.00	15.00	5.00	10.00	10.00	10.00
Better						
M	11.87	5.00	3.12	3.12	3.75	2.50
SD	6.09	3.46	3.47	3.42	4.84	2.82
SE_m	6.84	8.83	5.86	4.67	4.33	4.67
Range	20.00	10.00	10.00	10.00	15.00	10.00
Poorer						
M	13.75	10.62	3.12	5.62	3.12	5.00
SD	5.99	3.70	3.47	3.12	4.96	4.47
SE_m	4.33	5.00	5.59	4.67	5.30	6.37
Range	20.00	10.00	10.00	10.00	15.00	10.00
No Info.						
M	11.25	11.25	3.75	4.37	1.87	2.50
SD	2.16	2.60	4.14	3.70	3.95	2.82
SE_m	6.61	3.95	6.12	3.06	5.00	3.06
Range	5.00	10.00	10.00	10.00	10.00	10.00

* In dB re ISO 1964 standard

degree of repeatability indicates that previous audiometric information did not cause sophisticated audiologists to influence significantly normal hearing children's pure-tone

air-conduction thresholds. According to these results, it seems that previous audiometric information is only slightly more influential in creating a bias among well-trained testers when they are measuring pure-tone air-conduction thresholds on children than when they are making these same measurements on adults.

FIGURE 8.--Mean Air-Conduction Thresholds for Eight Normal Hearing Children with Normal Articulation Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

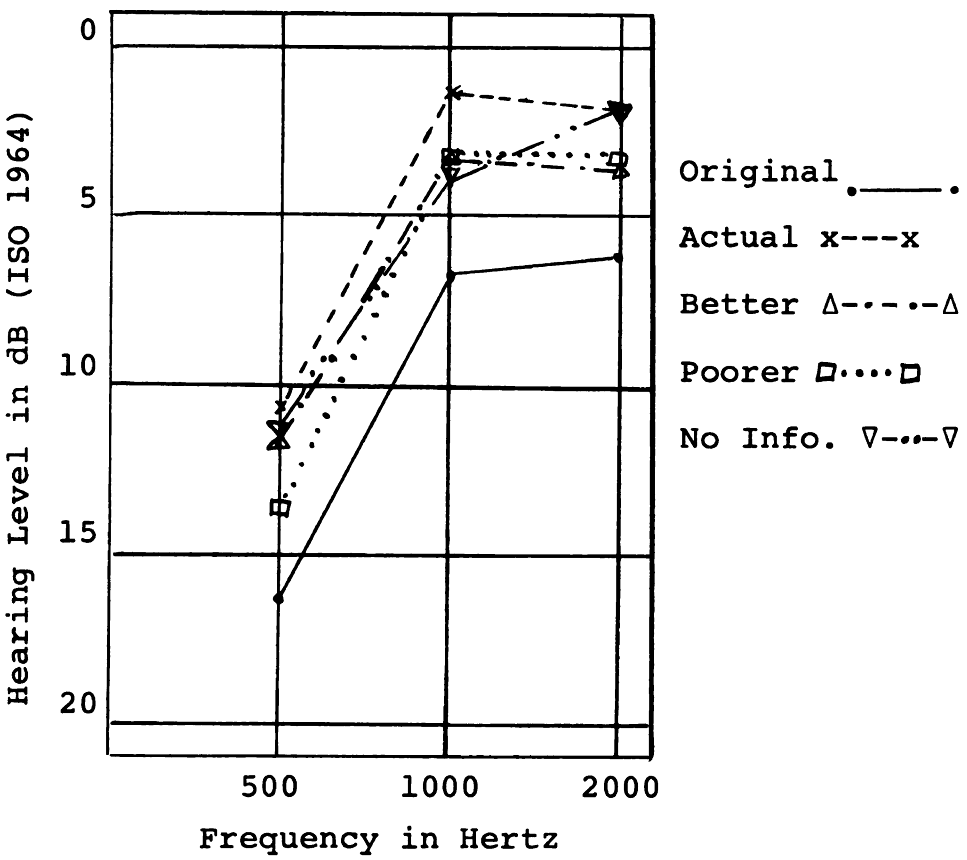


FIGURE 9.--Mean Air-Conduction Thresholds for Eight Normal Hearing Children with Defective Articulation Obtained by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

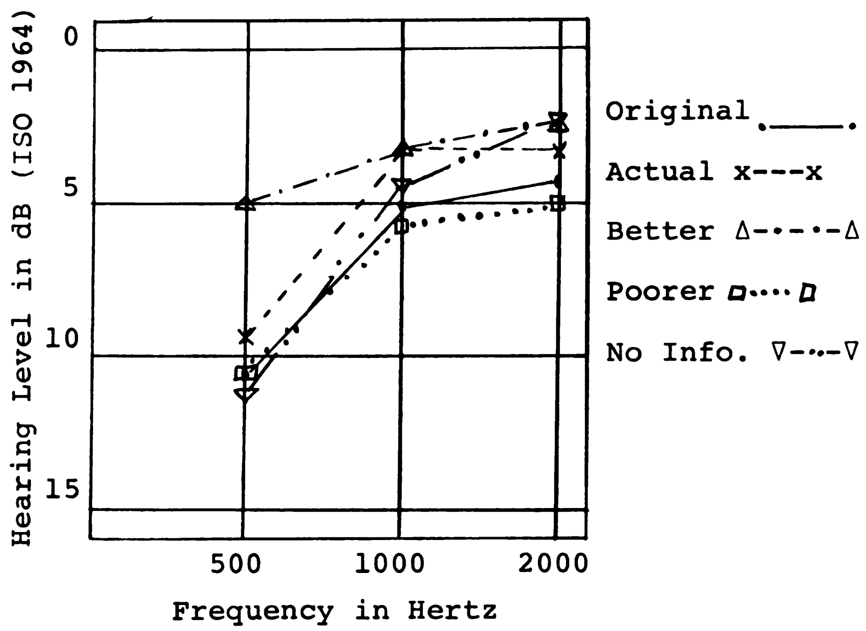


Figure 8 shows a composite audiogram of the mean air-conduction thresholds of the children with normal articulation, while Figure 9 illustrates the same information for the children with defective articulation.

The above figures also show the relatively high degree of repeatability of the mean air-conduction thresholds for three frequencies under four test conditions with the exception at 500 Hz for the actual condition among the children with defective articulation. It should also be noted that the mean original audiometric thresholds for the children with normal articulation are approximately 3 to 4 dB poorer than the thresholds obtained by the testers under the four conditions. The

for this difference is not entirely clear; however, this group of subjects may have shown some improvement between the first and second test sessions because of a learning or practice effect.

Speech Reception Thresholds.--The means, standard deviations, standard errors of measurement and ranges of the speech reception thresholds obtained for eight normal hearing children with normal articulation and for eight normal hearing children with defective articulation by four experienced testers under four conditions are presented in Table IX. The original SRTs obtained for these subjects by the investigator are also reported in the table.

TABLE IX.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Speech Reception Thresholds (SRT)* for Sixteen Normal Hearing Children--Eight with Normal Articulation (N.A.) and Eight with Defective Articulation (D.A.)--by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

		CONDITION			
Subjects	Original	Actual	Better	Poorer	No Info.
N.A.					
M	7.75	7.50	7.00	8.00	6.25
SD	1.56	4.09	3.74	2.82	2.33
SE_m		4.41	3.67	2.34	2.44
Range	4.00	16.00	12.00	8.00	8.00
D.A.					
M	4.50	6.50	6.00	6.50	7.50
SD	1.32	2.49	2.99	3.78	2.77
SE_m		3.46	3.00	4.00	3.80
Range	4.00	10.00	10.00	14.00	12.00

*In dB re audiometric zero (20 dB SPL)

for this difference is not entirely clear; however, this group of subjects may have shown some improvement between the first and second test sessions because of a learning or practice effect.

Speech Reception Thresholds.--The means, standard deviations, standard errors of measurement and the ranges of the speech reception thresholds (SRT) of sixteen normal hearing children with normal articulation and sixteen normal hearing children with defective articulation by four experienced testers under four conditions are presented in Table IX. The obtained SRT's [illegible] of these subjects by the investigator are also [illegible] in the table.

TABLE IX.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Speech Reception Thresholds (SRT)* for Sixteen Normal Hearing Children--Eight with Normal Articulation (N.A.) and Eight with Defective Articulation (D.A.)--by the Investigator and Four Pre-Doctoral Students under Four Conditions.

	CONDITION				
Subjects	Original	Actual	Better	Poorer	No Info.
N.A.					
M	7.75	7.50	7.00	8.00	6.25
SD	1.56	4.09	3.74	2.82	2.33
SE_m		4.41	2.67	2.31	2.44
Range	4.00	16.00	12.00	8.00	8.00
D.A.					
M	4.50	6.50	6.00	6.50	7.50
SD	1.32	2.49	2.99	3.78	2.77
SE_m		3.46	3.00	4.00	3.80
Range	4.00	10.00	10.00	14.00	12.00

*In dB re audiometric zero (20 dB SPL)

The data in the previous table reveal that the variability between the mean speech reception thresholds under all conditions are within the limits of clinical test-retest reliability for spondee words. The SE_m between the conditions indicate that previous audiometric information given to sophisticated testers did not cause them to influence significantly normal hearing children's speech reception thresholds. The only indication observed that suggests some greater variability among the four testers when compared to the investigator is that, in general, for all subjects in all four experimental conditions the ranges for the speech reception thresholds were quite large. No other differences were found. Therefore, according to these results, it seems that previous audiometric information did not produce a clinically significant bias effect when sophisticated testers obtained speech reception thresholds for children even when the children had defective articulation.

Speech Discrimination.--Table X presents the means, standard deviations, standard errors of measurement, and ranges of speech discrimination scores recorded for sixteen normal hearing children. Eight of these children had normal articulation, whereas eight had defective articulation. The original speech discrimination scores recorded for these subjects by the investigator are also reported in this table.

The data in the previous table reveal that the variability between the mean speech reception thresholds under all conditions are within the [illegible] of [illegible] test-retest reliability for spondee words. The SE_m between the conditions indicate that receiving audiometric information given to sophisticated testers did not [illegible] influence significantly normal hearing children's speech reception thresholds. The [illegible] received this suggests some greater variability [illegible] when compared to the investigator's [illegible] for all subjects in all four [illegible] conditions the ranges for the speech reception thresholds were quite large. No other differences were found. Therefore, according to these results, it seems that negative audiometric information did not produce a statistically significant bias effect when sophisticated testers obtained speech reception thresholds for children even when the children had defective articulation.

Speech Discrimination.--Table X presents the means, standard deviations, standard errors of measurement, and ranges of speech discrimination scores recorded for sixteen normal hearing children. Eight of these children had normal articulation, whereas eight had defective articulation. The original speech discrimination scores recorded for these subjects by the investigator are also reported in this table.

TABLE X.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Speech Discrimination Scores* for Sixteen Normal Hearing Children--Eight with Normal Articulation (N.A.) and Eight with Defective Articulation (D.A.)--by the Investigator and Four Pre-Doctoral Students Under Four Conditions.

		CONDITION			
Subjects	Original	Actual	Better	Poorer	No Info.
N.A.					
M	98.50	95.25	98.00	99.00	97.75
SD	1.32	6.92	2.23	1.41	3.38
SE_m		8.09	2.23	2.44	3.39
Range	4.00	20.00	6.00	4.00	10.00
D.A.					
M	92.75	88.75	93.50	93.50	92.50
SD	2.88	9.63	5.41	5.35	5.91
SE_m		11.70	6.00	4.53	7.93
Range	12.00	36.00	20.00	22.00	24.00

*In percent correct.

Table X reveals that, in general, the mean scores for children with normal articulation under all conditions were very comparable. The SE_m between the conditions show essentially the same amount of repeatability which is within the limits of test-retest reliability for speech discrimination scores except for the condition of actual test results. However, the mean score and the SE_m for this condition do not indicate that the testers were biased in the direction of the previous information since under this condition the poorest scores were reported as well as the largest standard errors of measurement. The reason for this exception may, in part, be found by

examining the scores of the individual subjects. Examination of the raw scores show that one tester recorded a speech discrimination score that was twenty percent poorer than recorded by any of the other testers for the same subject. This accounts for the poorer mean score and the larger SE_m but does not explain why this score was obtained.

Table X also shows that the mean scores for children with defective articulation are in very close agreement under all conditions except for the condition of actual test results. This same relationship is revealed also for the SE_m. Again, it is difficult to explain why the greatest amount of variability occurred under the actual condition. It is not felt that it occurred because of the bias phenomenon since the condition of "no information" shows a poorer mean score and a greater amount of variability than either the better or poorer test result conditions. Therefore, the testers did not record speech discrimination scores in the direction of the previous information nor did they record more accurate scores with no information than they did with better or poorer information.

The SE_m, except for the actual condition noted above, vary essentially from 2 to 8%. However, it should be noted that for the group with normal articulation, the SE_m are smaller than 4%, whereas for the group with

defective articulation, the SE_m range from 4.5 to nearly 8%. This greater variability in the speech discrimination scores among the subjects with defective articulation is also reflected in the standard deviations and ranges.

In summary, the descriptive results of this experiment show that the differences obtained between means for pure-tone air-conduction thresholds and speech reception thresholds were within the limits of clinical reliability for both groups of children. The average standard error of measurement for pure-tone air-conduction thresholds for the children with normal articulation was 6.7 dB which is close to the 5 dB test-retest reliability acceptable for pure-tone air-conduction audiometry. The children with defective articulation had an average SE_m of 4.9 dB for pure-tone air-conduction thresholds.

The average standard errors of measurement for speech reception thresholds were 3.21 dB and 3.56 for the children with normal articulation and those with defective articulation respectively. Both of these SE_m are approximately 2 dB better than the test-retest reliability that is clinically acceptable for spondee thresholds.

The average standard errors of measurement for speech discrimination scores for the children with normal articulation was 4.03% which is in very close agreement to the 4% amount of error acceptable in speech discrimination testing. The children with defective articulation

had an average SE_m of 7.3% which is not in good agreement with the test-retest reliability of speech discrimination scores.

From these results it can be seen that skilled audiologists are not biased by previous audiometric results, and they do not clinically influence air-conduction thresholds, speech reception thresholds, or speech discrimination scores obtained from normal hearing children even children who have defective articulation. However, it is evident from the data in Table X that sophisticated testers recorded scores that had more variability when obtaining speech discrimination scores for children with defective articulation than they did when administering this audiometric test to children with normal articulation. This variability seems to be a function of defective articulation rather than a function of the bias phenomenon. In other words, the speech discrimination scores obtained for the children with defective articulation were seemingly based on the testers' personal interpretation of the spoken word rather than on previous test results. Therefore, the sophisticated audiologists were not biased by the previous audiometric information, and they did not influence these subjects' speech discrimination scores.

Statistical Significance

Pure-Tone Air-Conduction Thresholds.--In order to test the differences among mean pure-tone air-conduction thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among mean pure-tone air-conduction thresholds obtained on normal hearing children as a function of the following variables: testers, conditions, and articulation.

A four-factor analysis of variance with repeated measures on one factor was employed to determine the significance of differences among the variables in the principle comparisons for pure-tone air-conduction audiometry.

The analysis of variance for the mean pure-tone air-conduction thresholds was the same design as for the pure-tone thresholds described in Experiment II. Thus, an A x B x C x D four-entry table was used to record the data. The pure-tone air-conduction thresholds were the criterion measures entered in each cell.

The means used in comparing differences between testers, conditions, articulation (normal articulation and defective articulation), and frequencies are shown in Table VIII, page 109, and a summary of the analysis is given in Table XXI of Appendix C.

The only factor that was statistically significant was the frequency factor. Levels of confidence within

testers, within conditions, within articulation and interactions were not statistically significant. Therefore, the above null hypothesis was rejected. The reason for the frequency factor being significant is probably the same in this case as it was in Experiment I. Thus, the testers did not significantly influence the pure-tone air-conduction thresholds obtained under any of the four conditions for either the children with normal or defective articulation.

Speech Reception Thresholds.--In order to test differences among speech reception thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among the mean speech reception thresholds obtained on normal hearing children as a function of the following variables: testers, conditions, and articulation.

The analysis of variance for the mean speech reception thresholds was the same design as for the speech reception thresholds described in Experiment II. An A x B x C three-entry table contained the data recorded for this experiment.

The mean speech reception thresholds used in the comparison are shown in Table IX, page 112. A summary of the analysis is shown in Appendix C, Table XXII.

The F-statistic was not significant for the three principle variables or the interactions among the variables.

testers, within conditions, [illegible] and interactions were not statistically significant. There-fore, the above null hypotheses [illegible] for the frequency factor [illegible] the same in this case as [illegible] the testers did not significantly [illegible] air-conduction thresholds [illegible] conditions for either the [illegible] articulation.

Speech Reception [illegible] differences among speech reception [illegible] this experiment the following [illegible] lated:

> There are no significant differences among the mean speech reception thresholds obtained on normal hearing children as a function of the following variables: testers, conditions, and articulation.

The analysis of variance for the mean speech reception thresholds was the same design as for the speech reception thresholds described in Experiment II. An A x B x C three-entry table contained the data recorded for this experiment.

The mean speech reception thresholds used in the comparison are shown in Table IX, page 117. A summary of the analysis is shown in Appendix C, Table XXII.

The F-statistic was not significant for the three principle variables or the interactions among the variables.

Thus, the above null hypothesis could not be rejected. In other words, the testers did not significantly influence the speech reception thresholds measured under any of the four conditions for either the children with normal or defective articulation.

Speech Discrimination Scores.--In order to test the differences among speech discrimination scores obtained in this experiment, the following null hypothesis was postulated:

> There are no significant differences among the mean speech discrimination scores obtained on normal hearing children as a function of the following variables: testers, conditions and articulation.

The three-way analysis for the mean speech discrimination scores was the same design as for the speech reception thresholds described in Experiment II. The mean speech discrimination scores used in the comparison are shown in Table X, page 114. A summary of the analysis is given in Table XXIII of Appendix C.

Levels of confidence within testers, within conditions and interactions were not statistically significant. However, the factor of articulation was statistically significant. This means that even though testers did not obtain significantly different scores under the four conditions, they did obtain speech discrimination scores from the children with normal articulation that were significantly different from the scores obtained for children with defective articulation. Therefore, based on these

Thus, the above null hypothesis could not be rejected.
In other words, the testers did not significantly influ-
ence the speech reception thresholds obtained under any
of the four conditions for [illegible]
or defective articulation.

Speech Discrimination Scores. [illegible]
differences among speech [illegible]
this experiment, the [illegible]
lated:

> There are no significant [illegible] speech discrimination [illegible] ing children as a function of [illegible] testers, conditions and articulation.

The three-way analysis [illegible]
crimination scores was the same design [illegible]
reception thresholds described in Experiment [illegible]
mean speech discrimination scores [illegible]
are shown in Table X, page [illegible] summary of the analysis
is given in Table XXIII of Appendix D.

Levels of confidence within testers, within condi-
tions and interactions were not statistically significant.
However, the factor of articulation was statistically
significant. This means that even though testers did not
obtain significantly different scores under the four
conditions, they did obtain speech discrimination scores
from the children with normal articulation that were sig-
nificantly different from the scores obtained for children
with defective articulation. Therefore, based on these

findings, the above null hypothesis was rejected. However, this is not to assume that the rejection of the null hypothesis indicates that the testers exhibited a bias because, as shown, the means obtained by the four testers under the four conditions were not statistically significant.

The four-way analysis of variance on the pure-tone air-conduction thresholds and the three-way analyses of variance on the speech reception thresholds and speech discrimination scores computed for this experiment indicate that sophisticated audiologists were not influenced significantly by previous audiometric results. These analyses illustrate that skilled testers obtained mean air-conduction thresholds, speech reception thresholds, and speech discrimination scores that were in close agreement under all conditions.

Based on the statistical results of this experiment, it can be seen that experimenter (tester) bias was not exhibited significantly by sophisticated audiologists when obtaining pure-tone air-conduction thresholds, speech reception thresholds, and speech discrimination scores on normal hearing children with either normal articulation or defective articulation. However, it should be recalled that there was considerably more variability in the speech discrimination scores of the children with defective articulation than among the children with normal articulation.

Experiment IV

In this experiment four speech and hearing therapy undergraduate students were used as testers. Each of them tested sixteen hard-of-hearing adults. Eight of the adults had a sensorineural hearing loss while eight had a conductive hearing impairment. Each tester obtained pure-tone air and bone-conduction thresholds bilaterally under one of four test conditions. Although each subject was tested bilaterally by air and bone-conduction, analysis was made on a single ear. The conditions were the same as those described in Experiment I. The frequencies tested were 250, 500, 1000, 2000, and 4000 Hz.

Clinical Significance

Pure-Tone Air-Conduction Thresholds.--Table XI shows the means, standard deviations, standard errors of measurement, and ranges of the pure-tone air-conduction thresholds recorded for eight adults with a sensorineural hearing loss. Table XII contains the same information for the adult subjects with conductive hearing impairments.

Table XI shows relatively close agreement between the mean pure-tone air-conduction thresholds obtained by the testers under all conditions. The table also reveals that the SE_m are in close agreement with each other under all conditions; however, the SE_m are quite large at 250 Hz under all conditions except for the condition of no

TABL

Cond

Orig
M
SD
Ra

Act
M
S
S
P

Bet
N
S

Po

N

TABLE XI.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Air-Conduction Thresholds* Obtained for Eight Sensorineural Adults by the Investigator and Four Undergraduate Students Under Four Conditions.

	Frequency in Hertz				
Condition	250	500	1000	2000	4000
Original					
M	23.13	25.00	20.63	35.63	63.75
SD	13.67	14.14	12.10	7.26	13.16
Range	50.00	50.00	35.00	20.00	40.00
Actual					
M	20.00	21.15	20.00	33.75	64.38
SD	11.18	10.82	13.69	8.92	14.67
SE_m	8.47	7.07	4.67	4.67	5.86
Range	40.00	40.00	40.00	30.00	40.00
Better					
M	22.50	23.13	19.38	33.75	62.50
SD	13.91	11.43	14.01	8.65	12.99
SE_m	10.75	5.30	4.33	4.67	3.53
Range	45.00	40.00	40.00	25.00	40.00
Poorer					
M	17.50	23.13	20.00	33.13	62.50
SD	11.45	10.58	13.46	9.66	15.81
SE_m	11.85	6.84	3.06	4.33	5.00
Range	35.00	35.00	40.00	30.00	35.00
No Info.					
M	21.88	21.25	20.00	35.00	64.37
SD	8.99	10.82	14.14	12.74	16.28
SE_m	6.08	6.61	3.06	6.84	4.67
Range	35.00	35.00	40.00	40.00	50.00

*In dB re 1964 ISO standard.

information. Although the SE_m were large at 250 Hz, the thresholds recorded at this frequency were not in the direction of the previous test results given to the testers (experimental conditions). Based on these results

it is doubtful that the testers were truly biased by the previous test results.

The mean pure-tone air-conduction thresholds in Table XII are similar under all conditions. The table shows that they are approximately equivalent except for the condition of no information. It was for this condition at all frequencies that the greatest amount of variability was found (note the large SE_m). Again, the data in Table XII illustrate that the mean air-conduction thresholds obtained by the unskilled testers were not in the direction of previous audiometric results given to the testers (experimental conditions). Therefore, according to these results, it was not possible to determine whether the thresholds obtained were affected by the testers' biases. However, the data do indicate that unskilled testers obtain air-conduction thresholds that deviate more from the listeners' true thresholds when the tester has no previous information than when he has true or erroneous test results. This may mean that testers who are not familiar with audiometric testing techniques feel more secure in recording thresholds when they have previous test results to review. Except for the condition of no information, however, the SE_m were essentially within acceptable limits for pure-tone reliability.

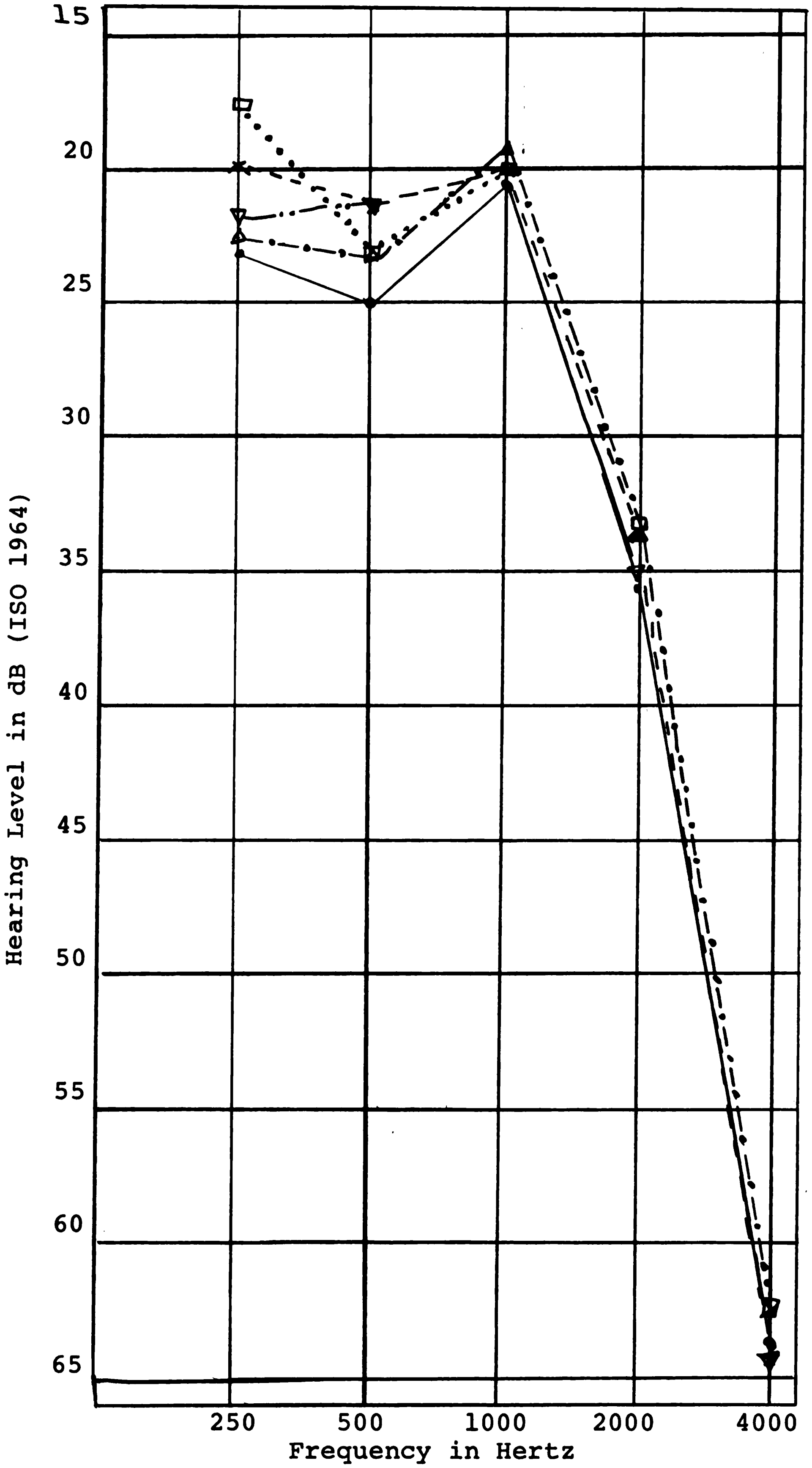
TABLE XII.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Air-Conduction Thresholds* Obtained for Eight Conductive Adult Subjects by the Investigator and Four Undergraduate Students Under Four Conditions.

	Frequency in Hertz				
Condition	250	500	1000	2000	4000
Original					
M	46.85	45.00	33.75	28.75	33.75
SD	15.79	16.95	16.15	14.52	16.15
Range	45.00	55.00	50.00	40.00	50.00
Actual					
M	43.75	41.87	33.75	26.25	31.25
SD	15.56	18.86	20.11	15.36	19.48
SE_m	5.86	5.30	4.33	5.00	5.00
Range	50.00	65.00	60.00	50.00	65.00
Better					
M	41.87	43.80	35.62	25.62	31.87
SD	14.98	18.83	19.11	12.85	18.36
SE_m	6.12	5.00	5.30	7.28	5.30
Range	45.00	55.00	60.00	45.00	55.00
Poorer					
M	43.75	43.75	31.87	24.37	34.37
SD	16.72	16.72	15.59	12.60	15.09
SE_m	5.30	3.53	4.67	6.37	3.06
Range	45.00	55.00	50.00	45.00	45.00
No Info.					
M	39.37	40.00	34.37	23.75	32.50
SD	16.47	18.70	21.99	16.66	18.70
SE_m	9.01	7.50	6.08	9.35	5.59
Range	50.00	65.00	70.00	50.00	50.00

*In dB re ISO 1964

Figures 10 and 11 show composite audiograms of the mean air-conduction thresholds of the adults with sensori-neural hearing disorders and conductive hearing impairments respectively. They illustrate the degree of repeatability

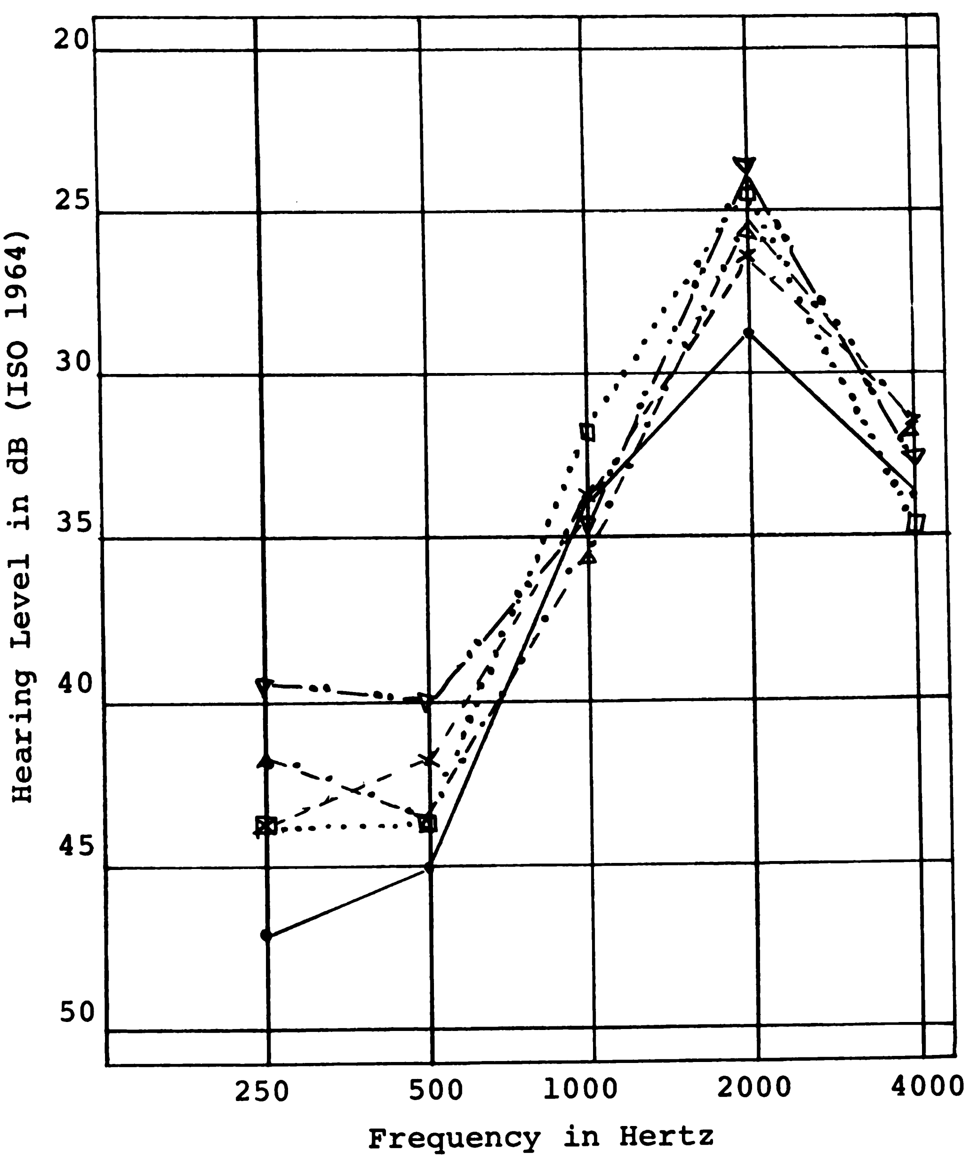
FIGURE 10.--Mean Air-Conduction Thresholds for Eight Adults with a Sensorineural Hearing Disorder Obtained by the Investigator and Four Undergraduate Students Under Four Conditions.



Original •_____• Actual x---x Better Δ-•-•-Δ Poorer □....□
No Info. ∇-••-∇

of the mean air-conduction thresholds for five frequencies under four test conditions. The mean thresholds labeled "original" are those obtained by the investigator.

FIGURE 11.--Mean Air-Conduction Thresholds for Eight Adults with a Conductive Hearing Disorder Obtained by the Investigator and Four Undergraduate Students Under Four Conditions.



Original •——• Actual x---x Better Δ-•-•-Δ Poorer □••••□

No Info. ∇-••-∇

These figures show that for the subjects with a sensorineural loss the greatest amount of variability between means was found in the low frequency region. Further, in general, there was more variability between the mean air-conduction thresholds among the conductive hearing impaired subjects than among the sensorineurals.

Pure-Tone Bone-Conduction Thresholds.--The means, standard deviations, standard errors of measurement, and ranges of the pure-tone bone-conduction thresholds for eight adults with sensorineural hearing disorders are contained in Table XIII. The same information for adults with conductive hearing impairments is listed in Table XIV.

The mean bone-conduction thresholds in Table XIII show very good agreement under all conditions. The SE_m for all frequencies under all conditions are within the limits of test-retest reliability for bone-conduction audiometry. Clinically, this high degree of repeatability indicates that previous audiometric test results do not significantly cause unskilled testers to influence sensorineural adults' pure-tone bone-conduction thresholds. According to these results, it is evident that the previous audiometric information reviewed by the testers did not elicit a bias that affected clinically the repeatability of bone-conduction audiometry.

TABLE XIII.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Bone-Conduction Thresholds* Obtained for Eight Sensorineural Adult Subjects by the Investigator and Four Undergraduate Students Under Four Conditions.

	Frequency in Hertz				
Condition	250	500	1000	2000	4000
Original					
M	18.13	26.25	20.63	31.88	58.75
SD	11.16	10.82	11.02	7.04	7.39
Range	30.00	35.00	35.00	15.00	20.00
Actual					
M	16.25	26.25	18.75	31.88	58.75
SD	8.56	9.27	10.53	8.63	8.56
SE_m	5.30	4.33	3.06	4.33	2.50
Range	30.00	30.00	35.00	20.00	20.00
Better					
M	16.25	26.25	21.25	31.88	58.75
SD	7.80	9.92	12.18	9.66	7.39
SE_m	5.68	2.50	3.95	5.00	0.00
Range	25.00	35.00	35.00	25.00	20.00
Poorer					
M	15.00	29.38	17.50	32.50	60.63
SD	7.50	9.16	10.00	10.89	6.34
SE_m	5.68	4.67	5.30	6.37	3.06
Range	25.00	30.00	30.00	30.00	15.00
No Info.					
M	17.50	29.38	18.75	33.13	57.50
SD	6.61	9.82	8.56	7.88	9.35
SE_m	5.30	4.67	3.95	4.33	3.53
Range	25.00	30.00	25.00	20.00	25.00

*In dB re ISO 1964

TABLE XIV.--Means (M), Standard Deviations (SD), Standard Errors of Measurement (SE_m), and Ranges of Pure-Tone Bone-Conduction Thresholds* Obtained for Eight Conductive Adult Subjects by the Investigator and Four Undergraduate Students Under Four Conditions.

	Frequency in Hertz				
Condition	250	500	1000	2000	4000
Original					
M	5.62	17.50	12.50	6.25	15.62
SD	5.26	9.66	9.01	9.60	15.29
Range	15.00	25.00	30.00	30.00	45.00
Actual					
M	3.12	14.37	11.25	4.27	13.75
SD	3.47	9.16	7.80	9.16	13.16
SE_m	4.33	4.67	4.33	4.67	6.37
Range	10.00	30.00	25.00	30.00	40.00
Better					
M	6.87	18.12	15.00	5.00	15.00
SD	5.55	8.26	9.35	10.00	12.50
SE_m	4.33	3.95	5.00	3.53	4.67
Range	15.00	25.00	35.00	35.00	40.00
Poorer					
M	5.00	18.75	16.25	6.87	17.50
SD	5.00	8.92	7.90	12.97	12.24
SE_m	4.33	4.33	6.61	4.67	7.70
Range	15.00	30.00	30.00	35.00	40.00
No Info.					
M	3.75	15.00	13.75	8.12	16.87
SD	5.44	9.35	9.60	11.97	12.97
SE_m	4.67	5.59	5.59	6.37	7.07
Range	15.00	25.00	30.00	35.00	45.00

*In dB re ISO 1964

The data in Table XIV are comparable to the thresholds shown in Table XIII. The SE_m are within the limits of bone-conduction reliability and are approximately the same under all conditions. Based on these results, it can be seen that inexperienced testers were not influenced significantly by the previous audiometric test results. It can be assumed from the data in Table XIV that previous audiometric information did not produce a bias that was clinically observable.

The following figures are composite audiograms of mean bone-conduction thresholds. Figure 12 gives the mean thresholds of the sensorineural subjects under all conditions obtained by four undergraduate students and the investigator of this study. Figure 13 shows the same data for the conductive subjects.

These figures show that for the conductive hearing impaired subjects the greatest amount of variability between means was found for the frequencies of 500, 1000, and 2000 Hz. Also, there was more variability between the mean bone-conduction thresholds among the conductive hearing impaired subjects than among the sensorineurals.

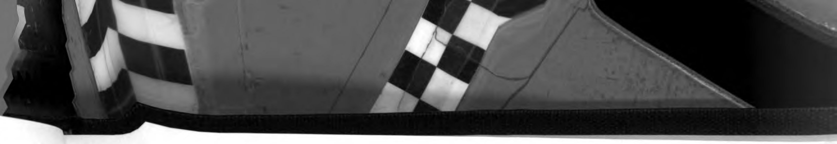


FIGURE 12.--Mean Bone-Conduction Thresholds for Eight Adults with a Sensorineural Hearing Disorder Obtained by the Investigator and Four Undergraduate Students Under Four Conditions.

s for Eight
earing Discri
and Four Under-
Conditions.

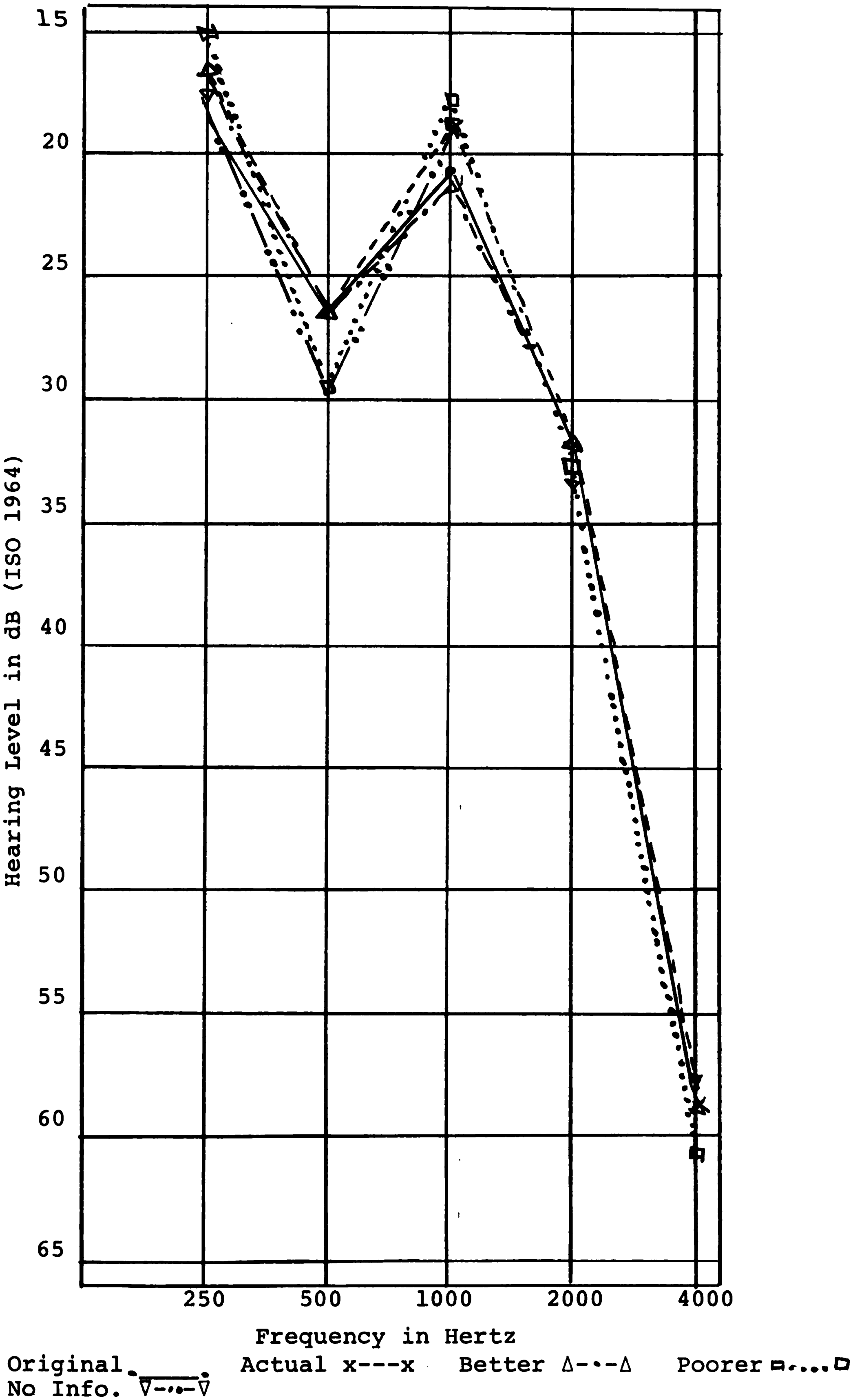
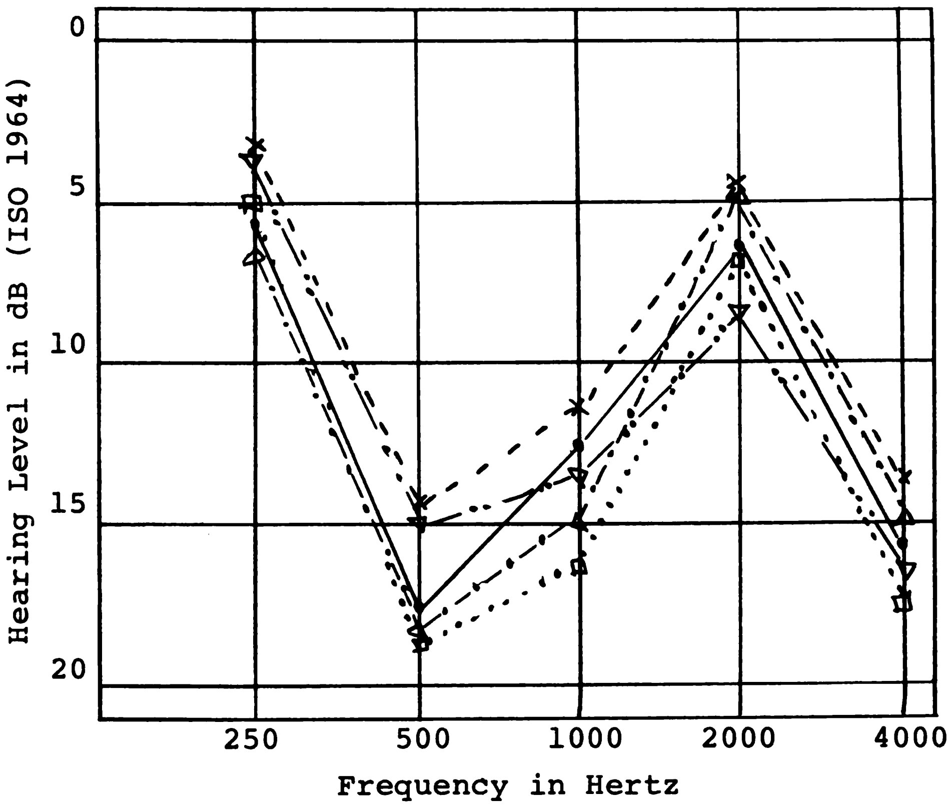


FIGURE 13.--Mean Bone-Conduction Thresholds for Eight Adults with a Conductive Hearing Disorder Obtained by the Investigator and Four Undergraduate Students Under Four Conditions.



Original •———• Actual x---x Better Δ-•-•-Δ Poorer □····□
No Info. ∇-••-∇

The descriptive results of this experiment show that the standard errors of measurement for pure-tone air and bone-conduction thresholds obtained by four unsophisticated testers under four conditions are within the limits of clinical reliability for both pathological groups. The average standard errors of measurement of air-conduction thresholds for the eight sensorineural subjects and for the eight conductive subjects were 5.9 dB and 5.7 dB, respectively. Both of the SE_m are approximately the same as the 5 dB acceptable amount of error allowed in air-conduction audiometry. The average SE_m of bone-conduction thresholds under all conditions was 4.2 dB for the subjects with sensorineural hearing disorders and 5.1 dB for the subjects with conductive hearing impairments. Both of these SE_m are within the limits of bone-conduction test-retest reliability. Although it cannot be thoroughly explained, the bone-conduction thresholds for both pathological groups showed less variability than the air-conduction thresholds for the same groups. However, Carhart and Hayes reported that bone-conduction audiometry is as reliable or more so than air-conduction audiometry.[1]

According to these results, it seems that unskilled testers were no more influenced by previous audiometric results when testing pathological ears than were sophisticated audiologists. This is not to say that unskilled testers are as adept in obtaining pure-tone thresholds that are as clinically accurate as those obtained by sophisticated audiologists; however, within the limitations of this study, the unskilled testers demonstrated

---

[1] Carhart and Hayes, "Clinical Reliability of Bone Conduction," pp. 1084-1101.

tha

es

by

St

te

th

hy

s

t

c

(

that they were capable of obtaining thresholds that were essentially as free from errors as the thresholds obtained by the skilled audiologists.

Statistical Significance

Pure-Tone Air-Conduction Thresholds.--In order to test the differences among pure-tone air-conduction thresholds obtained in this experiment the following null hypothesis was postulated:

> There are no significant differences among pure-tone air-conduction thresholds obtained for hard-of-hearing adults as a function of the following variables: testers, conditions, pathologies, and frequencies.

A four-factor analysis of variance that was the same design as the one described for the air-conduction thresholds in Experiment II was used to test the significance of differences among the variables in the principle comparison for pure-tone air-conduction audiometry. The air-conduction thresholds were the criterion measures entered in each cell.

The means used in comparing differences between testers, conditions, frequencies, and pathologies are contained in Tables XI and XII on pages 123 and 125 respectively. A summary of the analysis is contained in Table XXIV of Appendix D.

The factors of pathology and frequency and the interaction between these were statistically significant. Levels within testers and within conditions were not

si
di
co
n
t
t
t
d
i
l

significant. This indicates that even though the testers did not obtain different thresholds as a function of the conditions, they did obtain thresholds from the sensori-neural subjects that were significantly different from the thresholds obtained from the conductive subjects at the frequencies tested. Thus, according to these results, the above null hypothesis was rejected. However, this does not mean that the rejection of the null hypothesis indicates that the testers were influenced significantly by the previous test results. As previously mentioned, the means obtained by the four testers under the four conditions were not statistically significant within each pathology under each condition. Therefore, these data indicate that the testers did not obtain different air-conduction thresholds as a function of conditions.

Pure-Tone Bone-Conduction Thresholds.--In order to test the differences among pure-tone bone-conduction thresholds obtained in the experiment the following null hypothesis was postulated:

> There are no significant differences among pure-tone bone-conduction thresholds obtained for hard-of-hearing adults as a function of the following variables: testers, conditions, pathologies, and frequencies.

A four-factor analysis of variance that was the same design as the one described for the air-conduction thresholds in Experiment II was used to test the significance of differences among the variables. The bone-conduction thresholds were the criterion measures entered in each cell.

The means used in comparing differences between testers, conditions, frequencies, and pathologies are contained in Tables XIII and XIV on pages 129 and 130 respectively. A summary of the analysis is listed in Table XXV of Appendix D.

The factors of pathology and frequency and the interactions between these were statistically significant. Levels within testers, within conditions, and interactions between them were not statistically significant. These results indicate that testers did not obtain thresholds that were significantly different as a function of the conditions; however, the testers did obtain thresholds from the sensorineural subjects that were significantly different from the thresholds obtained from the conductive subjects at the frequencies tested. Therefore, based on these results the above null hypothesis was rejected; however, the variables that were significant do not indicate that the testers were influenced significantly by the previous test results. As mentioned earlier, the means obtained by the four testers were not statistically significant within each pathology for the four conditions. Therefore, these data indicate that the testers did not obtain different bone-conduction thresholds as a function of conditions. According to the statistical results of this experiment, it is evident that tester bias was not exhibited significantly by unskilled testers when obta

harc

ind

by

auc

su

ic

T

e

m

obtaining pure-tone air and bone-conduction thresholds on hard-of-hearing subjects. Statistically these results indicate that unskilled testers were no more influenced by previous audiometric information than were skilled audiologists.

General Discussion

It is felt that it is appropriate at this time to subjectively analyze the reasons for not obtaining clinical or statistical significance in this investigation. This seems warranted especially since the phenomena of experimenter bias and effect have been observed experimentally in a variety of areas.

1. This was an extremely structured study that limited the testers in making subjective judgments. The testers were restricted to obtaining audiometric scores and did not have the opportunity of giving the subjects instructions or placing the transducers on the subjects. The testers were given specific instructions on how to obtain pure-tone air and bone-conduction thresholds, speech reception thresholds, and speech discrimination scores. Therefore, if they were biased by the previous audiometric results, they were limited in their opportunity to deviate from the required techniques in establishing audiometric scores that were in accord with the previous test results.

obtaining pure-tone air and bone-conduction thresholds on hard-of-hearing subjects. Statistically these results indicate that unskilled testers were no more influenced by previous audiometric information than were skilled audiologists.

General Discussion

It is felt that it [illegible] to subjectively analyze the results of this study [illegible] logical or statistical [illegible] this investigation. This seems warranted especially since the phenomenon of experimenter bias and effect has been [illegible] experimentally in a variety of areas.

1. This was an extremely structured study that limited the testers in making subjective judgments. The testers were restricted to obtaining audiometric scores and did not have the opportunity of giving the subjects instructions or placing the transducers on the subjects. The testers were given specific instructions on how to obtain pure-tone air and bone-conduction thresholds, speech reception thresholds, and speech discrimination scores. Therefore, if they were biased by the previous audiometric results, they were limited in their opportunity to deviate from the required techniques in establishing audiometric scores that were in accord with the previous test results.

The only time that the testers had an opportunity to make a subjective judgment was when obtaining speech audiometry scores. However, the only testers that obtained speech audiometry scores were the skilled audiologists who seemingly were more concerned in obtaining good agreement between pure-tone audiometric results and speech audiometric results than in obtaining scores that were in accord with the previous test results.

2. Each tester knew that he was participating in a thesis project and, in all probability, was extremely careful in obtaining audiometric scores that were as accurate as he could obtain.

3. Even though the testers were instructed to study the previous test results and were given ample time to review them, there was no measurable means of determining whether the testers reviewed the scores on the audiograms or merely scanned the audiograms to determine which audiometric tests had been given or which audiometric test they were to replicate.

4. It seemed that the testers were more concerned about adhering to the procedures they were to follow than with the previous audiometric information. This was particularly evident when masking was employed. Each tester adhered strictly to the procedures given to him when obtaining masked thresholds and recorded the subjects' thresholds only after following the masking procedure given to him.

It should be pointed out however, that if a tester does not employ an audiometric procedure that is highly structures in defining a threshold or discrimination score, it is conceivable that the tester may be biased by previous test result information.

In summary it is felt that the non-significances found in this investigation were meaningful based on the findings of other studies. The results illustrated that the bias phenomenon is not observable when making audiometric assessments while following stringent clinical procedures. Therefore, based on these results, the audiologist should review any pertinent audiometric test results that the patient may have had and not be unduly concerned that these previous test results may influence his audiometric findings.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of this investigation was to determine whether certain types of previous audiometric information caused the tester to influence pure-tone and speech audiometric results. Specifically, this study was concerned with the possibility of testers' expectancy affecting audiometric responses obtained from normal hearing adults, hard-of-hearing adults having either conductive or sensori-neural hearing disorders, and normal hearing children having either normal or defective articulation.

Summary

This study was composed of four individual experiments. In each experiment four testers were employed to administer pure-tone or pure-tone and speech audiometric tests to adults and children under four test conditions. The adult subjects were selected on the basis of their hearing sensitivity, whereas the children were selected on the basis of their ability to articulate English phonemes. The four experimental conditions were as follows: actual test results, better than actual test

CHAPTER [illegible]

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of this investigation was to determine whether certain types of [illegible] caused the tester to influence pure-tone and speech audiometric results. Specifically, this study was concerned with the possibility of testers' expectancy influencing audiometric responses of normal hearing adults, hard-of-hearing adults having conductive or sensori-neural hearing disorders, and normal hearing children having either normal or defective articulation.

[illegible] Summary

This study was conducted in the form of three experiments. In each experiment four testers were employed to administer pure-tone or speech audiometric tests to adults and children under four test conditions. The adult subjects were selected on the basis of their hearing sensitivity, whereas the children were selected on the basis of their ability to articulate English phonemes. The four experimental conditions were as follows: actual test results, better than actual test results, poorer than actual test results, and no test results.

The testers were divided into two groups. One group was composed of four pre-doctoral students majoring in audiology. The second group consisted of four under-graduate students majoring in speech and hearing therapy. These testers were classified as being sophisticated (skilled) audiologists and unsophisticated (unskilled) testers respectively.

A summary of each experiment is given below.

Experiment I

Conditions.--The four experimental conditions were: actual test results, better than actual test results, poorer than actual test results, and no test results.

Testers.--Four pre-doctoral students in audiology were employed as testers.

Subjects.--The subjects were eight normal hearing adults ranging in age from twenty to twenty five years with a mean age of 21.4 years and a median age of 20.5 years. All of the subjects were female with the exception of one.

Tests Administered.--One ear of each subject was tested by each tester by pure-tone air-conduction audiometry at the following frequencies: 250, 500, 1000, 2000, 4000, and 8000 Hz. This was followed by obtaining a speech reception threshold on the same ear.

Experiment II

<u>Conditions</u>.--The same four conditions were employed in this experiment as those in Experiment I.

<u>Testers</u>.--The testers were the same as those used in Experiment I.

<u>Subjects</u>.--The subjects consisted of sixteen hard-of-hearing adults. Eight subjects, six males and two females, had a sensorineural hearing loss. These subjects ranged in age from forty to sixty-three years with a mean age of 47.0 years and a median age of 43.5 years. The remaining eight subjects, six females and two males exhibited a conductive hearing impairment. These subjects ranged in age from eighteen to fifty-two years with a mean age of 35.6 years and a median age of 32.0 years.

<u>Tests Administered</u>.--Each of the subjects in this experiment was tested bilaterally by air-conduction at the frequencies of 500, 1000, and 2000 Hz and one ear was tested by bone-conduction at the same frequencies. Speech audiometry (speech reception thresholds and speech discrimination scores) was administered to the same ear that was tested by bone-conduction.

Experiment III

<u>Conditions</u>.--The conditions in this experiment were the same four as those utilized in Experiments I and II.

Testers.--The same four sophisticated testers participated in this experiment as in Experiments I and II.

Subjects.--The subjects were sixteen normal hearing children. Eight of these children, five females and three males, ranged in age from five to seven years with a mean age of 5.8 years and a median age of 5.0 years. These subjects were judged to have normal speech articulation. The remaining eight children, six males and two females, ranged in age from five to nine years with a mean age of 6.3 years and a median age of 6.0 years. These children had defective articulation characterized by substitutions, omissions, and/or distortions.

Tests Administered.--Each of these subjects was given a pure-tone air-conduction test at 500, 1000, and 2000 Hz for one ear and speech audiometry (speech reception threshold and speech discrimination) was administered to the same ear that was tested by air-conduction.

Experiment IV

Conditions.--The conditions used in this experiment were the same four as those employed in Experiments I, II, and III.

Testers.--The testers were four undergraduate students who were majoring in speech and hearing therapy. Each of these testers was enrolled in an introductory

audiology course at the time of this study. Thus, they were unsophisticated testers in comparison to the pre-doctoral students.

Subjects.--Sixteen hearing impaired adults served as subjects. Eight of the subjects, two females and six males, were diagnosed as having a sensorineural hearing loss. They ranged in age from thirty-seven to sixty-five years with a mean age of 50.3 years and a median age of 49.5 years. Eight of the subjects exhibited a conductive hearing impairment. All of these subjects were females with the exception of one and they ranged in age from sixteen to thirty-five years with a mean age of 31.4 years and a median age of 30.0 years.

Tests Administered.--Each subject was tested by pure-tone air and bone-conduction audiometry at the following frequencies: 250, 500, 1000, 2000, and 4000 Hz.

Conclusions

Within the limits imposed by the design of this study, the following conclusions appear warranted.

1. There are no significant differences between the pure-tone air-conduction thresholds obtained by sophisticated and unsophisticated testers as a function of the type of previous audiometric information (true, erroneous, of no information) reviewed by these testers. However, there are significant differences in the pure-tone

air-conduction thresholds obtained by these testers as a function of the subjects' age and hearing sensitivity (adults with normal hearing, adults with a sensorineural hearing loss, adults with a conductive hearing loss and children with normal hearing). Sophisticated testers obtain pure-tone air-conduction thresholds for adults, regardless of the subjects' hearing sensitivity, with essentially the same degree of repeatability. However, these same testers obtain pure-tone air-conduction thresholds from children with normal hearing that are slightly less reliable than the same thresholds obtained for normal hearing adults. Further, although the unskilled testers obtain pure-tone air-conduction thresholds that reveal a relatively high degree of repeatability, regardless of the subjects' hearing sensitivity, the degree of repeatability is slightly less than it is for the sophisticated testers.

2. There are no significant differences between pure-tone bone-conduction thresholds obtained by sophisticated and unsophisticated testers as a function of the type of previous audiometric information (true, erroneous, or no information) reviewed by these testers. However, there are significant differences in the pure-tone bone-conduction thresholds obtained by these testers as a function of the subject's hearing sensitivity (adults with a sensorineural hearing disorder and adults with a conductive hearing impairment). Sophisticated testers

obtain pure-tone bone-conduction thresholds for adults, regardless of the subject's hearing sensitivity, with essentially the same degree of repeatability. The unsophisticated testers obtain pure-tone bone-conduction thresholds that reveal good test-retest repeatability, regardless of the subject's hearing sensitivity; however, this degree of reliability is slightly less than it is for the sophisticated testers.

3. There are no significant differences between the speech reception thresholds obtained by sophisticated testers as a function of the type of previous audiometric information (true, erroneous, or no information) reviewed by these testers. However, there are significant differences in speech reception thresholds obtained as a function of the subject's hearing sensitivity (normal hearing adults, adults with a sensorineural hearing loss, and adults with a conductive hearing loss). Sophisticated testers tend to obtain speech reception thresholds that show a higher degree of repeatability for adults with normal hearing or a conductive hearing loss than they do for subjects with a sensorineural hearing loss. Also, there tends to be a slightly higher degree of repeatability for normal hearing adults than for normal hearing children. Further, there are significant differences in speech reception thresholds obtained as a function of the subject's ability to articulate English phonemes (normal hearing children having

normal articulation and children with normal hearing having defective articulation). These same testers obtain speech reception thresholds that reveal slightly less variability for children with normal articulation than for children with defective articulation.

4. There are no significant differences between the speech discrimination scores obtained by sophisticated testers as a function of the type of previous audiometric information (true, erroneous, or no information) reviewed by these testers. However, there are significant differences in speech discrimination scores obtained as a function of the subject's age and hearing sensitivity and ability to articulate English Phonemes (adults with a sensorineural hearing loss, adults with a conductive hearing loss, children with normal hearing having normal articulation, and children with normal hearing having defective articulation). Sophisticated audiologists obtain discrimination scores that reveal greater variability for adults with a sensorineural hearing disorder than they do for adults with a conductive hearing loss or children with normal speech articulation. These same testers obtain speech discrimination scores that show slightly greater variability for children with defective articulation than for children with normal articulation.

5. Essentially all of the audiometric results obtained in this study by the four sophisticated

audiologists and by the four unskilled testers under four experimental conditions (actual test results, better than actual test results, poorer than actual test results, and no test results) show a high degree of test-retest reliability. In other words, the testers employed in this study obtained audiometric scores that were clincally and statistically reliable, regardless of the type of previous information concerning the subjects. Therefore, having previous test result information failed to elicit an observable bias among the testers regardless of their academic or clinical experience.

Recommendations for Further Research

The present study should be repeated in a less structured situation. The testers should have the opportunity of giving the subjects instructions, placing the transducers on the subjects, and obtaining audiometric scores in any manner they deem acceptable. This may cause the testers to give the subjects instructions that would cause the subjects to respond in a manner that would be more in accord with the testers' expectancies.

BIBLIOGRAPHY

Books

Blalock, Hubert M. *Social Statistics*. New York: McGraw-Hill Company, Inc., 1963.

Chaiklin, Joseph B. and Ventry, Ira M. "Functional Hearing Loss," *Modern Developments in Audiology*, James Jerger, editor. New York: Academic Press, Inc., 1963.

O'Neill, John J. and Oyer, Herbert J. *Applied Audiometry*. New York: Dodd, Mead & Company, Inc., 1966.

Rosenthal, Robert. *Experimenter Effects in Behavioral Research*. New York: Appleton-Century-Crofts, 1966.

Ward, Dixon. "Auditory Fatigue and Masking," *Modern Developments in Audiology*. James Jerger, editor. New York: Academic Press, Inc., 1963.

Whyte, William Foote. *Street Corner Society*. Chicago: University of Chicago Press, 1943.

Winer, B. J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill Book Company, Inc., 1962.

Reports

Harris, J. Donald, and Myers, Cecil. "Experiments on Fluctuations of Auditory Acuity," Medical Research Laboratory U.S. Naval Submarine Base, New London, Connecticut. Volume XI No. 13, Report No. 196 (22 June 1952) Project NM 003 041.21.08.

Tillman, Tom W., and Carhart, Raymond. "An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words: Northwestern University Auditory Test No. 6 USAF School of Areospace Medicine, Areospace Medical Division (AFSC) Brooks Air Force Base, Texas (1966), 1-12.

Periodicals

Back, Kurt W. "Influence Through Social Communication," Journal of Abnormal and Social Psychology, 46 (1951), 9-23.

Bellak, Leopold. "The Concept of Projection: An Experimental Investigation and Study of the Concept," Psychiatry, 7 (1944), 353-370.

Blakenship, Albert B. "The Effect of the Interviewer Upon the Response In a Public Opinion Poll," Journal of Consulting Psychology, 4 (1940), 134-136.

Carhart, Raymond. "Individual Differences in Hearing for Speech," Annals of Otology, Rhinology, and Laryngology, 55 (1946), 223-267.

Carhart, Raymond. "Speech Reception in Relation to Pattern of Pure Tone Loss," Journal of Speech Disorders, 11 (1946), 97-108.

Carhart, Raymond. "Problems in the Measurement of Speech Discrimination," Archives of Otolaryngology, 82 (1965), 253-260.

Carhart, Raymond and Hayes, Claude. "Clinical Reliability of Bone Conduction Audiometry," Laryngoscope, 50 (1940), 1084-1101.

Carhart, Raymond and Jerger, James. "Preferred Method of Determination of Thresholds," Journal of Speech and Hearing Disorders, 24 (1959), 330-345.

Chaiklin, Joseph B. and Ventry, Ira M. "Spondee Threshold Measurement: A Comparison of 2- and 5- dB Methods," Journal of Speech and Hearing Disorders, 29 (1964), 47-59.

Cochran, William G. and David J. Watson. "An Experiment on Observer's Bias in the Selection of Shoot-Heights," Empire Journal of Experimental Agriculture, 4 (1936), 69-76.

Corado, Lucion and Ison, James R. "Observer Bias in Conditioning of Planarian," Psychological Reports, 13 (1963), 787-789.

Fletcher, Harvey. "Auditory Patterns," Review of Modern Physics, 12 (1940) 47-65.

Fletcher, Harvey. "A Method of Calculating Hearing Loss for Speech from an Audiogram," Journal of the Acoustical Society of America, 22 (1950), 1-5.

Fletcher, Harvey and Munson, W. A. "Relation Between Loudness and Masking," Journal of the Acoustical Society of America, 9 (1937), 1-10.

Freidman, Neil, Kurland, Daniel, and Rosenthal. "Experimenter Behavior as an Unintended Determinant of Experimental Results," Journal of Projective Techniques, 29 (1965), 479-490.

Hanson, Robert H., and Marks, Eli S. "Influence of the Interviewer on the Accuracy of Survey Results," Journal of the American Statistical Association, 53 (1958), 635-655.

Harvey, S. M. "A Preliminary Investigation of the Interview," British Journal of Psychology, 28 (1938), 263-287.

High, Wallace S., Glorig, Aram, and Nixon, James. "Estimating the Reliability of Auditory Threshold Measurements," Journal of Auditory Research, 1 (1961), 247-262.

Hirsh, Ira, Davis, Hallowell, Silverman, S. Richard, Reynolds, Elizabeth G., Eldert, Elizabeth, Benson, Robert. "Development of Materials for Speech Audiometry," Journal of Speech and Hearing Disorders, 17 (1952), 321-337.

Hood, J. D. "Principles and Practice of Bone Conduction Audiometry," Laryngoscope, 70 (1960), 1211-1228.

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

Jerger, James F. "Comparative Evaluation of Some Auditory Measures," Journal of Speech and Hearing Research, 5 (1962), 3-17.

Jerger, James F., and Carhart, Raymond, Tillman, Ton W. "Some Relations Between Normal Hearing for Pure Tones and for Speech," Journal of Speech and Hearing Research, 2 (1959), 126-140.

Lehiste, I. and Peterson, G. E. "Linguistic Considerations in the Study of Speech Intelligiliby," Journal of the Acoustical Society of America, 31 (1959), 280-286.

Lord, Edith. "Experimentally Induced Variations in Rorschach Performance," Psychological Monographs, 64 (1961), 321-333.

Luft, Joseph. "Interaction and Projection," Journal of Projective Techniques, 17 (1953), 489-492.

Lybarger, S. F. "Interim Bone Conduction Thresholds for Audiometry," Journal of Speech and Hearing Research, 9 (1966), 483-487.

Masling, Joseph. "The Effects of Warm and Cold Interaction on the Administration and Scoring of an Intelligence Test," Journal of Consulting Psychology, 23 (1959), 336-341.

Menzel, Otto J. "Error in Audiometry," Eye Ear Nose and Throat Monthly, 42 (1963), 74.

Peterson, G. E. and Lehiste, I. "Revised CNC Lists for Auditory Tests," Journal of Speech and Hearing Disorders, 27 (1962), 62-70.

Rhum, Howard B. and Carhart, Raymond. "Objective Speech Audiometry," Journal of Speech and Hearing Research, 1 (1958), 169-178.

Rintelmann, William F. amd Harford, Earl R. "Type V Bekesy Pattern: Interpretation and Clinical Utility," Journal of Speech and Hearing Research, 10 (1967), 733-744.

Rosenthal, Robert. "Experimenter Attributes as Determinants of Subjects' Responses," Journal of Projective Techniques, 27 (1963), 324-331.

Rosenthal, Robert. "Experimenter Modeling Effects as Determinants of Subjects' Responses," Journal of Projective Techniques, 27 (1963), 467-471.

Rosenthal Robert and Fode, Kermit L. "The Effects of Experimenter Bias on the Performance of the Albino Rat," Behavioral Science, 8 (1963), 183-189.

Rosenthal, Robert, Kohn, Paul, Greenfield, Patricia M., and Carota, Noel. "Data Desirability, Experimenter Expectancy, and the Results of Psychological Research," Journal of Personality and Social Psychology, 3 (1966), 20-27.

Rosenthal, Robert and Lawson, Reed. "A Longitudinal Study of the Effects of Experimenter Bias on the Operant Learning of Laboratory Rats," Journal of Psychiatric Research, 2 (1964), 61-72.

Rosenthal, Robert, Murly, Ray C., Persinger, Gordon W., and Vikan-Kline, Linda. "Changes in Experimental Hypotheses as Determinants of Experimental Results," Journal of Projective Techniques, 28 (1964), 465-469.

Sanders, Jay W. and Rintelmann, William F. "Masking in Audiometry," Archives of Otolaryngology, 80 (1964), 541-556.

Sarason, Irwin G. "Individual Differences, Situational Variables, and Personality Research," Journal of Abnormal and Social Psychology, 23 (1959), 336-341.

Sataloff, Joseph. "Pitfalls in Routine Hearing Tests," Archives of Otolaryngology, 73 (1961), 717-726.

Shapiro, Arthur. "A Contribution to a History of the Placebo Effect," Behavioral Science, 5 (1960), 109-135.

Stream, Richard W. and McConnell, Freemann. "A Comparison of Two Methods of Administration in Bekesy-Type Audiometry," Journal of Auditory Research, 4 (1961), 263-271.

Studebaker, Gerald A. "Clinical Masking of Air- and Bone-Conducted Stimuli," Journal of Speech and Hearing Disorders, 29 (1964), 23-35.

Tillman Tom W. and Jerger, James F. "Some Factors Affecting the Spondee Threshold in Normal-Hearing Subjects," Journal of Speech and Hearing Research, 2 (1959), 141-146.

Tillman, Tom W., Johnson, Robert M. and Olsen, Wayne O. "Earphone versus Sound Field Threshold Sound Pressure Levels for Spondee Words," Journal of the Acoustical Society of America, 39 (1966), 125-133.

Witting , E. G. and Hughson, Walter. "Inherent Accuracy of a Series of Repeated Clinical Audiograms," Laryngoscope, 50 (1940), 259-269.

Unpublished Material

Haskins, Harriet. "A Phonetically Balanced Test of Speech Discrimination for Children" (Unpublished Master's Thesis, Northwestern University, 1949).

Rintelmann, William F. and Jetty, Albert J. "Reliability of Speech Discrimination Testing Using CNC Monosyllabic Words" (Unpublished Study, Michigan State University, 1968).

Tokay, F. Harry. "Validity and Reliability of Bekesy Audiometry with Preschool Age Children" (Unpublished Doctor's Thesis, Michigan State University, 1967).

Other Sources

________. "American Standard Specification for Audiometers for General Diagnostic Purposes," American Standards Association, Incorporated, No. S3-3-1960.

________. "Operating Manual for the Beltone Artificial Mastoid, M5 Series," Beltone Electronics Corporation, Chicago, Illinois, 1968.

APPENDICES

TA

Sc

A

B

Al

Z

C

A

B

A

Z

T

APPENDIX A

RESULTS OF PURE-TONE AIR-CONDUCTION THRESHOLDS

TABLE XV.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), and Frequency (C) for Eight Normal Hearing Adults by Pure-Tone Air-Conduction Audiometry.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	32.682	3	10.894	0.171
B	70.182	3	23.394	0.368
AB	807.421	9	89.713	1.414
Z=D+AD+BD+ABD	1014.583	16	63.411	
C	2833.463	5	566.692	16.059*
AC	143.098	15	9.539	0.270
BC	93.098	15	6.206	0.175
ABC	869.921	45	19.331	0.547
Z=CD+ACD+BCD+ABCD	2822.916	80	35.286	
TOTAL	8687.369	191		

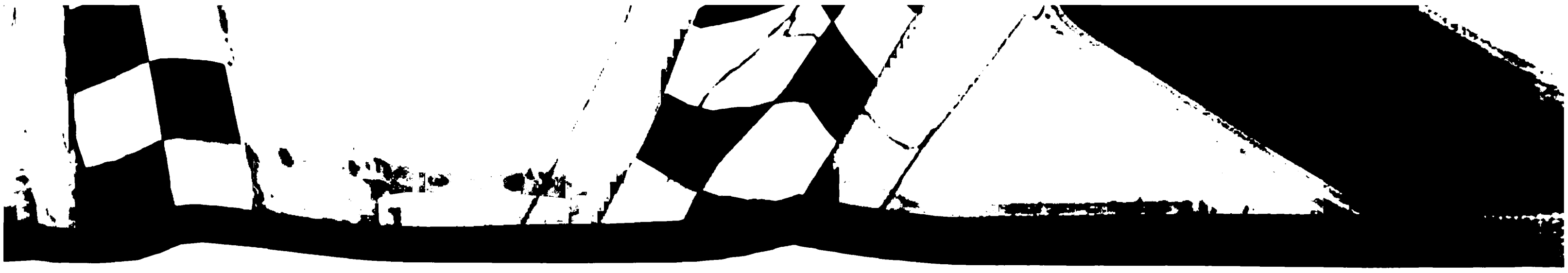
*Significant beyond the 0.01 level.

APPENDIX A (Cont.)

RESULTS OF SPEECH RECEPTION THRESHOLDS

TABLE XVI.--Summary of Analysis of Variance Comparing Differences of Testers (A), and Conditions (B) for Eight Normal Hearing Adults for Speech Reception Thresholds.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	2.000	3	0.666	0.333
B	5.000	3	1.666	0.833
AB	63.000	9	7.000	3.500
Z=C+AC+BC+ABC	32.000	16	2.000	
TOTAL	102.000	31		



APPENDIX B

RESULTS OF PURE-TONE AIR-CONDUCTION THRESHOLDS

TABLE XVII.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Pathologies (C), and Frequency (D) for Sixteen Hard-of-Hearing Adults by Pure-Tone Air-Conduction Audiometry.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	116.015	3	38.671	0.079
B	22.265	3	7.421	0.015
C	9422.005	1	9422.005	19.250*
AB	1655.338	9	183.926	0.375
AC	37.890	3	12.630	0.025
BC	25.390	3	8.463	0.017
ABC	4779.296	9	531.032	1.084
Z=E+AE+BE+CE+ABE+ ACE+BCE+ABCE	15662.499	32	489.453	
D	1254.166	2	627.083	5.112*
AD	56.250	6	9.375	0.076
BD	87.500	6	14.583	0.118
CD	5529.166	2	2764.583	22.539*
ABD	3052.083	18	169.560	1.382
ACD	37.500	6	6.250	0.050
BCD	25.000	6	4.166	0.033
ABCD	1775.000	18	98.611	0.803
Z=DE+ADE+BDE+CDE+ ABDE+ACDE+BCDE+ ABCDE	7849.999	64	122.656	
TOTAL	51387.369	191		

*Significant beyond the 0.01 level.

APPENDIX B (Cont.)

RESULTS OF PURE-TONE BONE-CONDUCTION THRESHOLDS

TABLE XVIII.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Pathologies (C), and Frequency (D) for Sixteen Hard-of-Hearing Adults by Pure-Tone Bone-Conduction Audiometry.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	15.104	3	5.034	0.019
B	52.604	3	17.534	0.066
C	49088.020	1	49088.020	186.631*
AB	1040.104	9	115.567	0.439
AC	77.604	3	25.868	0.098
BC	42.187	3	14.062	0.053
ABC	2004.687	9	222.743	0.846
Z=E+AE+BE+CE+ABE+ ACE+BCE+ABCE	8416.666	32	263.020	
D	460.156	2	230.078	5.106*
AD	16.927	6	2.821	0.062
BD	57.552	6	9.592	0.212
CD	4231.510	2	2115.755	46.962*
ABD	2865.364	18	159.186	3.533*
ACD	93.489	6	15.581	0.345
BCD	25.781	6	4.296	0.095
ABCD	1799.218	18	99.956	2.218
Z=DE+ADE+BDE+CDE+ ABDE+ACDE+BCDE+ ABCDE	2833.339	64	45.052	
TOTAL	73170.312	191		

*Significant beyond the 0.01 level.

APPENDIX (Cont.)

RESULTS OF SPEECH RECEPTION THRESHOLDS

TABLE XIX.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), and Pathologies (C) for Sixteen Hard-of-Hearing Adults for Speech Reception Thresholds.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	3.187	3	1.062	0.002
B	6.687	3	2.229	0.006
C	976.562	1	976.562	2.702
AB	2902.562	9	311.395	0.861
AC	12.187	3	4.062	0.011
BC	18.687	3	6.229	0.017
ABC	1355.562	9	150.618	0.416
Z=D+AD+BD+CD+ABD+ ACD+BCD+ABCD	11562.000	32	361.312	
TOTAL	16737.437	63		

APPENDIX B (Cont.)

RESULTS OF SPEECH DISCRIMINATION SCORES

TABLE XX.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), and Pathologies (C) for Sixteen Hard-of-Hearing Adults for Speech Discrimination Scores.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	373.500	3	124.500	0.675
B	97.500	3	32.500	0.176
C	5402.250	1	5402.250	29.320*
AB	3498.000	9	388.666	2.109
AC	76.250	3	25.416	0.137
BC	140.250	3	46.750	0.253
ABC	2683.250	9	298.138	1.618
Z=D+AD+BD+CD+ABD+ ACD+BCD+ABCD	5896.000	32	184.250	
TOTAL	18167.000	63		

*Significant beyond the 0.01 level.

APPENDIX C

RESULTS OF PURE-TONE AIR-CONDUCTION THRESHOLDS

TABLE XXI.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Articulation (C) and Frequency (D) for Sixteen Normal Hearing Children by Pure-Tone Air-Conduction Audiometry.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	109.375	3	36.458	1.538
B	128.125	3	42.708	1.802
C	4.687	1	4.687	0.197
AB	370.833	9	41.203	1.738
AC	25.520	3	8.506	0.358
BC	92.187	3	30.729	1.296
ABC	419.270	9	46.585	1.965
Z=E+AE+BE+CE+ABE+ ACE+BCE+ABCE	758.333	32	23.697	
D	2259.375	2	1129.687	58.228*
AD	18.750	6	3.125	0.161
BD	78.125	6	13.020	0.671
CD	153.125	2	76.562	3.946
ABD	197.916	18	10.995	0.566
ACD	47.916	6	7.986	0.411
BCD	53.125	6	8.854	0.456
ABCD	366.666	18	20.370	1.049
Z=DE+ADE+BDE+CDE+ ABDE+ACDE+BCDE+ ABCDE	1241.666	64	19.401	
TOTAL	6325.000	191		

*Significant beyond the 0.01 level.

APPENDIX C (Cont.)

RESULTS OF SPEECH RECEPTION THRESHOLDS

TABLE XXII.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), and Articulation (C) for Sixteen Normal Hearing Children for Speech Reception Thresholds.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	83.687	3	27.895	2.672
B	5.687	3	1.895	0.181
C	7.562	1	7.562	0.724
AB	214.062	9	23.784	2.278
AC	19.687	3	6.562	0.628
BC	11.687	3	3.895	0.373
ABC	130.062	9	14.451	1.384
Z=D+AD+BD+CD+ABD+ ACD+BCD+ABCD	334.000	32	10.437	
TOTAL	806.437	63		

APPENDIX C (Cont.)

RESULTS OF SPEECH DISCRIMINATION SCORES

TABLE XXIII.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), and Articulation (C) for Sixteen Normal Hearing Children for Speech Discrimination Scores.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	143.687	3	47.895	0.814
B	175.187	3	58.395	0.992
C	473.062	1	473.062	8.043*
AB	457.062	9	50.784	0.863
AC	109.687	3	36.562	0.621
BC	8.187	3	2.729	0.046
ABC	260.062	9	28.895	0.491
Z=D+AD+BD+CD+ABD+ ACD+BCD+ABCD	1882.000	32	58.812	
TOTAL	3508.937	63		

*Significant beyond the 0.01 level.

APPENDIX D

RESULTS OF PURE-TONE AIR-CONDUCTION THRESHOLDS

TABLE XXIV.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Frequency (C), and Pathologies (D) for Sixteen Hard-of-Hearing Adults by Pure-Tone Air-Conduction Audiometry.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	49.609	3	16.536	0.012
B	24.609	3	8.203	0.006
D	828.828	1	828.828	0.638
AB	1262.578	9	140.286	0.108
AD	212.734	3	70.911	0.054
BD	88.984	3	29.661	0.022
ABD	8993.203	9	999.244	0.769
Z=E+AE+BE+DE+ABE+ ADE+BDE+ABDE	41547.500	32	1298.359	
C	17652.656	4	4413.1640	37.12685*
AC	155.468	12	12.955	0.108
BC	155.468	12	12.955	0.108
CD	32966.093	4	8241.523	69.333*
ABC	2258.906	36	62.747	0.527
ACD	197.031	12	16.419	0.138
BCD	239.531	12	19.960	0.167
ABCD	2179.843	36	60.551	0.509
Z=CE+ACE+BCE+CDE+ ABCE+ACDE+BCDE+ ABCDE	15215.000	128	118.867	
TOTAL	124028.046	319		

*Significant beyond the 0.01 level.

APPENDIX D (Cont.)

RESULTS OF PURE-TONE BONE-CONDUCTION THRESHOLDS

TABLE XXV.--Summary of Analysis of Variance Comparing Differences of Testers (A), Conditions (B), Frequency (C), and Pathologies (D) for Sixteen Hard-of-Hearing Adults by Pure-Tone Bone-Conduction Audiometry.

Source of Variance	Sum of Squares	df	Mean Square	F-Statistic
A	43.437	3	14.479	0.046
B	190.312	3	63.437	0.205
D	30225.312	1	30225.312	97.846*
AB	813.437	9	90.381	0.292
AD	8.437	3	2.812	0.009
BD	91.562	3	30.520	0.098
ABD	2214.687	9	246.076	0.796
Z=E+AE+BE+DE+ABE+ ADE+BDE+ABDE	9885.000	32	308.906	
C	25739.375	4	6434.8437	67.707*
AC	167.500	12	13.958	0.146
BC	170.625	12	14.218	0.149
CD	15120.000	4	3780.000	39.773*
ABC	1422.500	36	39.513	0.415
ACD	186.875	12	15.572	0.163
BCD	191.250	12	15.937	0.167
ABCD	1786.875	36	49.635	0.522
Z=CE+ACE+BCE+CDE+ ABCE+ACDE+BCDE+ ABCDE	12165.000	128	95.039	
TOTAL	100422.187	319		

*Significant beyond the 0.01 level.

APPENDIX E

PURE-TONE AIR-CONDUCTION THRESHOLDS* OBTAINED IN EXPERIMENT I

	Frequency in Hertz					
	250	500	1000	2000	4000	8000
Subjects			ORIGINAL			
1	5	5	0	0	0	20
2	5	0	-10	-5	5	0
3	0	5	0	0	5	0
4	10	15	-5	-5	5	20
5	5	5	-5	-10	5	15
6	10	10	-5	0	5	10
7	0	0	-5	0	-5	5
8	-5	0	0	0	0	-5
			ACTUAL			
1	0	5	0	-5	10	0
2	5	0	-10	-10	5	5
3	0	0	-5	0	5	-10
4	10	10	-5	-10	5	15
5	5	0	-10	-15	5	10
6	5	5	0	5	5	0
7	0	0	-5	-5	-5	5
8	-5	-5	-5	0	0	-5
			BETTER			
1	5	0	-5	-5	10	5
2	5	0	-15	-15	0	5
3	0	0	0	0	5	-5

APPENDIX E (Cont.)

	Frequency in Hertz					
	250	500	1000	2000	4000	8000
Subjects			BETTER (Cont.)			
4	10	10	-5	-10	10	15
5	0	0	-10	-10	0	5
6	10	10	0	5	0	10
7	0	0	-5	0	-5	5
8	0	-5	-5	-5	0	-5
			POORER			
1	0	0	5	0	10	10
2	5	0	-10	-5	0	0
3	0	0	-5	0	10	-10
4	10	10	-5	-5	5	15
5	0	0	-5	-15	5	20
6	10	15	0	5	5	5
7	5	0	-5	0	0	5
8	0	0	-5	-5	0	-10
			NO INFORMATION			
1	0	0	-5	-10	5	0
2	0	0	-10	-10	5	0
3	0	0	0	0	10	-5
4	5	10	0	-10	10	20
5	0	5	-10	-10	5	15
6	10	5	-5	0	0	10
7	-5	0	-10	-5	-5	5
8	-5	-5	-5	-5	0	-5

*In dB re ISO 1964

APPENDIX E (Cont.)

SPEECH RECEPTION THRESHOLDS*
OBTAINED IN EXPERIMENT I

	Original	Actual	Better	Poorer	No Information
Subjects					
1	-4	-4	-6	-4	-6
2	-4	-6	-8	-4	-4
3	-4	-2	-2	-4	-4
4	0	-2	0	0	-2
5	-4	-6	-6	-4	-6
6	0	-2	-4	-4	-4
7	-4	-2	-2	-4	-4
8	-2	-2	-4	2	-4

*In dB re Audiometric zero.

APPENDIX F

PURE-TONE AIR AND BONE-CONDUCTION THRESHOLDS* OBTAINED IN EXPERIMENT II

Frequency in Hertz	500		1000		2000	
	Air	Bone	Air	Bone	Air	Bone
Sensorineural Subjects			ORIGINAL			
1	45	55	60	50	60	55
2	20	30	20	20	35	35
3	50	45	40	40	45	45
4	55	50	45	60	60	60
5	50	40	45	40	50	50
6	10	20	40	40	80	65
7	50	45	60	55	65	65
8	10	20	50	50	60	60
			ACTUAL			
1	50	55	60	65	70	65
2	20	25	25	20	35	35
3	45	35	50	45	45	40
4	55	55	45	50	55	55
5	50	40	50	40	55	45
6	5	20	40	50	80	65
7	50	45	50	55	55	65
8	10	20	45	50	60	60
			BETTER			
1	50	50	60	50	60	60
2	25	35	20	20	25	25

APPENDIX F (Cont.)

Frequency in Hertz	500		1000		2000	
	Air	Bone	Air	Bone	Air	Bone
Sensorineural Subjects			BETTER (Cont.)			
3	45	40	50	45	45	40
4	55	50	45	55	50	50
5	50	40	50	45	50	45
6	5	20	40	45	65	65
7	55	40	60	55	65	65
8	5	20	50	50	60	60
			POORER			
1	45	55	60	50	60	55
2	25	35	20	20	25	20
3	45	40	45	40	45	40
4	60	55	50	60	55	60
5	50	40	50	40	55	50
6	5	15	40	45	75	65
7	50	35	60	55	65	60
8	10	20	50	50	60	60
			NO INFORMATION			
1	45	55	60	50	60	55
2	25	25	20	20	25	20
3	45	35	45	45	45	40
4	55	55	50	60	60	60
5	45	40	50	45	55	50
6	5	20	40	50	80	65

APPENDIX F (Cont.)

Frequency in Hertz	500 Air	500 Bone	1000 Air	1000 Bone	2000 Air	2000 Bone
Sensorineural Subjects			NO INFORMATION (Cont.)			
7	50	45	55	55	65	65
8	10	20	50	50	60	60
Conductive Subjects			ORIGINAL			
1	50	20	55	20	50	20
2	50	15	30	15	35	10
3	35	10	30	5	30	0
4	40	25	30	10	30	0
5	15	20	10	0	15	0
6	35	5	35	5	30	0
7	45	15	35	15	30	20
8	25	15	25	15	25	15
			ACTUAL			
1	50	30	65	20	50	20
2	40	15	35	10	30	5
3	40	15	20	15	25	10
4	35	25	30	10	30	0
5	15	15	10	0	15	5
6	35	10	30	5	35	0
7	45	15	40	15	30	15
8	20	10	25	15	20	10

APPENDIX F (Cont.)

Frequency in Hertz	500		1000		2000	
	Air	Bone	Air	Bone	Air	Bone
Conductive Subjects			BETTER			
1	50	15	55	20	40	25
2	40	20	30	15	25	10
3	45	20	40	15	30	0
4	35	20	25	15	25	0
5	10	15	10	5	10	0
6	35	10	35	5	35	0
7	50	15	35	15	30	15
8	20	10	25	15	25	15
			POORER			
1	45	30	60	25	45	20
2	40	15	30	15	25	15
3	35	10	20	10	30	0
4	35	25	30	10	25	0
5	20	20	10	5	10	0
6	35	10	35	5	35	5
7	40	15	35	15	30	20
8	25	15	25	15	20	10
			NO INFORMATION			
1	50	25	65	35	45	25
2	50	10	35	15	35	10
3	40	20	30	15	30	15
4	30	20	25	10	30	0

APPENDIX F (Cont.)

Frequency in Hertz	500		1000		2000	
	Air	Bone	Air	Bone	Air	Bone
Conductive Subjects		NO INFORMATION (Cont.)				
5	10	20	10	5	10	0
6	40	10	30	10	30	5
7	50	20	35	20	30	15
8	20	10	25	15	25	15

*In dB re ISO 1964

APPENDIX F (Cont.)

SPEECH RECEPTION THRESHOLDS* AND
SPEECH DISCRIMINATION SCORES#
OBTAINED IN EXPERIMENT II

Condition	Original		Actual		Better		Poorer		No Info.	
	SRT	PB	SRT	PB	SRT	PB	SRT	PB	SRT	PB
Sensorineural Subjects										
1	56	86	56	84	56	88	58	88	50	90
2	20	82	20	80	18	96	18	88	20	94
3	48	88	40	84	42	96	42	96	44	96
4	58	76	56	80	56	80	54	88	56	86
5	46	80	46	92	48	78	46	84	46	84
6	8	50	6	50	8	34	2	28	8	66
7	56	68	46	76	52	74	52	68	48	76
8	12	50	8	44	12	48	10	44	12	46
Conductive Subjects										
1	44	94	46	84	50	92	48	100	44	94
2	30	96	30	92	32	98	32	94	32	92
3	28	96	32	94	28	90	28	98	24	92
4	26	96	24	98	26	98	24	98	26	100
5	6	92	6	90	4	88	6	90	4	90
6	30	96	34	98	32	96	34	94	30	100
7	32	98	32	96	28	100	32	98	32	96
8	22	90	22	90	20	86	22	90	20	86

*In dB re Audiometric zero.
#In percent correct.

APPENDIX G

PURE-TONE AIR-CONDUCTION THRESHOLDS* OBTAINED IN EXPERIMENT III

	Frequency in Hertz		
	500	1000	2000
Normal Artic. Subjects		ORIGINAL	
1	15	0	5
2	15	15	0
3	20	10	5
4	20	5	10
5	15	10	15
6	20	5	10
7	20	10	0
8	5	0	5
		ACTUAL	
1	10	0	0
2	10	5	0
3	20	0	0
4	5	0	0
5	5	0	10
6	20	0	0
7	10	5	0
8	5	0	5
		BETTER	
1	0	0	5
2	15	5	0
3	15	0	0

APPENDIX G (Cont.)

Normal Artic. Subjects	Frequency in Hertz 500	1000	2000
		BETTER (Cont.)	
4	15	0	5
5	15	5	15
6	20	10	0
7	10	5	0
8	5	0	5
		POORER	
1	10	0	0
2	20	5	5
3	20	0	0
4	15	0	0
5	15	10	15
6	15	5	5
7	15	5	0
8	0	0	0
		NO INFORMATION	
1	10	0	0
2	10	5	0
3	15	0	0
4	10	0	5
5	10	0	10
6	15	10	0

APPENDIX G (Cont.)

	Frequency in Hertz		
	500	1000	2000
Normal Artic. Subjects		NO INFORMATION (Cont.)	
7	10	10	0
8	10	5	0
Defective Artic. Subjects		ORIGINAL	
1	10	0	5
2	5	5	5
3	10	5	0
4	15	10	10
5	15	5	5
6	10	10	10
7	15	0	0
8	5	5	0
		ACTUAL	
1	5	0	5
2	10	5	10
3	15	10	5
4	10	0	5
5	10	0	0
6	15	10	0
7	10	0	0
8	0	0	0

APPENDIX G (Cont.)

	Frequency in Hertz		
	500	1000	2000
Defective Artic. Subjects		BETTER	
1	5	0	0
2	10	5	5
3	10	10	5
4	10	10	10
5	0	0	0
6	0	0	0
7	0	0	0
8	5	0	0
		POORER	
1	15	10	15
2	10	0	0
3	15	10	10
4	5	5	0
5	15	5	5
6	5	10	10
7	15	0	0
8	5	5	0
		NO INFORMATION	
1	15	0	5
2	10	0	0
3	10	10	0
4	10	10	5

APPENDIX G (Cont.)

	Frequency in Hertz		
	500	1000	2000
Defective Artic. Subjects			
5	15	0	0
6	5	10	10
7	15	0	0
8	10	5	0

*In dB re ISO 1964

APPENDIX G (Cont.)

SPEECH RECEPTION THRESHOLDS* AND
SPEECH DISCRIMINATION SCORES#
OBTAINED IN EXPERIMENT III

Condition	Original		Actual		Better		Poorer		No Info.	
	SRT	PB	SRT	PB	SRT	PB	SRT	PB	SRT	PB
Normal Artic. Subjects										
1	8	98	6	100	4	96	4	100	2	90
2	6	98	8	92	6	98	8	100	6	96
3	10	98	16	90	16	94	12	100	10	100
4	6	100	8	100	6	100	6	96	4	100
5	10	100	0	80	4	100	10	100	8	100
6	6	98	8	100	8	96	6	98	6	96
7	8	96	6	100	8	100	6	100	8	100
8	8	100	8	100	4	100	12	98	6	100
Defective Artic. Subjects										
1	6	96	4	80	6	86	4	96	12	76
2	4	96	8	96	4	100	4	100	10	100
3	6	88	12	60	12	80	16	90	8	96
4	6	92	10	94	6	92	6	92	6	94
5	6	90	4	92	10	100	10	78	8	88
6	4	100	6	96	8	98	6	100	8	100
7	2	94	6	98	0	94	4	96	4	90
8	2	96	2	94	2	98	2	96	0	96

*In dB re Audiometric zero.
#In percent correct.

APPENDIX H

PURE-TONE AIR-CONDUCTION THRESHOLDS* OBTAINED IN EXPERIMENT IV

	Frequency in Hertz				
	250	500	1000	2000	4000
Sensorineural Subjects			ORIGINAL		
1	25	15	5	45	70
2	25	30	30	30	40
3	55	55	35	50	75
4	15	15	15	30	55
5	15	30	30	35	75
6	20	20	20	30	65
7	25	30	30	30	80
8	5	5	0	35	50
			ACTUAL		
1	20	15	0	40	65
2	10	15	35	30	45
3	40	45	40	50	85
4	20	20	20	30	55
5	20	25	25	40	75
6	20	20	15	20	55
7	30	25	25	25	85
8	0	5	0	35	50
			BETTER		
1	15	15	0	45	65
2	10	20	35	25	40
3	45	45	40	45	75

APPENDIX H (Cont.)(Air-conduction thresholds)

	Frequency in Hertz				
	250	500	1000	2000	4000
Sensorineural Subjects			BETTER (Cont.)		
4	15	15	10	30	50
5	30	30	25	40	75
6	35	25	20	20	60
7	30	30	25	30	80
8	0	5	0	35	55
			POORER		
1	5	15	5	45	65
2	10	20	30	25	35
3	35	40	40	50	80
4	20	20	10	25	55
5	15	30	30	35	75
6	25	20	15	20	55
7	30	35	30	30	85
8	0	5	0	35	50
			NO INFORMATION		
1	25	10	0	45	65
2	20	20	30	25	35
3	40	40	40	60	85
4	20	15	10	20	55
5	20	30	30	40	75
6	20	20	20	20	65

APPENDIX H (Cont.) (Air-conduction thresholds)

	Frequency in Hertz				
	250	500	1000	2000	4000
Sensorineural Subjects		NO INFORMATION (Cont.)			
7	25	30	30	35	85
8	5	5	0	35	50
Conductive Subjects			ORIGINAL		
1	65	70	45	40	60
2	55	55	30	25	20
3	55	50	55	50	35
4	25	15	5	5	10
5	20	25	20	15	45
6	65	60	55	45	40
7	45	40	30	20	45
8	45	45	30	30	15
			ACTUAL		
1	60	70	50	40	65
2	45	45	30	30	20
3	50	50	60	45	30
4	15	5	0	-5	0
5	25	25	15	15	40
6	55	60	60	40	40
7	45	40	30	20	45
8	45	40	25	25	10

APPENDIX H (Cont.) (Air-conduction thresholds)

Conductive Subjects	Frequency in Hertz 250	500	1000	2000	4000
			BETTER		
1	55	65	50	45	60
2	50	55	35	25	15
3	45	60	65	40	45
4	20	10	5	0	5
5	15	20	15	20	45
6	60	60	55	30	35
7	45	40	35	20	40
8	45	40	25	25	10
			POORER		
1	65	65	45	35	60
2	55	55	30	25	20
3	45	50	45	45	40
4	20	10	5	0	15
5	20	30	15	15	45
6	65	60	55	30	40
7	35	40	35	20	40
8	45	40	20	25	15
			NO INFORMATION		
1	60	70	55	50	55
2	50	45	25	20	15
3	45	50	70	40	40
4	10	5	0	0	5

APPENDIX H (Cont.)(Air-Conduction Thresholds)

	Frequency in Hertz				
	250	500	1000	2000	4000
Conductive Subjects		NO INFORMATION (Cont.)			
5	15	20	5	10	45
6	50	50	55	25	35
7	45	45	30	20	55
8	40	35	25	25	10

*In dB re ISO 1964

APPENDIX H (Cont.)

PURE-TONE BONE-CONDUCTION THRESHOLDS*
OBTAINED IN EXPERIMENT IV

	Frequency in Hertz				
	250	500	1000	2000	4000
Sensorineural Subjects			ORIGINAL		
1	20	15	10	40	65
2	20	35	30	25	45
3	40	50	35	45	65
4	25	15	20	25	55
5	10	25	30	30	65
6	20	20	15	30	60
7	0	25	25	25	65
8	10	25	0	35	50
			ACTUAL		
1	15	20	10	40	65
2	20	35	30	20	45
3	35	45	35	45	65
4	20	20	15	25	55
5	15	30	25	40	60
6	10	15	15	25	65
7	5	25	20	25	65
8	10	20	0	35	45
			BETTER		
1	15	15	5	45	65
2	20	30	35	20	45
3	30	50	35	45	65

APPENDIX H(Cont.) (Bone-conduction thresholds)

Sensorineural Subjects	Frequency in Hertz 250	500	1000	2000	4000
	BETTER (Cont.)				
4	25	20	25	25	55
5	15	25	30	30	65
6	10	20	20	20	60
7	5	25	20	30	65
8	10	25	0	40	50
	POORER				
1	15	20	10	40	65
2	20	35	30	20	50
3	30	50	30	45	65
4	20	25	10	15	60
5	10	30	20	40	65
6	10	20	15	30	65
7	5	30	25	25	65
8	10	25	0	45	50
	NO INFORMATION				
1	15	20	10	40	65
2	20	35	25	25	45
3	30	50	30	45	65
4	20	20	15	20	55
5	15	35	30	40	65
6	20	20	15	30	60

APPENDIX H (Cont.) (Bone-conduction thresholds)

	Frequency in Hertz				
	250	500	1000	2000	4000
Sensorinaural Subjects			NO INFORMATION (Cont.)		
7	5	25	20	30	65
8	15	30	5	35	40
Conductive Subjects			ORIGINAL		
1	0	15	25	20	45
2	10	25	10	0	0
3	0	30	25	20	20
4	5	5	-5	-10	0
5	10	20	10	5	30
6	0	5	10	10	10
7	5	15	15	5	20
8	15	25	10	0	0
			ACTUAL		
1	0	15	20	15	40
2	5	25	15	0	0
3	0	30	25	20	20
4	5	5	0	-10	0
5	0	15	5	0	20
6	0	0	5	0	0
7	5	10	10	10	20
8	10	15	10	0	10

APPENDIX H (Cont.) (bone-conduction thresholds)

	Frequency in Hertz				
	250	500	1000	2000	4000
Conductive Subjects			BETTER		
1	0	20	20	15	40
2	15	30	15	0	0
3	5	25	30	25	15
4	0	5	-5	-10	0
5	5	20	15	0	25
6	5	5	20	5	10
7	10	20	15	5	20
8	15	20	10	0	10
			POORER		
1	0	20	40	30	40
2	0	30	15	0	0
3	0	25	25	25	25
4	5	0	-5	-10	0
5	15	20	10	0	20
6	5	10	20	10	20
7	5	20	15	0	20
8	10	25	10	0	15
			NO INFORMATION		
1	0	25	25	15	45
2	0	20	15	0	0
3	0	25	30	25	15
4	0	0	0	-10	5

APPENDIX H (Cont.) (bone-conduction thresholds)

	Frequency in Hertz				
	250	500	1000	2000	4000
Conductive Subjects			NO INFORMATION (Cont.)		
5	5	15	5	0	20
6	0	0	15	25	10
7	10	15	5	10	25
8	15	20	15	0	15

*In dB re ISO 1964

APPENDIX I

A Copy of the Letter Sent to Each Adult Subject Who Participated in This Study

We here at the Michigan State University Speech and Hearing Clinic are conducting a research project in the area of hearing disorders.

Our records indicate that you have previously been given a hearing evaluation and that the evaluation indicated that you have a hearing loss. The project that we are conducting involves patients who have hearing losses; and the ultimate goal of our research is to determine a better way of serving the hard-of-hearing population.

We would sincerely appreciate it if you would volunteer as a subject.

If you do volunteer for this research it will mean: (1) that you will have a free series of hearing tests, (2) that you will have to come to the Michigan State University Speech and Hearing Clinic for a two hour test session, and (3) it is expected that the knowledge gained from this research project will be of benefit to individuals with a hearing loss in the future in that we are developing new methods for measuring and evaluating hearing impairments.

In a few days you will be receiving a phone call from our clinic to determine if you will be willing to participate in this very important research study.

Sincerely,

William F. Rintelmann, Ph.D.
Associate Professor
Nicholas M. Hipskind
Clinical Assistant