Open access • Journal Article • DOI:10.1044/JSHR.1703.531

Performance of severely hearing-impaired children on a closed-response, auditory speech discrimination test. — Source link

Kenneth O. Jones, Gerald A. Studebaker

Institutions: Brigham Young University, University of Memphis

Published on: 01 Sep 1974 - Journal of Speech Language and Hearing Research (American Speech-Language-Hearing Association)

Topics: Discrimination learning and Auditory perception

71-17,043

JONES, Kenneth Oscar, 1943-
THE PERFORMANCE OF SEVERELY HEARING-IMPAIRED CHILDREN ON A CLOSED-RESPONSE AUDITORY-SPEECH-DISCRIMINATION TEST.

The University of Oklahoma, Ph.D., 1970
Health Sciences, hearing

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE PERFORMANCE OF SEVERELY HEARING-IMPAIRED CHILDREN ON A CLOSED-RESPONSE AUDITORY-SPEECH-DISCRIMINATION TEST

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

KENNETH OSCAR JONES

Oklahoma City, Oklahoma

1970

THE PERFORMANCE OF SEVERELY HEARING-IMPAIRED CHILDREN ON A CLOSED-RESPONSE AUDITORY-SPEECH-DISCRIMINATION TEST

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to the following:

To Dr. Gerald A. Studebaker, advisor for this study, for his enthusiastic encouragement and meticulous interest in the details of this study.

To the dissertation reading committee and final examination committee, comprised of Drs. Howard B. Ruhm, Eugene O. Mencke, William A. Cooper, Jr., and Paul S. Anderson.

To the Veterans Administration Hospital, Oklahoma City, for providing financial support under the VA Training Program.

To the Speech and Hearing Center, University of Oklahoma Medical Center, for furnishing instrumentation and testing facilities for this study.

To my parents for their encouragement and hopes which ultimately led to this degree.

A special word of appreciation is expressed to my wife, Kathy, whose encouragement and assistance made this study possible.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF TABLES

LIST OF TABLES--Continued

THE PERFORMANCE OF SEVERELY HEARING-IMPAIRED CHILDREN ON A CLOSED-RESPONSE AUDITORY-SPEECH-DISCRIMINATION TEST

CHAPTER I

INTRODUCTION

In 1953, Hirsh (25, p. 114) stated,

> By far the most important sounds to which we listen in civilized life are the sounds of speech . . . the signals that travel from person to person and so permit the complex communicative society that we have evolved.

Much research has been done over the years in an attempt to develop reliable and valid speech tests for evaluating speech-understanding ability. Substantial research effort has been directed toward the development of tests which indicate the extent of a patient's disability and his ability to benefit from amplification. However, there remains an unfilled need for a speech-discrimination test specifically designed for use with the severely hearing-impaired population. When the commonly used speech-discrimination tests are administered to members of this population, the results are usually so poor as to indicate nothing more than what was already known, i.e., that the person has a very severe hearing loss.

A major problem is that existing speech-discrimination tests are inappropriate for these persons. Most were standardized on normal-hearing individuals and frequently require reading ability and a certain auditory-receptive vocabulary. Most tests were designed for people who once had normal hearing. When these tests are administered to severely hearing-impaired individuals, particularly to those with congenital-hearing losses, scores are often at or near 0 percent. The low scores in themselves testify to the inappropriateness of the test used.

The speech tests commonly used today for clinical-hearing evaluations and for hearing-aid evaluations are based upon test paradigms that were originally designed for testing communication systems. These clinical tests have generally taken the form of the monosyllabic, "phonetically balanced" (PB), meaningful word test item tests with phonemic content which is ostensibly representative of the statistical occurrence of the phonetic or phonemic elements within the parent language. The PB-word lists (each list 50 words in length) provide an overall indication of a patient's discrimination ability, but attempts to evaluate speech discrimination analytically with the PB-word lists have not been successful (8, 50, 57). In addition, investigators (8, 29, 30, 38, 55, 59, 60, 73) have noted the limitations of using PB-word lists to differentiate reliably between performances of hearing aids.

The above-mentioned limitations of PB-speech tests suggest the need for new methods for the evaluation of the severely hearing-impaired. Researchers concerned about the need for better tests to measure speech-discrimination ability of both normal and hearing-impaired listeners have developed some new test paradigms. Among these are the Rhyme Test (14), the Closed-Response Set (27), the Minimal Contrast Test (17), and the University of Oklahoma Speech Test #6 (UOST #6) (51, 64). These tests have been used in an attempt to study speech-discrimination abilities more analytically; that is, to go beyond the mere counting of the total number of errors by studying the pattern of the errors produced by individuals and by groups.

A closed-response set is a multiple-alternative forced-choice speech test. It constrains a subject's responses to the alternatives in the set. By comparison, an open-response test is one in which a subject may choose any response he wishes limited only by the constraints of the total vocabulary available to him. A principal limitation of the open-response test is that the number of response alternatives available may vary widely from subject to subject and from test item to test item.

The closed-response set can be used with relative ease and success (14, 17, 27, 48, 51, 64, 71). It has been demonstrated that this test paradigm can be used successfully with severely hearing-impaired children (64, 71).

The closed-response set seems particularly well suited for use with persons with severe hearing losses for the following reasons:

1. It is nearly a vocabulary free test, and the vocabulary which is needed can be learned quickly, if necessary, by nearly all school-age children.

2. Learning effects are minimal and asymptote very early (52, 53, 56), simplifying the evaluation of the results of repeated testing.

3. Intra-word context and word-frequency effects are virtually eliminated as influences (52, 53, 56).

4. The items within the closed-response sets can be graded with at least moderate precision with respect to the difficulty of the phonemic differentiation called for (9).

5. The closed-response sets can be made very easy, e.g., a set may call for no more than the differentiation between one and two syllable words (9).

6. The response can be simplified to pointing to objects or pictures without altering the fundamental aspects of the approach (31).

7. The results lend themselves to error-matrix analyses (51).

8. The sets can be arranged to test an individual's ability to identify consonants or vowels in any normal position in a word and to differentiate within and among speech sounds categorized into groups, such as voiced, voiceless, nasal, glides, plosives, front, back, etc. or according to any other grouping scheme.

The University of Oklahoma Speech Test #6 (UOST #6), a modified closed-response set, was recently standardized on a group of normal-hearing young adults at the Speech and Hearing Center, University of Oklahoma Medical Center (51). An early version of this test was tried on a preliminary basis on 15 severely hearing-impaired children (64). On the basis of this experience, the UOST #6 appears to offer considerable potential as a means of evaluating the speech-discrimination ability of severely hearing-impaired individuals, both quantitatively, in terms of total scores, and qualitatively, in terms of the kinds of errors made by individuals or groups.

Such a procedure could find immediate application in the clinical evaluation of children with severe hearing losses. The decision regarding the type of educational setting best suited to a child's needs or whether a child should continue in a particular type of program may be influenced by the results of this test. Information obtained from this test could also be used in evaluating the progress of children in auditory training or be of help in analyzing the probable causes for lack of success in a particular training procedure.

The purpose of the present investigation was to contribute needed information about the performance of severely hearing-impaired children on a closed-response auditory-speech-discrimination test (UOST #6). The data were compared with other tests of hearing ability performed in this study, and also with such characteristics of the children as chronological age, I.Q. (performance section of the Wechsler Intelligence Scale for Children, W.I.S.C.), I.Q. (Leiter International Performance Scale), existing-audiometric data, achievement-test data (California Achievement Test), and teacher evaluations.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

There are several types of speech-hearing tests, such as threshold tests, intelligibility tests, and articulation tests. Speech-hearing tests have been used as diagnostic tools.

Currently used auditory-speech-discrimination tests are less reliable than is desirable and have not been used successfully to evaluate the auditory-speech-discrimination abilities of the severely hearing-impaired population. Moreover, they have not been used successfully to evaluate auditory-speech-discrimination ability in an analytical manner, that is, a meaningful evaluation of the types of errors made by a subject.

The present investigation is concerned with the auditory-speech-discrimination abilities of severely hearing-impaired individuals. Consequently, the following review is restricted to those tests which were principally designed to measure auditory-speech-discrimination ability.

The results of the California Achievement Test, the performance section of the Wechsler Intelligence Scale for Children (W.I.S.C.), and the Leiter International Performance Scale were used in the present investigation. A brief description and review of the development and standardization of these tests, therefore, is included. Literature dealing with the ability of hearing-impaired children to perform on these and similar tests is also briefly reviewed.

Auditory-Speech-Discrimination Tests

Open-Response Tests

An open-response test is one in which a subject may choose any response he wishes within the constraints of the total vocabulary available to him. The response alternatives available for any test item may vary widely from subject to subject depending on the subject's vocabulary. Wide variation may also occur from test item to test item.

The speech-test materials originally used for the clinical evaluation of a subject's auditory-speech-discrimination ability were developed by Egan (12) at the Psycho-Acoustical Laboratory of Harvard University. The PAL PB-50-test lists consist of 1,000 monosyllabic words distributed equally among 20 lists. These lists were developed according to the criterion of phonetic balance. Phonetic balance refers to the phonetic composition in each word list which is representative of the frequency of usage of phonemes in spoken American English. According to Egan, these word lists are also equal in average difficulty and in range of difficulty and are composed of words in common usage. The PB-50 lists were developed for the evaluation of military communications equipment.

In 1948, Davis (10) reported on the use of the PB-50's as a clinical-diagnostic test of hearing ability. These lists were recorded at the Technisonic Studios in St. Louis, Missouri, using the voice of Rush Hughes. Davis found that as the intensity of the speech signal was increased, the number of words understood increased at a rate of 2 percent per dB. Moreover, the recordings were found to differentiate between subjects with conductive and subjects with non-conductive hearing losses. The ability of these recordings to make this differentiation was due apparently to the manner in which the lists were spoken and recorded rather than to any inherent characteristic of the lists themselves (61).

In 1952, Hirsh, et al. (26) published the CID (Central Institute for the Deaf) W-22 word lists. Hirsh (24) reported that the words used in the PAL PB-50 lists were unfamiliar to many patients and that the Rush Hughes recordings of them were not adequately standardized. The W-22 test consists of 200 monosyllabic words distributed equally among four lists, with six randomizations of each list. The lists were recorded on magnetic tape at the Technisonic Studios using the voice of Ira Hirsh. The words used in the CID W-22 word lists were more familiar to the subjects than the words used in the PAL PB-50 lists. The W-22 lists were checked for familiarity with the Thorndike and Lorge lists (65). The use of familiar words had the effect of reducing the influence of learning on the results. Greater control was also exercised over phonetic balance within the test. The phonetic composition of the words was based on Dewey's (11) study of the phonetic composition of newsprint and the Bell Telephone Laboratories' study (15) of business telephone calls in New York City. The number of W-22 words understood increased at an average rate of 5 percent per dB as the intensity of the speech signal was increased. According to Hirsh and his associates (26), there were no consistent differences among the scores obtained with the four different lists.

Elpern (13) investigated the CID W-22 word lists to determine if the four lists were equal in difficulty. He found that a difference existed both in average level of difficulty and average range of difficulty. Elpern, however, indicated that these differences were not great enough to be of significance in the clinical setting.

Soon after the W-22 word lists were published, users of this test indicated that the W-22 word lists did not distinguish between subjects with conductive and subjects with non-conductive hearing losses (61). Silverman and Hirsh (61) compared the original PB-50 recordings spoken by Hughes, W-22 lists spoken by Hirsh, and W-22 lists spoken by

Reynolds (a female talker) to investigate the effect of talker differences among the different recordings. The results of this investigation indicated only small differences in intelligibility among the recordings. It was concluded that these differences were due to differences in the talker presentations, but the specific speaking characteristics which were responsible for the differences were not determined.

In 1949, Haskins (22) developed the PB-K-50 test specifically for use with children. This test consists of a vocabulary of 200 monosyllabic words distributed equally among four lists. The words used were taken from the speaking vocabularies of young children. Each list was phonetically balanced based on the vocabularies of young children.

In 1959, Lehiste and Peterson (38) developed the initial-consonant, vowel-nucleus, final-consonant (CNC) word lists. These lists consist of a vocabulary of 500 monosyllabic CNC words distributed equally among 10 lists. These lists were phonemically balanced. The term "phonemic balance" was used rather than "phonetic balance," according to Lehiste and Peterson, because subjects could be expected to perceive a sound as falling within a particular phoneme category even though considerable phonetic variation occurred across different pronunciations of the same phoneme. The phonemic balance was achieved by using each phoneme within a list in proportion to the frequency of occurrence for each phoneme as found in Thorndike and Lorge's (65) list of 1,000 most common words.

The Northwestern Auditory Test #4 was developed by Tillman, Carhart, and Wilber (68) in 1963. It consists of a vocabulary of 100 CNC monosyllabic words distributed equally between two lists, with six randomizations for each list. The 100 words were taken from the original Lehiste-Peterson CNC word lists (38). The NU Auditory Test #4 lists were, according to the authors, more precisely phonemically balanced than the original Lehiste-Peterson CNC word lists.

Studies at Northwestern University utilizing the NU Auditory Test #4 indicated that the test materials were too restricted in number. Learning effects and possibly other variables produced differential results when the tests were used repeatedly with individual subjects. Consequently, Tillman and Carhart (67) developed the NU Auditory Test #6 as an expansion of the NU Auditory Test #4. Test #6 consists of a vocabulary of 200 CNC monosyllabic words distributed equally among four lists. Phonemic balance was maintained throughout. Tillman and Carhart indicated that the four lists were essentially identical, interchangeable, and yielded approximately the same articulation-gain function (5.6 percent per dB).

In 1967, Sommerville (63) recorded and standardized the NU Auditory Test #6 at the Speech and Hearing Center, University of Oklahoma Medical Center. The results of her investigation indicated that the University of Oklahoma recordings of the NU Auditory Test #6 are equally reliable and interchangeable and produce essentially the same percent-per-dB slope (approximately 5 percent per dB) as obtained by Tillman and Carhart (67) using NU Auditory Test #6 as well as that obtained by others (21, 26) using the W-22 word lists.

The tests previously discussed were developed for the purpose of assessing the overall-auditory-speech-discrimination ability of a subject. Attempts to use these tests to evaluate auditory-speech-discrimination ability analytically (evaluating the types of errors made by a subject) have met with relatively little success. In 1959, Oyer and Doudna (50) employed the W-22 word lists as the speech signal in an attempt to analyze the response made by hearing-impaired subjects with varying hearing-loss etiologies. The subjects' incorrect responses were analyzed with respect to the phoneme confusions which occurred and their frequency of occurrence. The authors reported that although vowel confusions occurred less frequently than did consonant confusions, a greater proportion of confusions occurred among vowels than among consonants. They also reported that essentially the same phoneme confusions occurred in all the etiological categories studied and that substitutions occurred more frequently than either omissions or additions, although omissions and additions occurred more frequently in the final than in the initial positions of the test word. The error information was too limited with this test to distinguish specific error patterns.

In 1964, Schultz (57) utilized the W-22 word lists to determine a pattern of errors in an auditory-speech-discrimination task. Schultz stated that a subject's incorrect response should yield a predictable pattern because responses should exhibit phonemic lawfulness. Schultz was not successful in his attempt to analyze error patterns utilizing W-22 word lists because of the large number of constraints imposed on the subject's responses which were not acoustic in character (the need to respond in words, the subject's vocabulary, etc.). Schultz concluded that vowels which had previously been reported as contributing little to auditory-speech discrimination were of sufficient significance to warrant further exploration. He stated that tests ought to be developed which would provide separate testing and analysis of consonants and vowel speech sounds.

In 1967, Bricker (5), using an open-response test, investigated the auditory-speech-discrimination errors of 90 normal-hearing preschool children between the ages of 3 and 6 years. The test consists of 22 consonants coupled with each of two vowels to form consonant-vowel nonsense syllables. The consonants are combined with the vowel /a/ in one list and the vowel /ɛ/ in a second list. Bricker indicated that errors made were inversely related to the frequency of the sound in the vocabulary of the subjects as well as to the sound's frequency of occurrence in the English language. He also reported that the number of errors decreased as the age of the subject increased. More errors were associated with the place of articulation than with either the manner of articulation or the voiced-voiceless dimension. The author summarized by suggesting that an auditory-speech-discrimination test focused on the place of articulation alone would be the best type of test to bring the majority of responses under appropriate stimulus control.

In 1968, Boothroyd (4) developed a mathematical procedure for calculating a predicted phoneme-discrimination score on the basis of the acoustical properties of a specific phoneme, the frequency of occurrence of the phoneme, and certain contextual influences, such as the articulatory influences of the adjacent phonemes. He developed an open-response test composed of consonant-vowel-consonant words. The author tested two groups of children, one with normal hearing and the other with a "partial" hearing loss. The subjects' ages ranged from 5 to 15 years. The test consists of 150 "short, isophonemic words," utilizing a total of 30 phonemes distributed equally among 15 lists. The vocabulary was not restricted to that of any particular age group; consequently, many of the test words were unfamiliar to the subjects. The subjects' responses were written down by the tester. Boothroyd indicated that an auditory-speech-discrimination score based on a percentage of recognized phonemes was a more valid measure of the ability to make phonemic classifications than was a score based on the percentage of recognized words.

In 1969, Abbs and Minifie (1) presented an open-response auditory-speech-discrimination test to 17 normal-hearing children between the ages of 5 and 8 years to determine their ability to discriminate among fricatives. Carrell and Tiffany (7) have listed the following phonemes as fricatives: /f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/, and /h/. The fricatives used in the Abbs and Minifie study were /s/, /z/, /f/, /v/, /θ/, and /ð/. They were presented in consonant-vowel and vowel-consonant combinations. The three vowels used were /a/, /i/, and /aɪ/. Each subject received syllable pairs for each possible fricative-vowel comparison. The investigators analyzed the recorded speech signal to determine the duration of the consonants and vowels, the amplitude of the fricatives, and the center frequency and band width of the resonance curves of each of the fricatives. They evaluated the contribution of these acoustic cues to the discrimination scores obtained among the fricatives. The highest discrimination-error rates were observed in the pairs /f/ and /θ/ and /v/ and /ð/. Considerably lower discrimination-error rates were observed when voicing differences were present between two fricatives or if there were significant spectural differences. Certain spectural characteristics of the fricatives /s/ and /z/ set them apart from the other fricatives, thus making discrimination between /s/ and /z/ and any of the other fricatives considerably easier than discrimination among /f/, /v/, /θ/, and /ð/.

In 1969, Lawrence and Byers (35) investigated the ability of hearing-impaired subjects to identify voiceless fricatives. Five male adults who had high-frequency sensori-neural hearing losses were the subjects. Their speech-reception thresholds ranged from 6 to 22 dB, and their speech-discrimination scores ranged from 72 to 84 percent as measured on the W-22 word lists. The open-response test used for this study consists of 16 consonant-vowel syllables which were formed by combining the fricatives /ʃ/, /s/, /f/, and /θ/ and the four vowels /i/, /e/, /o/, and /u/. The investigators reported that the fricatives were identified correctly as follows: /ʃ/, 87 percent; /s/, 83 percent; /f/, 77 percent; and /θ/, 72 percent. The fricatives /s/ and /ʃ/ were confused one for the other, and /f/ and /θ/ were confused one for the other. No vowel confusions were reported. Fricatives, however, were more often confused in association with the front vowels /i/ and /e/ than with the back vowels /u/ and /o/. Lawrence and Byers questioned whether high-frequency energy was a valid and important contributor, particularly for individuals with high-frequency hearing impairments. They summarized their investigation by suggesting that low-frequency energy, intensity, and duration of the fricatives were important to hearing-impaired subjects as cues for voiceless-fricative identification.

Closed-Response Tests

The term "closed-response" was first used by House, et al. (27) in 1965 to describe a multiple-choice auditory-speech-discrimination test which they had devised. A subject selected his response from among a relatively small number of alternatives (from two to eight, but usually four), all of which were known to him. The alternatives in a set were presented to the subject, usually on a printed page, just before the presentation of the test item.

Closed-response or multiple-choice, forced-response tests were used in a number of early speech-hearing tests (43) but were not utilized for the analytical study of auditory-speech-discrimination ability. One of the earliest procedures of this type was a hearing-screening test used in the 1930's (61), in which a subject selected one of three pictures in response to a recorded voice heard over earphones. The set of three pictures sometimes included rhyming words. This procedure lost popularity as a screening device with the development of pure-tone screening tests.

The Victoria University Education of the Deaf New Standard Testing List (43) (year unknown) utilized a six-choice nonsense-word set for auditory-speech-discrimination testing. This test was used to obtain only quantitative total scores.

In 1945, MacFarland (43) utilized the closed-response set as a means of evaluating quantitative auditory-speech-discrimination scores. This test is simply a series of pictures of familiar monosyllabic words. The task of the subject was to point to the appropriate picture when a test word was presented. There was no attempt to evaluate the errors analytically.

Following World War II, other studies appeared on the development and use of multiple-choice or closed-response auditory-speech-discrimination tests. Haagen (18) and Black (3) reported multiple-choice speech-intelligibility tests which were primarily designed to test the articulation ability of communications systems and talkers. In these studies, the investigators did not analyze phoneme difficulty or phoneme-error patterns. The vocabulary for these tests consists of pairs of one and two syllable words. These tests were designed primarily to evaluate the talker rather than the listener. Therefore, the authors made no attempt to evaluate the performance of those with defective hearing.

In 1953, Quick (54) constructed a simplified open-response auditory-speech-discrimination test for use with hearing-impaired subjects. The test consists of two lists of 25, three-alternative sets and requires the subject to read the response words. Homophenous monosyllabic words are used in each set; i.e., pink, pig, big. Quick reported that the test was reliable and that it could measure the effects of auditory training.

In 1953, Watson (71) constructed his M/J lists of monosyllabic words taken from the vocabulary of 5-year-old normal-hearing Scottish children. Watson standardized the words for use with "partially" hearing-impaired English children. From this list, Watson constructed a six-item multiple-choice or closed-response auditory-speech-discrimination test. This test consists of 25 cards, six pictures on each. The names of the pictures are printed under each picture. His subjects were required to point to one picture in response to the speech signal.

In 1954, Harold (20) adapted Watson's (71) test so that it could be used with children who had more severe hearing impairments than Watson's "partially" hearing subjects. Harold's test consists of 10 words and pictures on a card. Only five words on a card were presented, and the remaining five words served as dummies or distractors. The task of his subjects was the same as in Watson's test.

In 1952, Murray (45), under the auspices of the Commonwealth Acoustic Laboratories of Australia, developed a test similar to Watson's (71). Murray's test consists of five cards with 25 words on each. Only 10 words from each group of 25 were presented. Dale (9) reported that Murray's test could be more quickly administered than

Watson's test. He admitted, however, that a certain amount of accuracy was lost because of this factor.

In 1962, Dale (9) reported that the tests of Watson (71), Harold (20), and Murray (45) were unsatisfactory for use with severely hearing-impaired children. Consequently, he developed two closed-response tests designed specifically for testing discrimination among vowel sounds. These vowel sounds were grouped in threes and in pairs. Dale believed that by using his test one was able to ascertain whether a severely hearing-impaired child had considerable or negligible auditory discrimination for speech sounds. He reported that in many instances tests which were too difficult for children were administered to them, and, therefore, the results obtained were of little or no value. Dale reported that out of 27 children who had hearing losses in excess of 100 dB (reference level not reported but presumably re. threshold British Standard, which is near the ANSI S3.6 - 1969 standard), 17 were able to discriminate accurately between three or more pairs of vowels on his test.

In 1953, Kendall (31) developed a closed-response auditory-speech-discrimination test for young hearing-impaired children using sets of toys or objects. He developed five lists, each containing 12 paired monosyllabic nouns. Each of the lists was equated for difficulty. Twelve toys or objects representing each of the 12 test words were placed in a box. The tester removed the first object and asked the subject, "What is the name of that?" When all the toys or objects had been named by a subject and placed in front of him, the subject was required to put them back into the box one by one as requested. As the test progressed or as objects on the table were reduced, the test became progressively easier. To counteract this, only the first 10 words in each list were used in the test. The last two words were included as distractors. Kendall reported that it was possible to ask the child to "Show me the ___" rather than "Put the ___ in the box." In this way, it was a test of equal difficulty throughout. If this latter method were used, the test was a 12-item multiple-choice or 12-item closed-response test. Kendall indicated that this method never seemed as successful for the younger children as did the activity of putting the toys or objects into the box. He reported that this test, although necessarily more inaccurate than tests for older children, proved to be very effective with younger children whose vocabularies were restricted.

In 1955, Miller and Nicely (44) investigated the use of the closed-response concept in an auditory-speech-discrimination test they devised. Their test is comprised of 16 consonants, each combined with the vowel /a/ to form a nonsense syllable. Each subject was familiarized with the test items and the response vocabulary prior to the data collection. The response range was sufficiently large so that the format placed little restriction upon the subject's response. Miller and Nicely investigated the phoneme-error-substitution patterns of normal-hearing subjects in a format where the constraints of language and vocabulary were removed. They also investigated the influences of noise and various filter settings upon the test signals. Matrix analysis of individual subject errors was not carried out in their study.

In 1958, Fairbanks (14) developed the Rhyme Test. In this "semi-closed" set, each test stimulus and all of the acceptable responses for it have common phoneme roots, hence, the term "Rhyme Test." The Rhyme Test consists of a vocabulary of 250 common monosyllabic words which contain 50 sets of five rhyming words including 18 consonants which have been reported to account for 90 percent of all consonant occurrences in the English language (15). Fairbanks developed an answer sheet which displays the rhyming portion of the test word with a blank before the common ending. The subject responds by adding the appropriate consonant to form the word he heard. The Rhyme Test uses essentially the same format as the test developed by Miller and Nicely (44). The Rhyme Test, however, uses meaningful words and limits the response by imposing the use of rhyming words, and hence the application of the term "semi-closed." The use of words increased the face validity according to Fairbanks. He indicated that chance probability varied from subject to subject according to each subject's vocabulary. He believed that the Rhyme Test might be suitable for determining auditory-speech-discrimination ability. Even though the test was not developed for detailed analysis of the substitution pattern, Fairbanks indicated that both the acoustical characteristics of the consonant itself and those of the vowel-consonant transition influenced identification of the test word.

In 1968, Kopra, Blossner, and Waldron (32) compared results obtained using the Fairbanks Rhyme Test (14) and the CID Auditory Test W-22 (26) on 15 normal-hearing and 15 hearing-impaired subjects. The ages of the subjects in the hearing-impaired group ranged from 15 to 65 years. Their speech-reception thresholds ranged from 8 to 62 dB. The ages of the normal-hearing subjects ranged from 19 to 37 years. Each subject's better ear, as defined by a speech-reception-threshold score, was used in the experimental-test session. Each subject was presented the Fairbanks Rhyme Test and the CID Auditory Test W-22 at -4, 0, 8, 16, 24, 32, and 40 dB sensation levels. The authors suggested that neither test appeared to be superior as a diagnostic indicator. Similar quantitative results were obtained using both tests, and they both differentiated between groups of normal-hearing subjects and subjects with sensori-neural hearing impairments. The performance-intensity functions for the two tests were similar for both groups of subjects. The authors summarized their investigation by suggesting that the Fairbanks Rhyme Test was the preferred test.

In 1965, a modification of Fairbanks' Rhyme Test (14) was presented by House, et al. (27). This new test is called the Modified Rhyme Test (MRT) and is composed of consonant-vowel-consonant (CVC) words. This test is by definition a true closed-response set because it limits the number of possible responses available to a subject for each test item presentation and presents the complete vocabulary of each closed-response set on a response sheet. The test is composed of 50 sets of six "rhyming" CVC words which commonly occur in the English language. In 25 of the sets, the words differ only in the initial-consonant position; in the remaining 25 sets, the words differ only in the final-consonant position. The task of a subject is to select the word presented to him from each closed-response set. The investigators reported the following data obtained from the MRT: (a) a performance-intensity function of 5.0 percent per dB, (b) more discrimination errors for voiced than unvoiced consonant sounds, and (c) more discrimination errors for the final-consonant test words. These data were obtained from normal-hearing subjects.

In 1967, Griffiths (17) presented a "diagnostic-articulation" test designed as the Rhyming Minimal-Contrast Test. He based his work on the House, et al. (27) study and on the Miller and Nicely (44) study. The Rhyming Minimal-Contrast Test consists primarily of CVC monosyllabic words. There is a total of 250 words, 150 of them taken from the House, et al. vocabulary. There are 50 sets of five words, each of which have a common consonant-vowel or vowel-consonant phoneme composition and differ only in the consonant to be tested. The initial-consonant represents the test item in 25 sets while the final consonant is the test item in the other 25 sets. The sets were designed in such a way as to offer minimal-contrast comparison. That is, Griffiths categorized the difference between phonemes according to place (position in the mouth most involved in the phoneme production), manner (the release of the vocal air stream), or voicing (voiced, voiceless, or nasal). Griffiths' results disagreed with earlier studies (14, 27, 44, 50) in that he found final-consonant items were not more difficult to recognize than were initial-consonant items. In an analytical evaluation of his data, he found a large number of confusions occurring between the /v/ and /ð/ phonemes and between the /f/ and /θ/ phonemes in both quiet and signal-to-noise listening conditions. Griffiths further indicated that the Minimal-Contrast Test was a valid and reliable diagnostic tool which offered an analytical evaluation of the errors produced by a speech-communication system.

In 1968, a further modification of the House, et al. (27) MRT was proposed by Kruel, et al. (33). This auditory-speech-discrimination test includes essentially the same vocabulary as the House, et al. test, but the test items are arranged in three original forms with six randomizations for each form. Included in the test set are six rhyming test items which vary either in the initial-consonant or the final-consonant position. The investigators produced recordings of the speech signals at different signal-to-noise levels (the noise and words were both recorded on tape) and used different talkers. Performance scores which fell below 90, 70, and 60 percent at three different signal-to-noise levels were considered abnormal. Analyses of the individual-subject-error patterns were not done. The authors, however, suggested that an error-pattern analysis would offer further information.

In 1959, Hutton, Curry, and Armstrong (28) presented a closed-response auditory-speech-discrimination test in which phonetic composition and word familiarity were considered. Information was obtained on a group of 30 subjects ranging from 8 to 71 years of age. The subjects' hearing varied from normal to moderate hearing losses, and their discrimination scores, as measured on the W-22 word lists, ranged from 66 to 100 percent. The closed-response test was presented free field at 30 dB above the level of the speech-reception threshold. This test was designed for subjects with mild to moderate hearing impairments and, therefore, was evaluated on subjects with hearing losses up to 60-dB hearing level (HL). The authors reported their test to be sensitive to different kinds of hearing losses and to yield reliable estimates of discrimination ability. The authors did not include error-matrix studies in their investigation.

In 1963, Myatt and Landes (46) presented a closed-response auditory-speech-discrimination test using pictures on cards for response purposes. Eight pictures, four to a card, were chosen to represent monosyllabic words judged to be within the vocabulary of the subjects tested. Some pictures were specifically grouped to present phonetically dissimilar words; i.e., fish, cup, blue, and book; whereas other groups were chosen to portray words of marked phonetic similarity; i.e., coat, cone, coke, and comb. A third type of grouping included elements of both similarity and dissimilarity; i.e., boy, wing, watch, ring. Each test word was presented with the carrier phrase "Point to the ___." After completion of the entire test, each subject who had made an error was shown the page on which the error had been made and was asked to name each picture on the page as the examiner pointed to it. The authors stressed that this was done to verify the subject's recognition of the pictures. The authors presented this test to four groups of 20 children each. There was a "normal" group, a "speech-defective" group, an "educable-mentally-retarded" group, and a "trainable-retarded" group. Children in all groups had normal hearing, and their ages ranged from 4 to 15 years. The investigators concluded that the phonetic variety of the picture groupings yielded valuable information from the number and types of errors a subject made. The authors further suggested that because of the non-verbal response required, their picture test would be applicable to those children who cannot give intelligible verbal responses due to immature speech or speech defects. They also concluded that intelligence was not a contributing factor to performance on their test; that is, most individuals with an I.Q. of at least 50 could perform the necessary task.

Ling and Druz (42) in 1967, and Ling (41) in 1968, attempted to develop speech-hearing tests which could be used to evaluate the effectiveness of auditory training and the effects of frequency transposition of speech signals. They devised several closed-response speech tests with which to accomplish this. The authors, however, did not evaluate these speech tests. Recently, other investigators have experimented with frequency-transposition and visual and tactile coding of speech. Many of them have used the closed-response-test paradigm, but in each instance, the speech test devised to evaluate the new equipment or method was not in itself evaluated.

In 1968, Owens and Schubert (48) reported on a closed-response test designed for use with hearing-impaired subjects. This test consists of CVC words arranged in four alternative sets. The variable phoneme is a consonant in either the initial or final position. The subjects who participated in this test were English-speaking adults (ages unreported) whose performance on the W-22 word lists varied from 20 to 70 percent. The test items were presented to one ear of each subject at a level 40 dB above his speech-reception threshold or, in case of a tolerance problem, at the highest level he would accept. Confusions between voiced and unvoiced consonants rarely occurred. Nasal and non-nasal sounds were rarely confused. A greater percentage of phoneme errors occurred in the final position than in the initial position. The "liquids" /r/ and /l/ were seldom confused with other phonemes. Owens and Schubert stated that the most efficient closed-response set would be one originating from words differing only in specific phoneme characteristics. They highly advocated the closed-response-set test as an auditory-speech-discrimination test for use with hearing-impaired subjects.

In 1968, Owens, Talbott, and Schubert (49) reported on a test designed for hearing-impaired subjects. This test specifically measures vowel discrimination among hearing-impaired subjects and includes all common vowels and diphthongs in the English language. The words are all monosyllabic and occur in the same closed-response-set format, differing only in the medial-vowel position. For the purposes of their study, differentiation was made between a vowel and a diphthong. One ear of each subject (ages unreported) was tested, and the presentation level was 40 dB above the subject's speech-reception threshold or, in case of a tolerance problem, at the maximum level he would accept. The subjects' scores on the W-22 test ranged from

20 to 70 percent with a mean score of 46 percent. On the closed-response-set test, the scores ranged from 83.7 to 98.6 percent with a mean of 93.6 percent. The authors suggested that vowel items would not in general make efficient auditory-speech-discrimination items in a closed-response test because the probability of error was too low.

In 1969, Berger (2) reported on a closed-response auditory-speech-discrimination test using five multiple-choice key words in sentences (KSU Speech Discrimination Test). The test consists of five phonetically similar multiple-choice key words for each sentence, any one of which can be a logical part of the sentence. The key word never appears as the initial word in the sentence. The task of a subject is to cross out on a printed answer form the key word which he believes the tester read. There are 8 equal forms with 13 sentences on each form. Each sentence becomes progressively more difficult from an auditory-discrimination standpoint, and the subject's score is based on the difficulty of the sentence to which he successfully responds. No effort was made to equate word difficulty. The key words are one or two syllables, but each group of five key words has the same number of syllables. The test sentences vary from four to nine words in length. Berger recommended that the examiner read the sentences in a conversational manner and peak the key words as near 0 on the VU meter as possible to reduce test-retest variability.

Berger evaluated the test on normal-hearing subjects between the ages of 18 and 32 years under three different types of difficult listening conditions chosen to yield an error score of about 20 percent. He compared scores obtained from the W-22 test and the KSU Speech Discrimination Test and found that the KSU test produced higher scores. The difference between the two tests, however, was never significant at the 5 percent level of confidence. Berger suggested that his test was useful with persons having a foreign accent and with young children. No attempt was made to analyze error patterns.

In 1968, Studebaker (64) developed an auditory-speech-discrimination-test paradigm based on the House, *et al.* (27) test. This test was designed to investigate vowel, initial-consonant, and final-consonant discrimination. The selection of the consonants for this test was based on the work of Heinz and Stevens (23). Studebaker classified phonemes on the basis of articulation categories, in keeping with the motor theory of speech perception (37), rather than on a purely acoustical basis. Studebaker's test compares four-word closed-response sets in which two articulation parameters are compared in pairs with given sets. He used a total of four parameters, including voicing, manner of production, place of articulation, and nasality. Studebaker's test results were found to be largely in agreement with previous studies (14, 17, 27, 44, 48). He noted that a more useful analysis would be allowed if the items within a set were varied only with respect to place of articulation (front, mid, or back area of the mouth). Consequently, he developed another initial-consonant test and a final-consonant test (with four-word forced-choice items) to evaluate this concept. Each test item within a test set utilizes a common word root and varies only in one consonant position; for example, pair, tear, care, and air. The manner or production in each set is constant. The test consonants differ only with respect to the place of articulation.

The results of Studebaker's study indicated that the test format was reasonably flexible and efficient so that a clinical and/or research auditory-speech-discrimination test would be possible. One major area of concern was the potential usefulness of the closed-response set with hearing-impaired subjects. Therefore, both normal-hearing and hearing-impaired subjects were run on the closed-response test. The results suggested the ability of the closed-response test, even in an early stage of development, to meaningfully rank severely hearing-impaired children with regard to their hearing ability. Studebaker also indicated that the closed-response test was better able to produce this ranking than pure-tone or PB auditory-speech-discrimination tests.

In 1969, Pederson (51) collected normative data on a test he constructed based on the Studebaker test paradigm (64). This study also refined and broadened the scope of the test format. The following is a brief description of Pederson's study and the University of Oklahoma Speech Test #6 (UOST #6) that he constructed as a part of that study.

The three subtests that make up UOST #6 are (1) an initial-consonant subtest, (2) a final-consonant subtest, and (3) a medial-vowel subtest. The names of the subtests refer to the position of the phoneme which varies in the items of a closed-response set. The following criteria were applied to the selection of the test stimuli:

1. Meaningful words were selected as test items whenever possible because they were easier for the subjects than nonsense materials. Meaningful words reduce learning effects by reducing the time needed for a subject to familiarize himself with the test items.
2. All test items are consonant-vowel-consonant (CVC) words because CVC's conveniently lend themselves to the varying of the sound in one phoneme position at a time within a closed-response set, while still maintaining a meaningful word.
3. All test items within a given closed-response set vary only in one phoneme position.
4. Test items were chosen for familiarity, when a choice was possible, in order to reduce to a

minimum the time necessary for familiarization by the subject.

Two additional requirements that were applied in the case of the consonant subtests were:

1. The variable phonemes in each closed-response set were selected in so far as possible to vary only in the place of articulation. It was necessary, however, to deviate from this criterion for one word within each consonant-subtest set. In this instance, the variable-phoneme position is left vacant.
2. Identical test phoneme sets are used in both initial-consonant and final-consonant subtests (51.)

The initial-consonant and final-consonant subtests have identical designs, but the medial-vowel subtest is different from the consonant subtests. The initial-consonant and final-consonant subtests each consist of five, four-item closed-response sets. In each subtest, each of the four items appear as a presentation stimulus four times (four items by five sets by four replications), producing an 80-item subtest. A four-item closed-response set was chosen because four items reduce chance performance to manageable levels. The four-item closed-response set also allows for inclusion of the /-/[1] alternative because most

[1]The symbol /-/ is used to indicate that the phoneme test position is left vacant in a closed-response set.

of the place-of-articulation phoneme groups occur in groups of threes.

The medial-vowel subtest consists of one, eight-item closed-response set. Each word is used as the test item eight times during the subtest, producing a 64-item medial-vowel subtest. A detailed description of these subtests, including a list of the phonemes and words used in each, is contained in Chapter III.

Pederson presented UOST #6 to 20 normal-hearing subjects (10 males and 10 females) at five different signal-to-noise ratios for each subject. These normative data were then analyzed quantitatively (by signal-to-noise ratios, subtest halves, subtest scramblings, and sex groups) and analytically in error matrices. Pederson reported the following quantitative findings. Intelligibility of the test items for all subtests increased successively as more favorable signal-to-noise ratios were used. The rate of increase observed for the final-consonant subtest was 2.3 percent per dB, for the initial-consonant subtest, 2.7 percent per dB, and for the medial-vowel subtest, 6.4 percent per dB. Each of the five scramblings of the UOST #6 subtest produced comparable results as main effects. First-half and second-half test scores of all subtests were not significantly different, indicating that comparable quantitative information might be derived from each half of the subtest. If only quantitative total scores were desired, the length of a clinically used UOST #6 probably could be cut in half.

Pederson reported a well defined heirarchy of error patterns for each subtest. The error patterns, particularly those for the consonant subtests, were in good agreement with earlier works. Pederson, however, indicated some important differences which will be discussed later in connection with presentation of the results of the present investigation. While Pederson reported good agreement between the UOST #6 consonant subtest and the results reported by previous investigators (14, 17, 27, 44, 48, 64), the results of the vowel subtest were not as consistent with earlier data.

The results of Pederson's study indicated that the UOST #6 was reasonably efficient so that a research auditory-speech-discrimination test might be used with a hearing-impaired population. The purpose of the present investigation was to contribute needed information about the performance of severely hearing-impaired children on this test.

Achievement and I.Q. Tests

Development, Description, and Standardization

In 1957, Tiegs and Clark (66) designed the California Achievement Test. This test consists of a series of comprehensive subtests designed for the threefold purpose of facilitating evaluation, educational measurement, and diagnosis. There are four alternative equivalent forms of the California Achievement Test, and each form is composed of three subtests: reading, arithmetic, and language. These three subtests are further divided into two parts. The reading test consists of reading vocabulary and reading comprehension. The arithmetic test consists of arithmetic reasoning and arithmetic fundamentals. The language test consists of mechanics of English and spelling. A subject's score illustrates, in terms of grade placement, his achievement in reading, arithmetic, and language. An examination of a subject's profile reveals areas in which the subject is strong, average, or weak relative to a chosen criterion.

In 1940, the Leiter International Performance Scale (39) was published as a standardized non-language test for the measurement of general intelligence, personality, and special abilities. The test has been revised several times since it was first designed by Leiter in 1927. It was developed for the testing of children between the ages of 2 and 18 years. The test consists of perceptual matching, analogy, memory, and other tasks. The instructions to the subject are simple and can be spoken or pantomimed. The mean I.Q. score on the Leiter test is 95, and the standard deviation is 20. The test has been reported to be especially well suited for language-handicapped children.

The Wechsler Intelligence Scale for Children (W.I.S.C.) was developed in 1949 by David Wechsler (72). The W.I.S.C. differs from other tests of intelligence for children by completely renunciating the concept of mental age as a basic measure of intelligence. An I.Q. score on the Wechsler test is obtained by comparing each subject's test performance, not with a composite age group, but exclusively with the scores obtained by subjects within a single age group. The W.I.S.C. consists of 12 tests which are divided into a verbal and performance portion. Subheadings under the verbal portion are information, comprehension, arithmetic, similarities, vocabulary, and digit span. Subheadings under the performance portion are picture completion, picture arrangement, block design, object assembly, coding, and mazes. In the standardization process of the W.I.S.C., all 12 tests were given to every subject. In the actual testing of a subject, however, only 10 tests are used, five verbal and five performance. The W.I.S.C. was standardized on a sample of 100 boys and 100 girls at each age level from 5 through 15 years (all white children). The mean I.Q. score on the W.I.S.C. is 100, and the standard deviation is 15.

Performance of Hearing-Impaired Subjects on Achievement and I.Q. Tests

Studies in the past few years have indicated that if non-language tests of intelligence are used, hearing-impaired children are not generally inferior to normal-hearing children in intelligence (16, 34, 36, 40, 62, 69). It is becoming clear that hearing-impaired children should not be considered inferior to normal-hearing children in intelligence.

In 1959, Larr and Cain (34) attempted to verify the usefulness of the W.I.S.C. (72) in measuring mental ability of hearing-impaired children. Two hundred and forty-eight students in a residential school for the deaf participated as subjects in this investigation. Their ages ranged from 8 to 19 years, and they had hearing losses exceeding a pure-tone average of 75 dB in the speech range. Approximately 55 percent of the subjects were congenitally hearing impaired, and 45 percent were adventitiously hearing impaired. The lowest W.I.S.C. score was 61, and the highest was 138. The mean I.Q. was found to be 97.8. Larr and Cain supported the notion that the W.I.S.C. was a useful tool for assessing the non-verbal abilities of hearing-impaired children. They also suggested that the hearing-impaired children who participated in their investigation demonstrated average mental ability as measured by the W.I.S.C.

A revolutionary approach to I.Q. testing with hearing-impaired children was used by Levine in 1963 (40). The test used was the Wechsler-Bellevue Intelligence Test which was revolutionary because it employs a performance as well as a verbal scale. The test and method of administration were easily adapted to the language and communication abilities of his hearing-impaired subjects, and the procedure permitted a much deeper probe of intelligence than was possible prior to the test's development. Levine's study revealed that although hearing-impaired subjects were quantitatively equal to normal-hearing subjects on the basis of I.Q. scores, there were distinctive and significant deficiences in patterns of thinking and reasoning. The author suggested that these deficiences appeared to resemble a picture of underdevelopment of mental potential that could be corrected through more effective educational procedures.

In 1962, Smith (62) tested both the verbal and performance portions of the W.I.S.C. with congenitally hearing-impaired children, children with "minor" hearing impairments, and children with "late onset" hearing impairments. His conclusions agreed with Levine's (40).

In other similar studies, Goetzinger and Rousey (16) in 1947, and Lavos (36) in 1962, using the performance scale of the W.I.S.C., confirmed the quantitative similarity in I.Q. between hearing-impaired and normal-hearing subjects.

In 1964, Vernon and Brown (69) evaluated two of the intelligence tests most commonly used with hearing-impaired children. They indicated that, at the time of their writing, the W.I.S.C. was the best test for hearing-impaired children between the ages of 9 and 16 years. It yielded a relatively valid I.Q. score and offered opportunities for qualitative interpretation of factors such as brain injury or emotional disturbance. It had good interest appeal and was relatively easy to administer and reasonable in cost.

Vernon and Brown (69) indicated in their evaluation of the Leiter International Performance Scale that it had good interest appeal and could be used to evaluate some hearing-impaired children who could not otherwise be tested. Although it is somewhat extensive and lacking in validation, they reported it is an excellent test for hearing-impaired children. One disadvantage reported was in the interpretation of the I.Q. scores because the mean of the test is 95 and the standard deviation is 20. This means that the absolute normal score on the Leiter test is 95 instead of 100 as on other intelligence tests. The authors further reported that a score of 60, for example, does not indicate mental deficiency but corresponds more to a score of 70 on a test such as the W.I.S.C. Therefore, care must be taken in interpreting the Leiter I.Q. scores.

In 1969, Hanson, Hancock, and Kopra (19) attempted to determine whether significant inter-relationships existed among visual-motor-perception, linguistic skills, academic achievement, and audiological status of hearing-impaired children. They obtained data from 199 subjects between the ages of 5 and 11 years who attended the Texas School for the Deaf. All subjects functioned at least at a dull normal level of intelligence. Visual-motor-perceptual ability was measured by the Bender Gestalt Test for Young Children. Linguistic skills were measured by the Illinois Test of Psycholinguistic Abilities (ITPA). Academic achievement was determined by the Gates Primary Reading Tests. Intelligence was measured by the Columbia Mental Maturity Scale. Pearson Product Moment Correlation Coefficients were computed for the scores of all subjects (122) from whom complete data had been obtained. The authors reported significant correlations between visual-motor-perception and intelligence, visual-motor-perception and linguistic ability, and visual-motor-perception and academic status. They summarized that visual-motor perceptual dysfunction was more frequently found in hearing-impaired children than in normal-hearing children. This lack, however, did not appear to be the result of brain damage. The authors recommended that a refined teacher evaluation scale which could assess language in the hearing-impaired child should be developed and correlated with the previous findings. They determined the best binaural two-frequency average of hearing levels selected from 500, 1,000, and 2,000 Hz. This measure of hearing ability did not correlate significantly with any of the other variables. The investigators reported that one would expect the hearing ability of the subjects to correlate significantly with language acquisition and other measures of linguistic ability. Their sample, however, consisted primarily of severely hearing-impaired children with insufficient residual hearing for the purpose of acquiring language. A sample should include a broader distribution of hearing ability in order to have meaningful correlations.

Data obtained from subjects in the present investigation on the California Achievement Test, Leiter International Performance Scale, and the performance section of the Wechsler Intelligence Scale for Children (W.I.S.C.) are reported in Chapter IV.

CHAPTER III

INSTRUMENTATION AND PROCEDURE

Introduction

Most of the auditory-speech-discrimination tests currently used clinically were devised to measure the speech-hearing status of individuals who once had normal hearing and whose hearing loss was less than severe. These tests have not proved appropriate for the speech-discrimination testing of severely hearing-impaired persons. Recently, investigators (14, 17, 27, 30, 51, 64) have developed the multiple-alternative forced-choice test in an effort to produce an auditory-speech-discrimination test which would reduce the contaminating influence of word frequency, word familiarity, and learning effects. These tests are generally referred to as closed-response-set tests. The University of Oklahoma Speech Test #6 (UOST #6) is such a test. UOST #6 has been standardized on normal-hearing young adults and has been shown to be reliable, easily administered, and easily scored (51). Further, it has shown some promise as an instrument to permit the analytical evaluation of individual and group speech-discrimination-ability patterns

(51). Watson (71) and Studebaker (64) demonstrated that the closed-response set can be used successfully with severely hearing-impaired children. For these reasons, this investigation was designed to study the performance of severely hearing-impaired children on UOST #6.

Test Description

W-22 Word List

In addition to UOST #6, three scramblings of list 2 of the CID W-22 test were used in this study (47). The words in this list are as follows.

1. ail (ale)
2. air (heir)
3. and
4. bin (been)
5. by (buy)
6. cap
7. cars
8. chest
9. die (dye)
10. does
11. dumb
12. ease
13. eat
14. else
15. flat
16. gave
17. ham
18. hit
19. hurt
20. ice
21. ill
22. jaw
23. key
24. knee
25. live (verb)
26. move
27. new (knew)
28. now
29. oak
30. odd
31. off
32. one (won)
33. own
34. pew
35. rooms
36. send
37. show
38. smart
39. star
40. tare (tear)
41. that
42. then
43. thin
44. too (two, to)
45. tree
46. way (weigh)
47. will
48. with
49. yore (your)
50. young

University of Oklahoma Speech Test #6

The following is a brief description of UOST #6. For additional details concerning test development, assembly, recording procedures, equipment, etc. used in the genesis of this closed-response test, the reader is referred to the dissertation of Pederson (51).

The words which make up the initial-consonant and final-consonant subtests are listed below. (The words on one line, e.g., items 1, 6, 11, and 16, make up a set. There are five sets in each consonant subtest.) Note that the test phonemes are the same in each of the consonant subtests.

Initial-Consonant Subtest

1.	pair	6.	tear	11.	care	16.	air
2.	bale	7.	dale	12.	gale	17.	ale
3.	fin	8.	sin	13.	thin	18.	in
4.	vee	9.	zee	14.	thee	19.	ee
5.	chop	10.	shop	15.	stop	20.	hop

Final-Consonant Subtest

1.	pop	6.	pot	11.	pock	16.	pa
2.	robe	7.	rode	12.	rogue	17.	row
3.	roof	8.	ruse	13.	Ruth	18.	rue
4.	live	9.	lies	14.	lithe	19.	lie
5.	leech	10.	leash	15.	least	20.	lee

The medial-vowel subtest consists of one, eight alternative set. The words used in the medial-vowel subtest set are as follows.

Medial-Vowel Subtest

1. beat	5. bait
2. bit	6. boat
3. bet	7. but
4. bat	8. boot

One talker with a General-American dialect recorded all the test items. The test items were spoken with as little exaggeration, prolongation, or individual word-production variation as possible. Each subtest was introduced by the talker with the following introduction: "This is University of Oklahoma Speech Test #6, Initial-Consonant Test (Final-Consonant Test, Medial-Vowel Test). Are you ready?" The test item presentations were incorporated into a common carrier phrase in order that the test items would more closely approximate conversational utterances and to alert the subject to the approaching presentation. The carrier phrase is "The word is ___." All test items were recorded and then rerecorded for equal intensity. A five-second interval elapsed between the onset of each carrier phrase.

Three test scramblings (51) for each of the three subtests were constructed. Each word in the consonant subtests is presented four times within each test for a test-list length of 80 items. Each word in the vowel subtest is presented eight times in each test for a test-list length of 64 items. An equal number of presentations of each item is maintained in each half of each subtest scrambling. Each of the three subtests also has three different response sheets (51). This was designed to counterbalance response biases based on the position of the printed response item on the response sheet.

Subjects

Data from the present experimental study were collected from 23 hearing-impaired children (12 males and 11 females), ranging in age from 9 years 3 months to 16 years 6 months inclusive. The mean age of the subjects was 12 years 9 months. A pilot study was conducted by this experimenter utilizing UOST #6 on 10 normal-hearing children (siblings of the hearing-impaired children used in the principal study) between the ages of 6 and 14 years. Data obtained from this preliminary study demonstrated that the performance of these children on UOST #6 is essentially the same as the performance of normal-hearing adults (51) and that normal-hearing children of approximately age 9 years and above can be expected to perform the task required by UOST #6. All subjects who participated in the pilot study and experimental study had parental permission (Appendix A).

Ten of the children (four males and six females) who participated attended the University of Oklahoma Medical Center School for the Deaf which is a laboratory oral day-school operated by the Department of Communications Disorders, University of Oklahoma Medical Center, Oklahoma City, Oklahoma.

Twelve children (seven males and five females) were pupils in the Oklahoma City public schools hearing-impaired classes. One child (male) was attending the Oklahoma State School for the Deaf, Sulphur, Oklahoma.

Only the better ear of each subject was utilized in the collection of experimental data. The better ear of each subject was determined on the following basis: (a) by asking the subject which was his better ear, (b) by asking the subject's teacher which was the subject's better ear, (c) by comparing the pure-tone averages for 500, 1,000, and 2,000 Hz for each ear, (d) by comparing the speech-reception thresholds for each ear, and (e) by observing in which ear the hearing aid was worn. In the case of indicator disagreement, that ear was chosen which the majority of the indicators favored. If a better ear could not be determined, either ear was tested.

Test Environment

All recordings, preliminary-audiometric tests, and experimental tests were carried out in a two-room sound-isolated test suite with a low ambient noise level located in the Speech and Hearing Center, University of Oklahoma Medical Center. The subjects were individually tested while seated in the inner room of the test suite. Visual communication between subjects and experimenter was carried out through an acoustically-damped window located in the wall separating the two rooms of the suite. A microphone talk-back system was employed to allow the subject to communicate with the investigator.

Instrumentation

Recordings

All recordings were made on 1½ mil acetate recording tape at a tape speed of 15 inches per second. A recording of a nursery rhyme that was used for determining the most-understandable-listening level (MULL) (Appendix B) was made with a microphone (Altec, Model 682-A) and a dual-channel tape recorder (Ampex, Model 354). The talker's voice was monitored on an external VU meter (Weston Electronic Instruments, Model 802) which was set up so that a 0-dB VU reading would correspond to a -10-dB deflection on the VU meter of the tape recorder. The nursery rhyme was recorded at approximately -10 dB on the tape recorder's recording VU meter with minimum intensity and stress variation.

The W-22 word list scramblings had been previously recorded on a dual-channel tape recorder (Ampex, Model 601) from a new set of Technisonic Studios record discs. The W-22 word list recordings for this study were made with the Model 601 output connected through a 600-ohm resistance pad to the input of a dual-channel tape recorder (Ampex, Model 354). A 1,000-Hz calibration tone was recorded at the beginning of each recording (nursery rhyme, W-22 word list scramblings, and UOST #6 subtest scramblings) at an intensity level approximately 10 dB above the level of the speech signals. The difference between the levels of the calibration tone and the speech signals allowed a mid-meter calibration reference for accurate calibration control while the speech signals were presented at a lower level, minimizing distortion due to system saturation. The calibration tone was monitored at 0 dB on the VU meter of the speech audiometer. All recordings were obtained with the complete electronics system, receiving power through a constant-voltage transformer (Sola Electronics, Model SB) to eliminate any source power fluctuations.

Preliminary Audiometric Tests

Air-conduction and bone-conduction pure-tone thresholds and speech-reception thresholds were established using a "diagnostic" pure-tone audiometer (Beltone, Model 15-C) and a speech audiometer (Grason-Stadler, Model 162). The air-conduction receivers (Telephonics, TDH-39, 10 ohm) were mounted in cushions (MX-41/AR). The bone-conduction transducer (Radioear, Model B-70) was held in place on the mastoid by a standard clinical headband.

Experimental Tests

A dual-channel tape recorder (Ampex, Model 354) served as the source of the pre-recorded speech signals (W-22, nursery rhyme, and UOST #6). The output of the tape recorder was connected to the tape input of the speech audiometer

(Grason-Stadler, Model 162), which provided intensity control of the speech signals by means of a 120-dB range 2-dB-per-step attenuator. The test stimuli were presented to each subject monorally using an air-conduction receiver (Telephonics, TDH-39, 10 ohm) mounted in a cushion (MX-41/AR). An inactive earphone and cushion covered the nontest ear.

The pilot study indicated that some children were confused about which test item number they should be responding to. In order to alleviate this problem, a counter indicator (Veeder-Root, Inc., Form #77604) was placed in front of each subject to indicate the number of the test item. The counter indicator received its power from a direct-current power supply (Heath Built, Model EUW-15) and was manually controlled by the experimenter from the control room of the two-room test suite.

Calibration

The acoustic output of the experimental test instrumentation was calibrated to current ANSI standards each day an experimental test session was performed. A 1,000-Hz calibration tone from an audio oscillator (Hewlett Packard, Model 200 ABR) measured at 1 volt RMS on a vacuum-tube volt meter (Hewlett Packard, Model 400 DR) was fed into the external input of the speech audiometer (Grason-Stadler, Model 162) and peaked at 0 dB on the VU meter of the speech audiometer. The attenuator dial of the speech audiometer was set at 70-dB HL while the output to the earphone was measured at 89.5-dB SPL on an audiometer-calibration unit (Allison Labs, Model 300). A frequency-response check was carried out on the tape-recording system prior to its use in this study. The frequency response of the tape-recording system operating at 15 inches per second was flat (± 2 dB) from 100 to 10,000 Hz. The attenuators of the Grason-Stadler speech audiometer were linear within ± .4 dB over the attenuator range used. This was periodically rechecked during collection of the data. The equipment was switched on at least one hour before a test session began.

Procedure

Data Collection

Twenty-three severely hearing-impaired children were tested on the modified closed-response set (UOST #6), W-22 word lists, and other tests of hearing ability.

Preliminary Audiometric Tests. Air-conduction and bone-conduction pure-tone thresholds for each ear were carefully determined for each subject individually using the method recommended by Carhart and Jerger (6).

A speech-reception threshold for each ear was determined for each subject by utilizing the Zenith Spondee Picture Test. This test is composed of 12 words which correspond to 12 pictures (a 12-alternative set). A 96-item randomized scrambling utilizing these 12 words was devised (Appendix C). This was accomplished by printing each of the 12 words on a separate slip of paper and putting them into a box. A word was randomly picked out of the box and replaced after each drawing. This procedure was performed 96 times. Each subject was asked to name the pictures for the experimenter at the beginning of the experimental test session to ensure the subject's familiarity with them. The spondee words were presented using monitored live voice. The talker used the phrase, "Point to the ___," to present the items of the randomized spondee word list and monitored the spondee word at a 0-dB deflection on the VU meter of the speech audiometer. The subject responded by pointing to the appropriate picture. The lowest level at which two correct responses for three presentations at one intensity level were obtained was the criterion used for speech-reception threshold. Threshold was crossed in 2-dB steps a total of four times, two crossings via the ascending method and two crossings via the descending method. Threshold was taken as the mean of the four crossings.

Experimental Tests. Each subject received the recorded speech stimulus for determining the most-understandable-listening level. The nursery rhyme "Bah Bah Black Sheep" was chosen as the speech stimulus for this task because the children were familiar with this particular nursery rhyme (Appendix B). Each subject was asked to listen to the recorded nursery rhyme as many times as necessary to determine the intensity level where he best understood the nursery rhyme, hence the term most-understandable-listening level. Each subject did this by personally adjusting the 2-dB-per-step attenuator on the speech audiometer. The most-understandable-listening level was used as a starting intensity level for the W-22 and the UOST #6 presentations.

After the most-understandable-listening level was determined for each subject, printed (Appendix D) and verbal instructions were given to him along with a demonstration of the test procedures. Time was allowed for questions after the instructions had been given. When the task was understood, the experiment began. The subject's task for UOST #6 was to listen to each test-item presentation, select from the appropriate closed-response set the item that he thought had been presented, underline that item on the response sheet for the initial-consonant and final-consonant subtests, or place a check mark in the square under the correct word on the response sheet for the medial-vowel subtest. Each subject was asked at the beginning of the test session to read aloud to the experimenter the W-22 words and the test items used in UOST #6 in order to ensure familiarity with them.

Each subject received a minimum of 12 test runs, e.g., three W-22 word list scramblings and three scramblings of each of the three UOST #6 closed-response subtests (initial-consonant, final-consonant, and medial-vowel subtests). A recorded word list was presented to the subject at his most-understandable-listening level. Each subject then received the other two scramblings of that test or subtest; one at

4 dB above the most-understandable-listening level and the other at 4 dB below the most-understandable-listening level. When necessary, each subject received additional scramblings until a "maximum" score was obtained. A maximum score was considered to have been obtained when a lesser score occurred 4 dB below and 4 dB above that score. The criterion for a maximum score was not met in the case of only two subjects because the output limits of the speech audiometer were reached first.

There were three response sheets available for each UOST #6 subtest (51). A different response sheet was used each time a subject was presented a subtest. The subjects responded by repeating the words for the W-22 word list presentations because traditionally it is done in this manner.

Existing-Subject Data. When available, the following data were obtained from the individual subject's school files: (a) chronological age, (b) I.Q. (performance section of the Wechsler Intelligence Scale for Children, W.I.S.C.), (c) I.Q. (Leiter International Performance Scale), (d) achievement data (California Achievement Test), and (e) existing-audiometric data.

Teacher Evaluation. A special teacher evaluation took place involving four teachers who routinely came in contact with the 10 subjects used in this study who attended the University of Oklahoma Medical Center School for the Deaf. Each teacher was asked to evaluate the children with whom he or she was acquainted in the following categories: (a) speech production, (b) auditory speech understanding, (c) lipreading, (d) overall understanding, and (e) receptive problems other than hearing. The method was one previously tried with apparent success by Studebaker (64). A teachers' meeting was called prior to the start of the school day, at which time teachers were instructed to rate each child on a scale from 0 to 100 in each of the five above-mentioned categories, keeping each child properly ranked with respect to the other children. The teachers were given the entire day to observe the children in their usual classroom routine. They could take notes but were asked not to discuss the matter with others until all their judgments had been made. They were asked to fill out a prepared form at the end of the day (Appendix E).

Data Evaluation

The statistical analyses necessarily required the extensive use of nonparametric statistics because much of the data was in the form of ranks. Principally, these included Spearman rank correlation coefficients (58). These were carried out between the W-22 test results and each of the three experimental closed-response subtests as well as between each of the four speech tests and each of the following factors: chronological age, I.Q. (performance section of the Wechsler Intelligence Scale for Children, W.I.S.C.), I.Q. (Leiter International Performance Scale), California

Achievement Test, and existing-audiometric data. These latter data were a matter of record and because some were incomplete, unsystematic, and uncontrolled, interpretations were restricted in some instances. These restrictions are discussed in the next chapter as the data are presented. Spearman rank correlation coefficients were also carried out between the speech-test results and the audiometric data acquired as a part of this study. Because these data were acquired under controlled conditions and were complete across subjects, interpretations were less restricted.

Spearman rank correlation coefficients were also carried out between the speech-test results and the teacher-evaluation results and between the acquired-audiometric data and the teacher-evaluation results. Spearman rank correlation coefficients were also carried out among the teacher-evaluated categories.

Chapter IV also contains substantial, purely descriptive sections. The results of the three UOST #6 subtests are presented in the form of phoneme-error matrices and discussed descriptively. Results of the phoneme-error-pattern analysis are compared to the results of the Pederson (51) study and others (1, 17, 27, 35, 44, 49, 50, 64).

CHAPTER IV

RESULTS AND DISCUSSION

Introduction

The present investigation was designed to contribute needed information about the performance of severely hearing-impaired children on a closed-response auditory-speech-discrimination test (University of Oklahoma Speech Test #6). The performance of 23 hearing-impaired children on University of Oklahoma Speech Test #6 (UOST #6) was compared to their performance on other tests of hearing ability also performed as part of this study. These data were also compared to other characteristics of the children: (a) chronological age, (b) I.Q. (performance section of the Wechsler Intelligence Scale for Children, W.I.S.C.), (c) I.Q. (Leiter International Performance Scale), (d) achievement-test data (California Achievement Test), (e) existing-audiometric data, and (f) teacher evaluations.

This chapter is divided into two main sections. The first section, Quantitative Analyses, includes a description and statistical analyses of individual-subject data obtained from school files, existing-audiometric data, experimental-audiometric data acquired as part of this study, and experimental-speech-test data. The second part of this chapter, Qualitative Analyses, presents data for each UOST #6 subtest in error-matrix form.

Quantitative Analyses

Existing-Subject Data

Table 1 presents individual-subject data obtained from the subjects' school files. As illustrated, the subjects ranged in age from 9 years 3 months to 16 years 6 months with a mean age of 12 years 9 months. Leiter I.Q. scores ranged from 64 to 120 with a mean score of 99.4 which is 4.4 points above the normative mean of 95. Performance I.Q. (W.I.S.C.) scores ranged from 79 to 132 with a mean of 102.4 which is 2.4 points above the normative mean of 100. This finding agrees with other studies (16, 34, 36, 40, 62, 69), suggesting that when non-verbal intelligence scales are used with hearing-impaired children, their performance is not different than that of normal-hearing children. The California Achievement Test scores (grade placement) ranged from 1.7 to 7.3 with a mean of 4.28 which is, on the average, about three and one-half years behind the grade placement of normal-hearing children with an average age of this group. This finding seems to be in reasonable agreement with investigations (16, 69) in the area of education of the hearing-impaired which have shown that hearing-impaired children range from .75 to 4 years

TABLE 1

INDIVIDUAL-EXISTING DATA FROM SCHOOL FILES

Subject	Chronological Age		I.Q. (Leiter)		I.Q. (W.I.S.C.)		California Achievement Test (Grade Placement)	
	Age	Rank	Score	Rank	Score	Rank	Score	Rank
1	15 yr 8 mo	4	91	12.0	103	7.0	6.3	3.5
2	14 yr 8 mo	5	64	17.0	79	13.5	3.3	16.0
3	13 yr 5 mo	8	112	5.0	80	12.0	5.4	8.0
4	12 yr 6 mo	14	107	7.0	120	3.0	5.5	6.5
5	12 yr 1 mo	15	100	9.5	79	13.5	5.5	6.5
6	11 yr 6 mo	17	90	13.0	114	5.0	5.2	9.0
7	10 yr 8 mo	19	120	1.0	97	9.0	3.7	11.5
8	10 yr 3 mo	21	88	14.0	89	11.0	2.1	19.5
9	10 yr 0 mo	22	118	2.0	132	1.0	1.7	22.0
10	9 yr 3 mo	23	100	9.5	127	2.0	3.5	14.0
11	16 yr 6 mo	1	. .	. .	90	10.0	6.7	2.0

TABLE 1--Continued

Subject	Chronological Age		I.Q. (Leiter)		I.Q. (W.I.S.C.)		California Achievement Test (Grade Placement)	
	Age	Rank	Score	Rank	Score	Rank	Score	Rank
12	13 yr 5 mo	8	. .	. .	103	7.0	3.0	18.0
13	11 yr 5 mo	18	115	3.5	. .	. .	2.1	19.5
14	13 yr 5 mo	8	. .	. .	. .	. .	6.3	3.5
15	13 yr 3 mo	10	108	6.0	103	7.0	6.1	5.0
16	13 yr 6 mo	6	. .	. .	. .	. .	3.7	11.5
17	11 yr 8 mo	16	115	3.5	. .	. .	3.3	16.0
18	12 yr 10 mo	12	83	15.0	. .	. .	4.8	10.0
19	16 yr 4 mo	2	104	8.0	118	4.0	7.3	1.0
20	13 yr 2 mo	11	98	11.0	. .	. .	3.3	16.0
21	10 yr 4 mo	20	. .	. .	. .	. .	1.8	21.0
22	16 yr 2 mo	3	. .	. .	. .	. .	. .	. .
23	12 yr 7 mo	13	77	16.0	. .	. .	3.6	13.0
Range	9 yr 3 mo-16 yr 6 mo		64-120		79-132		1.7-7.3	
Mean	12 yr 9 mo		99.4		102.4		4.28	

. . = Unavailable Data

behind the grade placement of normal-hearing children of the same average age depending upon the degree of hearing impairment.

Each subject was ranked relative to the other subjects on each of the above measurements. A numerical rank of one (1) represents the oldest subject, the highest (best) I.Q. score, and the highest achievement-test grade placement. When tied scores occurred, each score was assigned the average of the ranks which would have been assigned had no ties occurred. Period leaders (. .) represent unavailable data. In such instances, the subject was not included in the rankings.

Existing- and Experimental-Audiometric Data

Table 2 presents individual-subject-audiometric data from school files (existing) and the audiometric-test results acquired as part of this study (experimental). It also designates which ear of each subject was used in the experimental-data collection.

NR stands for "no response" and means that the subject did not respond to 500, 1K, or 2K Hz at 110 dB HL. Some of the pure-tone averages (PTA) are followed by a plus (+) sign which means that the PTA is greater than the indicated value because a response was not obtained at one or two of the three frequencies (500, 1K, or 2K Hz) at 110 dB HL. If no response was obtained at any one or two of the three frequencies, it was assigned a value of 110 dB HL and averaged with the remaining frequency or frequencies. A plus sign was

TABLE 2

INDIVIDUAL-EXISTING AND EXPERIMENTAL-AUDIOMETRIC DATA
(dB H.L. RE. ANSI - 1969 NORMS)

Subject	Experimental Ear	Existing-Audiometric Data						Experimental-Audiometric Data					
		PTA			SRT			PTA			SRT		
		R	L	Rank[a]	R	L	Rank[a]	R	L	Rank[a]	R	L	Rank[a]
1	R	88	100	11.0	. .	. .	. .	92	98	12.0	CNE	CNE	19.0
2	R	85	NR	10.0	80	. .	. .	83	110	10.0	80	CNE	19.0
3	R	73	70	5.0	70	. .	. .	70	70	5.0	68	74	6.5
4	R	83	73	9.0	. .	. .	. .	75	78	8.0	76	76	9.0
5	L	67	60	1.0	65	60	. .	60	57	1.0	58	55	1.0
6	L	108+	108+	20.5	CNE	CNE	. .	110+	108+	21.5	CNE	CNE	19.0
7	L	93	93	12.5	85	85	. .	90	93	13.0	96	96	13.0
8	L	80	77	7.0	. .	. .	. .	78	70	5.0	66	64	4.0
9	L	103	103	16.5	. .	. .	. .	97	87	11.0	CNE	84	11.0
10	L	NR	107+	18.5	. .	. .	. .	108+	103+	19.0	CNE	CNE	19.0
11	L	97	93	12.5	. .	. .	. .	93	93+	14.0	95	85	12.0
12	L	75	75	6.0	70	70	. .	82	73	7.0	78	68	6.5
13	R	82	89	8.0	. .	. .	. .	77	90	9.0	74	98	8.0
14	L	92	68	4.0	. .	60	. .	93	65	3.0	CNE	64	4.0

TABLE 2--Continued

Subject	Experi-mental Ear	Existing-Audiometric Data						Experimental-Audiometric Data					
		PTA			SRT			PTA			SRT		
		R	L	Rank[a]	R	L	Rank[a]	R	L	Rank[a]	R	L	Rank[a]
15	R	108+	NR	20.5	. .	. .	. .	95	99+	15.0	CNE	CNE	19.0
16	L	108+	103	16.5	CNE	CNE	. .	107+	97	16.0	CNE	102	14.0
17	L	103+	107+	18.5	CNE	CNE	. .	110+	108+	21.5	CNE	CNE	19.0
18	R	62	62	2.5	. .	. .	. .	70	70	5.0	64	68	4.0
19	R	62	70	2.5	65	70	. .	58	72	2.0	60	74	2.0
20	R	100+	105	15.0	. .	. .	. .	98+	105+	18.0	CNE	CNE	19.0
21	R	NR	NR	22.5	. .	. .	. .	105+	105+	20.0	CNE	CNE	19.0
22	R	97+	97+	14.0	. .	. .	. .	97+	97+	17.0	CNE	CNE	19.0
23	R	NR	NR	22.5	. .	. .	. .	NR	NR	23.0	CNE	CNE	19.0
Range (dB)		67-108+	60-108+	X	65-85	60-85	X	60-110+	57-110	X	58-96	55-102	X
Mean (dB)		88.3	87.4		72.5	69.2		88.5	88.5		74.1	77.5	

[a]Only those scores obtained for the experimental ear were included in the rankings.

. . = Unavailable Data

NR = No Response to 500, 1K, or 2K Hz at 110 dB H.L.

CNE = Could Not Establish

\+ = Pure-tone average is greater than numerical value since a response was not obtained to 500 or 1K or 2K Hz at 110 dB H.L. (ANSI - 1969 Norms).

attached to the resultant value. CNE stands for "could not establish," meaning that it was not possible to obtain a speech-reception-threshold (SRT) score.

Subjects' existing (from school files) pure-tone averages for the right ear ranged from 67 to 108+ dB HL with a mean score of 88.3 dB HL. Three subjects gave no response (NR) to 500, 1K, or 2K Hz. Subjects' existing pure-tone averages for the left ear ranged from 60 to 108+ dB HL with a mean of 87.4 dB HL. Four subjects gave no response (NR) to 500, 1K, or 2K Hz.

Subjects' experimental (acquired as part of this study) pure-tone averages for the right ear ranged from 60 to 110+ dB HL with a mean of 88.5 dB HL. Experimental pure-tone averages for the left ear ranged from 57 to 110 dB HL with a mean of 88.5 dB HL. Only one subject gave no response (NR) to 500, 1K, or 2K Hz for both the right and the left ears.

Subjects' existing (from school files) speech-reception thresholds for the right ear ranged from 65 to 85 dB HL with a mean of 72.5 dB HL. This range of scores includes only six subjects. Speech-reception thresholds had either never been attempted or could not be established on 17 subjects. Existing speech-reception thresholds for the left ear ranged from 60 to 85 dB HL with a mean of 69.2 dB HL. This range also includes only six subjects for the same reason as stated above.

Subjects' experimental (acquired as part of this study) speech-reception thresholds for the right ear ranged from 58 to 96 dB HL with a mean of 74.1 dB HL. This range includes only 11 subjects inasmuch as a speech-reception threshold could not be established (CNE) for 12 subjects. Experimental speech-reception thresholds for the left ear ranged from 55 to 102 dB HL with a mean of 77.5 dB HL. This range includes only 13 subjects for the same reason as stated above. It should be noted that approximately twice as many subjects successfully achieved speech-reception-threshold scores as a part of this study than in previous routine clinical evaluations. The reason for the relatively poor success rate in the clinical evaluation is uncertain. However, inadequate procedures or a failure to attempt the test because of an advance presumption of failure must be considered as possibilities.

The existing-audiometric data, which was acquired from individual-subject school files, and the experimental-audiometric data agree very well (correlations are positive .91 and above). Pure-tone averages and speech-reception thresholds for individual subjects agree within a few decibels, which was not unexpected. The existing- and experimental-audiometric data indicate that this group of subjects was, indeed, severely hearing impaired.

Each subject was ranked relative to the other subjects on each of the audiometric measurements for the experimental ear only. Speech-reception-threshold scores obtained from school files were not ranked because scores did not exist for 16 of the 23 subjects. Consequently, these scores were not used in the statistical analyses discussed later. A numerical rank of one (1) represents the best hearing for both the pure-tone averages and speech-reception thresholds. All of the pure-tone-threshold averages and speech-reception-threshold values are expressed relative to the American National Standards Institute (ANSI) S3.6 - 1969 norms. When tied scores occurred, each score was assigned the average of the ranks which would have been assigned had no ties occurred. Period leaders (. .) represent unavailable data, and that subject was not included in the rankings.

Experimental-Speech Tests

Tables 3 and 4 present individual subjects' most-understandable-listening levels (MULL) expressed in dB HL. Percent scores obtained on the experimental-speech tests at various intensity levels are also shown. The subjects' most-understandable-listening levels ranged from 62 to 106 dB HL with a mean of 90.9 dB HL. It can be seen from Table 3 that those subjects who chose an MULL of 90 dB HL or above did not perform above 0 percent on the W-22 test. Subject #20 who had an MULL of 80 dB HL also performed at 0 percent on the W-22 test. All other subjects with an MULL of less than 90 dB HL produced a supra-0 percent score on the W-22 test. This would indicate that an MULL of 90 dB HL is

TABLE 3

INDIVIDUAL MOST-UNDERSTANDABLE-LISTENING LEVELS (MULL) AND SCORES (IN PERCENT) FOR THE W-22 TEST AND THE MEDIAL-VOWEL SUBTEST AT VARIOUS INTENSITY LEVELS

Subject	Most-Under-standable-Listening Level (dB re. ANSI-1969)	W-22 Test dB Levels re. MULL						Medial-Vowel Subtest dB Levels re. MULL				
		-8	-4	0	+4	+8	+12	-8	-4	0	+4	+8
1	96		0	0	0				13	22	19	
2	90		0	0	0			31	53	38	23	
3	88		20	16	24	18		36	53	44	41	
4	80		8	26	28	16		38	52	47	39	
5	62		10	20	26	34	24		61	72	92	89
6	106		0	0	0				19	22	25[a]	
7	104		0	0	0				9	13	11	
8	84		32	34	36	32		34	31	16	38	31
9	102		0	0	0				16	22	19	
10	104		0	0	0			6	16	13	6	
11	94		0	0	0				44	50	31	

TABLE 3--Continued

Subject	Most-Understandable-Listening Level (dB re. ANSI-1969)	W-22 Test dB Levels re. MULL						Medial-Vowel Subtest dB Levels re. MULL				
		-8	-4	0	+4	+8	+12	-8	-4	0	+4	+8
12	80		22	24	26	20			50	55	34	
13	80		10	20	16				84	91	72	
14	74		8	16	20	18			81	94	91	
15	100		0	0	0				16	22	13	
16	104		0	0	0				13	22	16	
17	106		0	0	0				19	19	25[a]	
18	80	4	14	10	6				63	72	66	
19	74		6	12	14	10			75	81	78	
20	80		0	0	0				11	25	6	
21	102		0	0	0				9	11	6	
22	94		0	0	0				6	11	13	11
23	106		0	0	0			11	19	13	6	

[a]Upper limits of audiometer.

TABLE 4

INDIVIDUAL MOST-UNDERSTANDABLE-LISTENING LEVELS (MULL) AND SCORES (IN PERCENT) FOR THE INITIAL-CONSONANT SUBTEST AND THE FINAL-CONSONANT SUBTEST AT VARIOUS INTENSITY LEVELS

Subject	Most-Understandable-Listening Level (dB re. ANSI-1969)	Initial-Consonant Subtest dB Levels re. MULL					Final-Consonant Subtest dB Levels re. MULL					
		-8	-4	0	+4	+8	-8	-4	0	+4	+8	+12
1	96		33	15	38	25	30	35	25	18		
2	90		20	33	28			25	23	33	23	
3	88	41	45	40	38			23	38	35		
4	80		41	46	41			36	41	33		
5	62		41	53	71	69		36	44	54	58	48
6	106		30	33	38[a]			10	25	35[a]		
7	104		28	33	25		23	25	20	23		
8	84	55	58	50	55		28	35	28	20		
9	102		18	20	18			20	23	25	18	
10	104		25	28	20			20	28	15		
11	94	46	58	48	45		25	33	23	25		

TABLE 4--<u>Continued</u>

Subject	Most-Under-standable-Listening Level (dB re. ANSI-1969)	Initial-Consonant Subtest dB Levels re. MULL					Final-Consonant Subtest dB Levels re. MULL					
		-8	-4	0	+4	+8	-8	-4	0	+4	+8	+12
12	80		60	55	65	63		43	45	60	40	
13	80		60	65	63			50	50	58	54	
14	74		70	78	83	73	65	70	63	55		
15	100		21	23	25	21		30	33	35	25	
16	104		20	23	20		23	30	15	28		
17	106	30	35	30	20			10	30	35[a]		
18	80	53	63	60	43			38	43	38		
19	74	63	68	85	65			38	38	45	35	
20	80		20	21	18			18	23	20		
21	102		15	30	23			16	30	25		
22	94		23	33	23			28	33	23		
23	106		29	30	15			21	25	20		

[a]Upper limits of audiometer.

approximately the cut-off point for this group of subjects which demarcates 0 percent and non-0 percent performance on the W-22 test. It is more difficult to determine such a demarcation point for any of the UOST #6 subtests because of the effect that chance performance exercises on any given subject's score, but the question is considered later.

Both the initial-consonant and final-consonant subtests are four-alternative tests. Therefore, on the average, a score of 25 percent should be obtained on the basis of chance alone. The medial-vowel subtest is an eight-alternative test, and, on the average, a score of 12.5 percent should be obtained on the basis of chance alone. Therefore, a quantitative comparison of W-22 test and UOST #6 subtest scores, when the UOST #6 scores are low or when the W-22 scores are 0 percent, is of limited value or meaning.

The investigator considered excluding from the rankings and subsequent statistical analyses any score at or below chance performance level. This procedure was not followed, however, for the following three reasons:

1. It was not known what contribution guessing made to scores above chance performance.
2. It was not known what contribution actual correct performance (due to perceiving the test item correctly) made on scores below chance performance.

3. Spearman rank order correlation coefficients were performed on the data using both ranking methods (i.e., excluding all scores at chance performance level or below versus ranking all scores as they appeared). Little or no difference was found between the results of the two methods.

Theoretically, if only random chance performance scores were used in a correlation procedure, the resultant correlation coefficient would be at or near zero. Several significant correlation coefficients were obtained in the present study (later section of this chapter) using all of the UOST #6 subtest scores, even those below chance performance. This suggests that those scores at or below chance performance contain at least some "real data"; that is, the scores are affected in some way by auditory perception of the test items.

There appears perhaps to be a demarcation in the performance scores at MULL values of about 94 dB HL for the initial-consonant subtest, 88 dB HL for the final-consonant subtest, and 94 dB HL for the medial-vowel subtest. Above these MULL values, in each case, there appears to be a leveling off of the scores at values near chance performance. These MULL values cannot definitely be termed demarcation points because some scores obtained at higher MULL values may contain some "real data." There is no way to discern how much of any given score is due to chance performance or how much is due to correct performance.

Table 5 presents individual W-22 test scores (in percent) as well as UOST #6 initial-consonant-subtest, final-consonant-subtest, and medial-vowel-subtest scores (in percent). The scores recorded in Table 5 all represent the maximum-performance scores for each subject. Also included are the ranks for each score.

The W-22 scores recorded in Table 5 ranged from 0 to 36 percent. Only nine subjects achieved a score above 0 percent on the W-22 test. The mean score of these nine subjects is 24 percent. A mean score of 9.4 percent was obtained when the data from all 23 subjects were included.

Individual-subject scores on the UOST #6 initial-consonant subtest ranged from 20 to 85 percent with a mean of 44.7 percent. Individual-subject scores on the UOST #6 final-consonant subtest ranged from 23 to 70 percent with a mean of 38 percent. The UOST #6 medial-vowel-subtest scores ranged from 11 to 95 percent with a mean of 41.9 percent.

Average scores for all subjects show their performance to be best on the initial-consonant subtest, followed by the medial-vowel subtest and final-consonant subtest. This greater error probability for phonemes in the final-consonant position than in the initial-consonant position agrees with earlier studies (14, 27, 44, 48, 50). The ability of hearing-impaired subjects to produce higher scores

TABLE 5

INDIVIDUAL-MAXIMUM-PERFORMANCE SCORES AND RANKINGS FOR THE EXPERIMENTAL-SPEECH TESTS

Subject	W-22 Test		Initial-Consonant Subtest		Final-Consonant Subtest		Medial-Vowel Subtest	
	Score (%)	Rank[a]	Score (%)	Rank	Score (%)	Rank	Score (%)	Rank
1	0	. .	38	11.5	35	11.0	22	16.5
2	0	. .	33	15.0	33	15.0	53	7.5
3	24	5.0	45	10.0	38	8.0	53	7.5
4	28	3.0	46	9.0	41	7.0	52	9.0
5	34	2.0	71	3.0	58	3.5	92	2.0
6	0	. .	38	11.5	35	11.0	25	13.0
7	0	. .	33	15.0	25	21.0	13	21.5
8	36	1.0	58	7.5	35	11.0	38	11.0
9	0	. .	20	23.0	25	21.0	22	16.5
10	0	. .	28	19.0	28	19.0	16	20.0
11	0	. .	58	7.5	33	15.0	50	10.0

TABLE 5--Continued

Subject	W-22 Test		Initial-Consonant Subtest		Final-Consonant Subtest		Medial-Vowel Subtest	
	Score (%)	Rank[a]	Score (%)	Rank	Score (%)	Rank	Score (%)	Rank
12	24	4.0	65	4.5	60	2.0	55	6.0
13	20	6.5	65	4.5	58	3.5	91	3.0
14	20	6.5	83	2.0	70	1.0	94	1.0
15	0	. .	25	20.0	35	11.0	22	16.5
16	0	. .	23	21.0	30	17.5	22	16.5
17	0	. .	35	13.0	35	11.0	25	13.0
18	14	8.5	63	6.0	43	6.0	72	5.0
19	14	8.5	85	1.0	45	5.0	81	4.0
20	0	. .	21	22.0	23	23.0	25	13.0
21	0	. .	30	17.5	30	17.5	11	23.0
22	0	. .	33	15.0	33	15.0	13	21.5
23	0	. .	30	17.5	25	21.0	19	19.0
Range Mean	0-36 24.0 (only 9 S's)		20-85 44.7		23-70 38.0		11-94 41.9	

[a]Only those subjects who scored above 0 percent on the W-22 Test were included in the rankings.

. . = Unavailable Data Not Included in the Rankings on a closed-response-test paradigm (UOST #6) than on an open-response-test paradigm (W-22 test) is demonstrated in the present study and is supported by the results of earlier investigations (9, 71). The absolute scores produced by the subjects on the W-22 test are well below those obtained on any of the UOST #6 subtests, but, of course, the factor of chance must be considered.

As noted earlier, each subject was ranked relative to the other subjects on each of the experimental speech tests. These rankings are recorded in Table 5. Only those subjects who achieved a score were included in the ranking.

For the W-22 test and the three UOST #6 subtests a numerical rank of one (1) was assigned to the highest percentage score (best performance). When tied scores occurred, each score was assigned the average of the ranks which would have been assigned had no ties occurred. Period leaders (. .) represent unavailable data which was not included in the rankings.

Table 6 presents mean scores (in percent) for the experimental-speech tests at three intensity levels relative to the level at which maximum performance was obtained. Individual scores from nine subjects were used to calculate the mean scores for the W-22 test and individual scores from 23 subjects were used to calculate the mean scores for each of the UOST #6 subtests.

TABLE 6

MEAN SCORES (IN PERCENT) FOR THE EXPERIMENTAL-SPEECH TESTS AT THREE INTENSITY LEVELS FOR 23 SUBJECTS[a]

Experimental-Speech Test	Levels (in dB) re. the Level at Which Maximum Performance was Obtained		
	-4 dB	0	+4 dB
W-22 Test[a]	18.7	24.0	18.2
Initial-Consonant Subtest	36.9	44.7	38.1
Final-Consonant Subtest	31.1	38.0	29.8
Medial-Vowel Subtest	32.4	41.9	35.6

[a]Nine subjects received scores above 0 percent for the W-22 test.

Table 7 presents the number of subjects who obtained their maximum-performance scores for the experimental-speech tests at five intensity levels relative to the MULL. Six of a total of nine subjects performed best at a level 4 dB above their chosen MULL on the W-22 test. More subjects performed best at their chosen MULL rather than above or below it for each of the three UOST #6 subtests. This result suggests that these children are well able to identify their MULL's, at least as judged by UOST #6 results. It also suggests that the procedure used is adequate to accomplish this task and that maximum performance on the W-22 test required about 4 dB more intensity than is required for maximum performance on UOST #6.

Teacher Evaluations

Table 8 presents individual subject rankings from a teacher evaluation which was carried out as part of this study (Chapter III). Only the 10 subjects used in this study who attended the University of Oklahoma Medical Center School for the Deaf were evaluated in this part of the study.

Four teachers ranked the subjects (three ranked all 10 subjects, one ranked five subjects) on "speech production," "auditory speech understanding," "lipreading," "overall understanding," and "receptive problems other than hearing." The scores assigned by the four teachers were averaged, and then these average ranks were ranked from 1 to 10. In the first four categories, a rank of one (1) represents the best

TABLE 7

THE NUMBER OF SUBJECTS WHO OBTAINED MAXIMUM-PERFORMANCE SCORES ON THE EXPERIMENTAL-SPEECH TESTS AT FIVE INTENSITY LEVELS RE. THE MULL

Experimental-Speech Test	Levels (in dB) re. MULL					Total Subjects
	-8	-4	0	+4	+8	
W-22 Test	0	1	1	6	1	9
Initial-Consonant Subtest	0	5	12	6	0	23
Final-Consonant Subtest	0	5	9	8	1	23
Medial-Vowel Subtest	0	5	13	5	0	23

MULL = Most-Understandable-Listening Level

TABLE 8

THE RANKS OF THE AVERAGED RANKS APPLIED BY THE TEACHERS TO THE PERFORMANCE OF THE SUBJECTS IN VARIOUS CATEGORIES

Subject	Speech Production	Auditory Speech Understanding	Lipreading	Overall Understanding	Receptive Problems Other Than Hearing
1	9.0	8.0	10.0	9.0	2.0
2	10.0	5.0	8.5	10.0	6.0
3	5.0	4.0	7.0	6.0	5.0
4	3.0	3.0	2.0	2.0	NP
5	2.0	1.0	1.0	1.0	3.0
6	6.0	10.0	4.0	5.0	NP
7	7.0	7.0	6.0	7.0	NP
8	1.0	2.0	5.0	3.0	1.0
9	8.0	6.0	8.5	8.0	4.0
10	4.0	9.0	3.0	4.0	NP

NP = No Problems (not included in rankings) performance (i.e., best speech production, best lipreading, etc.). In the "receptive problems other than hearing" category, an NP represents no problems (normal). Four subjects were ranked as having no problems in this category. The remaining six subjects were ranked with a rank of one (1) representing the child judged as having the least receptive problems other than hearing loss outside of those with no problems.

Statistical Analyses

The analyses of the data required the extensive use of nonparametric statistics (Spearman rank correlation coefficients) (58). When tied scores occurred, each score was assigned the average of the ranks which would have been assigned had no ties occurred. When tied ranks occurred, a correction factor was incorporated into the computation of the Spearman rank correlation coefficients (58, p. 206).

Table 9 presents Spearman rank correlation coefficients among the experimental-speech tests (three UOST #6 subtests and the W-22 test results). Spearman rank correlation coefficients, as shown in Table 9, indicate non-significant (although negative) correlations between each of the three UOST #6 subtests and the W-22 test results. Table 9 also shows positive correlations (significant at the .01 level) among each of the UOST #6 subtest results.

As Table 3 indicates, only 9 of the 23 subjects scored above 0 percent on the W-22 test. Consequently, only

TABLE 9

SPEARMAN RANK CORRELATION COEFFICIENTS AMONG THE EXPERIMENTAL-SPEECH-TEST RESULTS FOR 23 SUBJECTS[a]

Experimental-Speech Test	W-22 Test[a]	Initial-Consonant Subtest	Final-Consonant Subtest	Medial-Vowel Subtest
W-22 Test[a]	X			
Initial-Consonant Subtest	-.38	X		
Final-Consonant Subtest	-.28	.88[b]	X	
Medial-Vowel Subtest	-.43	.80[b]	.82[b]	X

[a]Nine subjects were included in the ranking for the W-22 test.

[b]Significant at .01 level.

these nine subjects received a rank on the W-22 test and the UOST #6 subtests. The Spearman rank correlation coefficients among the three UOST #6 subtest results were performed using all 23 subjects.

The positive significant correlations among each of the UOST #6 subtest results indicate that those subjects who performed best on one subtest performed best on the other subtests. The W-22 test results, however, do not correlate significantly with any of the UOST #6 subtest results. The correlation coefficients, even though non-significant, are negative, indicating that those subjects who performed very well on the UOST #6 subtests tended not to perform as well on the W-22 test relative to the other subjects.

Table 10 presents Spearman rank correlation coefficients between the three UOST #6 subtest results, W-22 test results, and the following data: (a) chronological age, (b) I.Q. (Leiter), (c) I.Q. (W.I.S.C., performance section), (d) California Achievement Test, and (e) the three-frequency (500, 1K, 2K Hz) pure-tone average for the ear used in the experimental-data collection. These data were obtained from the individual subject's school files.

None of the experimental-test results (W-22 test or UOST #6 subtests) are significantly correlated with chronological age. The results of the I.Q. tests, both the Leiter and the performance section of the W.I.S.C., do not correlate

TABLE 10

SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN THE EXPERIMENTAL-SPEECH-TEST RESULTS AND INDIVIDUAL-SUBJECT DATA FROM SCHOOL FILES

	W-22 Test 9 S's	Initial-Consonant Subtest 23 S's	Final-Consonant Subtest 23 S's	Medial-Vowel Subtest 23 S's
Chronological Age 23 Subjects	-.55	.25	.24	.27
I.Q. (Leiter) 17 Subjects	-.08	-.09	.08	-.12
I.Q. (W.I.S.C.), Performance Section 14 Subjects	-.31	-.31	-.16	-.40[a]
California Achievement Test 22 Subjects	-.29	.41[a]	.36[a]	.30
Pure-Tone Average[c] (Experimental Ear) 23 Subjects	-.31	.81[b]	.73[b]	.82[b]

[a]Significant at .05 level.

[b]Significant at .01 level.

[c]500, 1K, and 2K Hz

significantly with any of the experimental-speech tests with the exception of a significant-negative correlation between the medial-vowel subtest results and I.Q. as derived from the performance section of the W.I.S.C. results. This negative correlation is barely significant at the .05 level. The reason for it is unknown. Overall, these results are believed to indicate that there is no relationship between the performance of these subjects on the experimental-speech tests and performance I.Q.

The results of the initial-consonant and final-consonant subtests are positively correlated (significant at the .05 level) with the results of the California Achievement Test. This indicates that the higher the grade placement of a subject the better was his performance on the UOST #6 consonant subtests. Thus, while there appears to be no relationship between scores on the experimental-speech tests and age or between the experimental-speech tests and I.Q., there does appear to be a weak relationship between the consonant-speech tests and grade level.

The results of the three-frequency (500, 1K, 2K Hz) pure-tone average for the ear used in the experimental-test session are positively correlated (significant at the .01 level) with each of the three UOST #6 subtest results. The pure-tone averages, however, do not correlate with the W-22 test results, and, in fact, the correlation is weakly negative.

Table 11 presents Spearman rank correlation coefficients between the W-22 test results, UOST #6 subtest results and the audiometric data (speech-reception threshold and three-frequency pure-tone average, 500, 1K, and 2K Hz) acquired as part of this study on the ear used in the experimental collection of the data. There is a positive correlation (significant at the .01 level) between each of the three UOST #6 subtest results and the pure-tone-average and speech-reception-threshold results. These positive-significant correlations between the UOST #6 results and the existing-audiometric data and the experimental-audiometric data for the experimental ear are particularly significant because of the small range of severe hearing losses among this group of subjects and also because some subjects presumably performed at or below chance performance level on the UOST #6 subtests. The W-22 test results, on the other hand, do not correlate with the pure-tone average for the experimental ear, taken from individual subjects' school files, or the pure-tone average and speech-reception threshold for the experimental ear, acquired as a part of this study. Only 9 of the 23 subjects scored above 0 percent on the W-22 test, and these subjects obtained scores over a very narrow range of values (14 to 36 percent). In contrast, the scores on the initial-consonant subtest ranged from 20 to 85 percent. They ranged from 23 to 70 percent on the final-consonant subtest and from 11 to 94 percent on the medial-vowel subtest.

TABLE 11

SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN THE EXPERIMENTAL-SPEECH-TEST RESULTS AND AUDIOMETRIC DATA FOR EXPERIMENTAL EAR ACQUIRED AS PART OF THIS STUDY FOR 23 SUBJECTS[a]

	W-22 Test[a]	Initial-Consonant Subtest	Final-Consonant Subtest	Medial-Vowel Subtest
Pure-Tone Average[c]	-.03	.72[b]	.71[b]	.78[b]
Speech-Reception Threshold	-.05	.77[b]	.69[b]	.83[b]

[a]Nine subjects were included in the ranking for the W-22 test.

[b]Significant at .01 level.

[c]500, 1K, and 2K Hz

This broad range of values supports the conclusion that these subtests are more appropriate for this group of hearing-impaired subjects than the W-22 test. The W-22 test produced a restricted range of scores because it was too difficult for these subjects. The broad range of scores obtained with UOST #6 undoubtedly contributed to the high positive correlations between these subtests and other measures of hearing function. The correlation coefficients between the W-22 test results and the experimental-audiometric data are very near zero, indicating very little or no relationship between W-22 test scores and other measures of hearing function in these subjects. This lack of correlation and the narrow range and low values of the scores obtained with the W-22 test are further evidence for the inappropriateness of this test with subjects of this type.

Table 12 presents Spearman rank correlation coefficients between the experimental-speech-test results, acquired-audiometric data for the experimental ear, and the teacher-evaluation results. The only positive-significant correlation between the W-22 test results and any of the categories included in the teacher evaluation is between the W-22 test results and the teacher-evaluated category of "speech production" (significant at the .05 level). Caution should be stressed in drawing conclusions about correlations between the W-22 test results and any of the teacher-evaluation-category results because only four subjects included in the teacher

TABLE 12

SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN THE EXPERIMENTAL-SPEECH-TEST RESULTS, ACQUIRED-AUDIOMETRIC DATA FOR THE EXPERIMENTAL EAR, AND TEACHER-EVALUATION RESULTS FOR 10 SUBJECTS[a]

	W-22 Test[a]	Initial-Consonant Subtest	Final-Consonant Subtest	Medial-Vowel Subtest	Pure-Tone Average	Speech-Reception Threshold
Speech Production	1.00[b]	.69[b]	.56	.29	.48	.58[b]
Auditory Speech Understanding	.80	.70[b]	.59[b]	.72[b]	.97[c]	.98[c]
Lipreading	.40	.50	.50	.25	.19	.33
Overall Understanding	.40	.67[b]	.62[b]	.31	.39	.50
Receptive Problems Other Than Hearing	.40	.50	.03	-.52	.49	-.37

[a]Four subjects were included in the ranking for the W-22 test.

[b]Significant at .05 level.

[c]Significant at .01 level.

evaluation scored above 0 percent on the W-22 test. In order to achieve statistical significance even at the .05 level with an N size of four, the correlation coefficient must equal ± 1.00 (a perfect correlation).

There are also positive correlations (significant at the .05 level) between teacher-evaluated "speech production" results and the initial-consonant-subtest results, and between the teacher-evaluated "speech production" results and speech-reception-threshold results. The final-consonant-subtest results, medial-vowel-subtest results, and pure-tone-average results are near to being significantly positively correlated with the teacher-evaluated "speech production" results. A weak and positive relationship between speech-production ability and measures of auditory-speech-discrimination ability seems reasonable in a group of subjects with hearing impairments of this range.

The teacher-evaluated category of "auditory speech understanding" is positively correlated (significant at the .05 level) with each of the three UOST #6 subtest results and more highly correlated (significant at the .01 level) with the three-frequency pure-tone-average and speech-reception-threshold results. It would appear that teacher judgments of "auditory speech understanding" ability are based mostly on their evaluation of hearing sensitivity and only to a lesser extent on evaluations of auditory-speech discrimination per se. The near perfect correlation coefficient between the teacher-evaluated category of "auditory speech understanding" and pure-tone-average results and the near perfect correlation coefficient between the teacher-evaluated category of "auditory speech understanding" and speech-reception-threshold results are certainly noteworthy. The correlation coefficients between the UOST #6 subtest results and the teacher-evaluated category of "auditory speech understanding," even though significant (.05 level), are smaller (.59 to .72) than the correlation coefficients between teacher-evaluated "auditory speech understanding" and speech-reception threshold and teacher-evaluated "auditory speech understanding" and pure-tone average (.97 and .98). The correlation coefficients between the W-22 test results and teacher-evaluated "auditory speech understanding" is of the same order of magnitude as those correlation coefficients between the UOST #6 subtest results and teacher-evaluated "auditory speech understanding" (.80). Due to the small N size (four), however, it is not significant. It does not appear unreasonable to conclude that a larger N size would reveal a significant correlation.

Teacher-evaluated "lipreading" results and teacher-evaluated "receptive problems other than hearing" results are not significantly correlated with any of the experimental-speech-test results or the acquired-audiometric data. This indicates that the UOST #6 results and the W-22 test results, pure-tone averages, and speech-reception thresholds are not related to this group of subjects' lipreading ability and other receptive problems such as vision, perceptual problems, etc. This outcome is encouraging because it seems reasonable that these auditory tests should be unrelated to lipreading ability and probably to other receptive problems as well.

Teacher-evaluated "overall understanding" results are positively correlated (significant at the .05 level) with the initial-consonant and final-consonant-subtest results, indicating that the ability to identify phonemes on the UOST #6 consonant subtests is related to the overall understanding ability of these subjects. The W-22 test, the medial-vowel subtest, pure-tone-average, and speech-reception-threshold results all do not correlate significantly with "overall understanding."

Table 13 presents Spearman rank correlation coefficients among the teacher-evaluated categories. The consonant-subtest results are correlated with the teacher-evaluated "overall understanding" results only at the .05 level. This result suggests that some factor other than hearing ability contributed substantially to this group of subjects' "overall understanding."

The positive Spearman rank correlation coefficient between the teacher-evaluated "overall understanding" results and the teacher-evaluated "lipreading" results is significant at the .01 level. The positive Spearman rank correlation coefficient between teacher-evaluated "overall understanding"

TABLE 13

SPEARMAN RANK CORRELATION COEFFICIENTS AMONG THE TEACHER-EVALUATED CATEGORIES FOR 10 SUBJECTS

	Speech Production	Auditory Speech Understanding	Lipreading	Overall Understanding	Receptive Problems Other Than Hearing
Speech Production	X				
Auditory Speech Understanding	.55	X			
Lipreading	.82[a]	.31	X		
Overall Understanding	.95[a]	.48	.93[a]	X	
Receptive Problems Other Than Hearing	.39	-.35	.56	.47	X

[a]Significant at .01 level.

results and teacher-evaluated "auditory speech understanding" results is not significant.

It is concluded that the significant-positive correlation between teacher-evaluated "lipreading" results and teacher-evaluated "overall understanding" results and the lack of a significant-positive correlation between teacher-evaluated "auditory speech understanding" results and teacher-evaluated "overall understanding" results indicates that lipreading and not auditory reception is the main contributor to the overall understanding ability of these subjects. The lack of a significant-positive correlation between teacher-evaluated "overall understanding" results and the acquired-audiometric data (pure-tone average and speech-reception threshold) for the experimental ear would also indicate that the auditory channel is less important to these subjects' overall understanding ability. These results also demonstrate that these children are, in fact, severely hearing impaired or even "deaf" by most definitions of that term.

There is a positive correlation (significant at the .01 level) between teacher-evaluated "speech production" results and teacher-evaluated "overall understanding" results. There is also a positive correlation (significant at the .01 level) between teacher-evaluated "speech production" results and teacher-evaluated "lipreading" results. These results indicate a strong relationship among lipreading ability, speech production ability, and overall understanding ability of these hearing-impaired subjects as judged by their teachers. The remaining correlations among the teacher-evaluated categories are not significant.

Discussion

It is believed that the present investigation has demonstrated that the closed-response-set-test paradigm is an appropriate format for an auditory-speech-discrimination test for severely hearing-impaired individuals. Further, it has been demonstrated that it is a more appropriate test than the W-22 test in that the results are better correlated with other measures of hearing function. The fact that none of the 23 subjects used in this study had been tested using the W-22 test prior to this investigation, presumably because the clinician thought the test was inappropriate or too difficult, would in itself support this conclusion. Only through a very carefully controlled procedure was the present investigator able to obtain W-22 test scores from 9 of the 23 subjects. These obtained scores are very poor. The lack of W-22 test scores and the very low W-22 test scores obtained in themselves testify to the inappropriateness of the W-22 test for this population.

The relative inappropriateness of the W-22 test becomes most evident when the results of the UOST #6 and the W-22 test are compared to other data, particularly hearing-function-dependent data, obtained from the subjects used in this study. Through the use of rank order correlation analyses, it was determined that there is no relationship between the performance of this group of subjects on UOST #6 and the following characteristics of the subjects: (a) chronological age, (b) I.Q. scores (Leiter and performance section of the W.I.S.C.), (c) teacher-evaluated "lipreading" ability, and (d) teacher-evaluated "receptive problems other than hearing." The performance of the subjects on the W-22 test also is not related to any of the above. These findings are believed favorable for both tests because the ability of a subject to perform on any auditory-speech-discrimination test should be unrelated, as much as possible, to age, performance I.Q., lipreading ability, and any other sensory receptive problems he may have.

The strong relationship between the performance of the subjects on UOST #6 and performance on the auditory tests of sensitivity (pure-tone average and speech-reception threshold) and the lack of such a relationship between the performance of the subjects on the W-22 test and the pure-tone-average and speech-reception-threshold data suggests that the closed-response-set-test paradigm (UOST #6) is a more valid auditory-speech-discrimination test of persons in the severely hearing-impaired category than is the W-22 test. This conclusion may be further supported by the significant correlations between the performance of the subjects on UOST #6 and teacher-evaluated "auditory speech understanding" results and teacher-evaluated "overall understanding" results.

The correlations between auditory speech discrimination as measured by the W-22 test and these two teacher-evaluated characteristics are approximately the same size but are not statistically significant. It does not appear unreasonable, however, to conclude that a larger N size would reveal a significant relationship here also. The observation that greater numbers of subjects performed above chance performance level on the UOST #6 subtests than above 0 percent on the W-22 test seems to support use of UOST #6, but this result is to a substantial but unknown extent artifactual.

The strong relationship of the performance of the subjects on the three UOST #6 subtests suggests some degree of reliability for UOST #6 as a whole as well as for its subtests. This is further indicated by the fact that previous investigators (14, 27, 44, 48, 50) found largely the same heirarchy of error probability for the consonant subtests and phoneme categories as was found in the present study. This will be discussed in the next section.

In conclusion, the closed-response-set-test paradigm, as a test format for an auditory-speech-discrimination test, appears most appropriate with severely hearing-impaired subjects whose level of performance is very low (but not 0 percent) on the W-22 test. Performance scores for these subjects on the closed-response-set-test paradigm (UOST #6) occurred over a broader range than did their scores on the

W-22 test and are strongly related to other hearing-function-dependent measures.

Qualitative (Error-Pattern) Analyses

UOST #6 was also developed as a tool which, hopefully, could be used to evaluate auditory-speech-discrimination-error patterns. The following discussion is a descriptive analysis of the response errors (substitutions, additions, and omissions) which occurred within phoneme groups and within subtests by the subjects used in this study.

Only the errors of those subjects who performed above the chance level of performance on the initial-consonant and final-consonant subtests or medial-vowel subtest are included in this section of the study. This includes 13 subjects for the initial-consonant and final-consonant subtests and 18 subjects for the medial-vowel subtest. The results are compared to the results of previous studies.

Only outstanding agreements and/or disagreements between previous studies and the present study are explicitly discussed in order to decrease the amount of relatively unimportant information and, thereby, improve the clarity. It should be kept in mind that hearing-impaired children served as subjects in the present study. Some previous studies used normal-hearing subjects (Abbs and Minifie (1), Griffiths (17), House, *et al.* (27), Miller and Nicely (44), and Pederson (51)), while other studies used hearing-impaired subjects (Lawrence and Byers (35), Owens, Talbott, and Schubert (49), and Oyer and Doudna (50)). Studebaker (64) used both normal-hearing and hearing-impaired subjects.

Initial-Consonant Subtest

The error matrices for the different phoneme groups within the initial-consonant subtest are presented in Table 14. More errors were observed among the voiceless-fricative-phoneme group than among any other phoneme group. The next largest error frequency was observed in the voiced-plosive group, followed successively by the voiced-fricative group, the voiceless-plosive group, and the affricative-blend group. There were considerably fewer errors noted for the last group than for the other groups noted above.

The data of Pederson (51) revealed a different overall heirarchy of phoneme-group breakdown for his normal-hearing subjects. He found the least number of errors in the affricative-blend group which concurs with the results of the present study. Pederson, however, reported that more errors were observed among the voiceless-plosive phoneme group for the initial-consonant subtest than for any other group. The next largest error frequency occurred in the voiceless-fricative group, followed successively by the voiced-plosive group, the voiced-fricative group, and the affricative-blend group. His data yielded considerably fewer errors for each of the three latter phoneme groups than for either of the

TABLE 14

INITIAL-CONSONANT-SUBTEST-ERROR MATRICES
FOR 13 SUBJECTS (RAW SCORES)

Voiceless-Plosive

Presented-Responded

pp	pt	pk	p-	tt	tp	tk	t-	kk	kp	kt	k-	--	-p	-t	-k
31	4	13	4	26	12	11	3	32	7	8	5	36	4	6	6

Voiced-Plosive

Presented-Responded

bb	bd	bg	b-	dd	db	dg	d-	gg	gb	gd	g-	--	-b	-d	-g
30	12	8	2	32	11	7	2	21	8	13	10	32	4	6	10

TABLE 14--Continued

Voiceless-Fricative															
Presented-Responded															
ff	fs	fθ	f-	ss	sf	sθ	s-	θθ	θf	θs	θ-	--	-f	-s	-θ
14	17	19	2	30	9	8	5	28	12	12	0	20	6	15	11

Voiced-Fricative															
Presented-Responded															
vv	vz	vð	v-	zz	zv	zð	z-	ðð	ðv	ðz	ð-	--	-v	-z	-ð
22	12	15	3	32	7	11	2	35	10	7	0	29	16	1	6

Affricative-Blend															
Presented-Responded															
st st	st ʃ	st tʃ	st-	ʃ ʃ	ʃ st	ʃ tʃ	ʃ-	tʃ tʃ	tʃ st	tʃ ʃ	tʃ-	--	-st	-ʃ	-tʃ
40	2	10	0	35	8	9	0	32	5	12	3	41	4	6	1

first two groups noted above. The data of Miller and Nicely (44), also obtained from normal-hearing subjects, generally support Pederson's (51) heirarchy of phoneme-group-error frequency.

In the initial-consonant subtest, errors occurred with the greatest frequency within the voiceless-fricative group. More errors are noted for the /f/ presentations than for the /s/, /θ/, and /-/ presentations, with the /θ/ for /f/ substitution occurring more often than any other. Abbs and Minifie (1), Griffiths (17), Pederson (51) (who all used normal-hearing subjects), and Lawrence and Byers (35) (who used hearing-impaired subjects) also noted that the /θ/ for /f/ substitution occurred with great frequency. These findings are very consistent with those of Studebaker (64) from both normal-hearing and hearing-impaired subjects.

The next greatest number of errors occurred in the voiced-plosive group. The present results agree with Miller and Nicely (44), who used normal-hearing subjects, in that the /d/ for /g/ substitution was exhibited more often than the other errors in this phoneme group. The present results also agree with Pederson (51) and Studebaker (64), both of whom employed normal-hearing and hearing-impaired subjects, in that the /d/ for /b/ substitution was seen frequently.

In the voiced-fricative group, the /v/ for /-/ addition and the /ð/ for /v/ substitution occurred most often. The /ð/ for /v/ substitution occurred most often in the

Pederson (51) study of normal-hearing subjects and was frequently noted by Studebaker (64) from hearing-impaired and normal-hearing subjects. This substitution was frequently noted from normal-hearing subjects in studies by Abbs and Minifie (1), Griffiths (17), and Miller and Nicely (44).

Within the voiceless-plosive group, more total errors were noted for the /t/ presentations than for the /p/, /k/, and /-/ presentations. This agrees with the Studebaker (64) data from both normal-hearing and hearing-impaired subjects. Miller and Nicely (44) found that the single /k/ for /p/ substitution occurred more often than any other voiceless-plosive substitution error with normal-hearing subjects. The /k/ for /p/ substitution occurred more frequently than any other voiceless-plosive error in the present study also.

The affricative-blend-group presentations produced the least number of errors of any one of the above-mentioned phoneme groups. The most frequent substitution was /ʃ/ for /tʃ/. No other discernible pattern of errors could be identified in this group, possibly as a result of the low total error frequency.

Final-Consonant Subtest

The subjects obtained more total errors for the final-consonant subtest than for the initial-consonant subtest. This finding agrees with the results of Fairbanks (14), House, et al. (27), Miller and Nicely (44), Owens and Schubert (48), Oyer and Doudna (50), and Pederson (51). The error matrices for the different phoneme groups within the final-consonant subtest are presented in Table 15.

The heirarchial sequence of phoneme-group error frequency for the final-consonant subtest is generally the same as for the initial-consonant subtest. The voiceless-plosive and affricative-blend groups each contain fewer errors than any other one phoneme group for both subtests. The voiced-plosive and voiceless-fricative groups each contain more errors than any other one phoneme group. These findings disagree with the Pederson (51) study of normal-hearing subjects tested on the same test used in the present study in that he reported the voiceless phonemes were heard better than the voiced. The House, et al. (27) study of normal-hearing subjects tested on a different test than that used in the present study concurs with the Pederson study. The data from the present study agree with the Studebaker (64) data obtained from both hearing-impaired and normal-hearing subjects in that the voiced and voiceless phoneme groups do not differ significantly in error frequency. Studebaker employed a test similar to the one used in the present study.

In the voiced-plosive group, the /g/ presentations were responsible for the greatest number of errors. These errors are primarily /b/ for /g/ substitutions and secondarily /d/ for /g/ substitutions. Studebaker's (64) data, which was obtained from hearing-impaired subjects, exhibited the same

TABLE 15

FINAL-CONSONANT-SUBTEST-ERROR MATRICES
FOR 13 SUBJECTS (RAW SCORES)

Voiceless-Plosive

Presented-Responded

pp	pt	pk	p-	tt	tp	tk	t-	kk	kp	kt	k-	--	-p	-t	-k
31	10	8	3	36	8	4	4	15	9	24	4	28	10	6	8

Voiced-Plosive

Presented-Responded

bb	bd	bg	b-	dd	db	dg	d-	gg	gb	gd	g-	--	-b	-d	-g
22	11	4	14	24	7	10	11	10	25	15	3	24	10	10	8

TABLE 15--Continued

Voiceless-Fricative

Presented-Responded

ff	fs	fθ	f-	ss	sf	sθ	s-	θθ	θf	θs	θ-	--	-f	-s	-θ
16	10	20	6	22	12	4	14	22	8	17	5	20	14	8	10

Voiced-Fricative

Presented-Responded

vv	vz	ðv	v-	zz	zv	ðz	z-	ðð	ðv	ðz	ð-	--	-v	-z	-ð
24	8	5	15	33	2	12	6	10	8	12	22	30	8	6	8

Affricative-Blend

Presented-Responded

st st	st ʃ	st tʃ	st-	ʃ ʃ	ʃ st	ʃ tʃ	ʃ-	tʃ tʃ	tʃ st	tʃ ʃ	tʃ-	--	-st	-ʃ	-tʃ
30	12	4	6	20	7	21	4	23	11	10	8	30	4	10	8

error patterns within this phoneme group. Pederson's (51) data obtained from normal hearing subjects revealed that the most errors occurred with the /b/ presentations, with the /g/ for /b/ substitution appearing most frequently.

The voiceless-fricative group contains the same total error frequency as the voiced-plosive group. The greatest number of errors in this group occurred with the /f/ presentations, with the /θ/ for /f/ substitution occurring most frequently. Abbs and Minifie (1), Griffiths (17), Lawrence and Byers (35), and Pederson (51) also found that the /θ/ for /f/ substitution was very prominent.

Among the voiced-fricative group, the /v/ and /ð/ presentations produced the most errors. The /-/ for /v/ omission occurred most frequently. There are a large number of /-/ for /ð/ omissions also. This is identical to the data of Studebaker (64) obtained from both normal-hearing and hearing-impaired subjects. Pederson (51) also noted the large number of /-/ for /v/ omissions and /-/ for /ð/ omissions. Abbs and Minifie (1), Griffiths (17), and Lawrence and Byers (35) also observed a high frequency of errors with the /v/ and /ð/ presentations. Griffiths (17) contrasted these two phonemes and indicated that the /v/ and /ð/ are virtually indistinguishable by normal-hearing subjects on the basis of acoustical cues alone.

The affricative-blend group presentations produced fewer errors than any one of the above phoneme groups. Only the /tʃ/ for /ʃ/ substitution occurred frequently. Pederson (51) found that errors occurred in this phoneme group only occasionally, with no identifiable error patterns. At the time of this writing, there were no other available data on error-matrix patterns obtained with these phonemes.

The voiceless-plosive-group presentations produced the least number of errors of any other one phoneme group. The most frequent error within this group is the /t/ for /k/ substitution. This same substitution was reported by Pederson (51). There are very few other errors within the voiceless-plosive group. Studebaker (64) also found that his hearing-impaired subjects made very few errors in this phoneme group.

Medial-Vowel Subtest

The error-matrices for the different phonemes within the medial-vowel subtest are presented in Table 16. Presentation of the vowel /I/ produced substantially more errors than any of the other single vowels studied, while the second largest number of errors occurred with the presentation of /ɛ/. The presentation of the phonemes /e/, /ʌ/, /æ/, and /i/ produced the next greatest error counts in that order. The presentation of /u/ and /o/ produced fewer errors than presentations of other phonemes. Studebaker (64) and Pederson (51) both observed more errors with /I/ and /ɛ/ presentations than with any other phoneme presentations from

TABLE 16

MEDIAL-VOWEL-SUBTEST-ERROR MATRICES
FOR 18 SUBJECTS (RAW SCORES)

Presented-Responded							
ii	iI	iɛ	iæ	ie	iʌ	io	iu
76	8	18	16	8	4	12	2
Presented-Responded							
II	Ii	Iɛ	Iæ	Ie	Iʌ	Io	Iu
57	18	33	2	4	6	10	14
Presented-Responded							
ɛɛ	ɛi	ɛI	ɛæ	ɛe	ɛʌ	ɛo	ɛu
65	8	9	14	13	25	6	4
Presented-Responded							
ææ	æi	æI	æɛ	æe	æʌ	æo	æu
74	12	4	10	6	20	8	10

TABLE 16--*Continued*

Presented-Responded							
ee	ei	eI	eɛ	eæ	eʌ	eo	eu
69	18	12	23	12	4	4	2

Presented-Responded							
ʌʌ	ʌi	ʌI	ʌɛ	ʌæ	ʌe	ʌo	ʌu
72	12	12	5	14	15	6	8

Presented-Responded							
oo	oi	oI	oɛ	oæ	oe	oʌ	ou
81	8	16	4	14	4	5	12

Presented-Responded							
uu	ui	uI	uɛ	uæ	ue	uʌ	uo
80	7	9	3	1	15	9	20

normal-hearing subjects. The data of Oyer and Doudna (50) also revealed that the /I/ and /ɛ/ phonemes are frequently missed and that they are often confused for one another. Pederson (51) further found that the presentation of the phoneme /o/ produced considerably fewer errors than presentations of the other phonemes which agrees with the present study.

The major substitutions in the order of their frequency of occurrence are /ɛ/ for /I/, /ʌ/ for /ɛ/, /ɛ/ for /e/, /ʌ/ for /æ/, and /o/ for /u/. The substitutions /ɛ/ for /i/, /i/ for /I/, /i/ for /e/, /æ/ for /i/, and /I/ for /o/ occurred less often than the previous errors and occurred with approximately equal frequency. The remaining possible substitutions not mentioned here occurred only infrequently.

Pederson (51) found the /u/ for /i/ and /i/ for /u/ substitutions to be two of the most often observed errors in his study. The opposite appears true in the present study; that is, /u/ for /i/ and /i/ for /u/ substitutions occurred only rarely. The data of Owens, Talbott, and Schubert (49), obtained from hearing-impaired subjects, and that of Studebaker (64), obtained from normal-hearing subjects, agree with the present study in that they did not show this substitution as a major one.

One of the most frequently observed phoneme substitutions in this study, /ʌ/ for /ɛ/, is in agreement with Owens, Talbott, and Schubert's (49) data which was obtained from hearing-impaired subjects. It was also the most often observed substitution in the Pederson (51) study. This substitution, however, is not noted frequently in other works.

Oyer and Doudna (50) noted that the /I/ for /ɛ/ and /ɛ/ for /I/ substitutions occurred frequently. This partially agrees with the present study in that the /ɛ/ for /I/ substitution was the most frequent of any single error substitution. The /I/ for /ɛ/ substitution, on the other hand, occurred only rarely.

Discussion

There is partial agreement between the overall heirarchy of phoneme-group-error frequency for the initial-consonant and final-consonant subtests and phoneme-error frequency for the medial-vowel subtest observed between the present study and earlier investigations. The lack of perfect agreement may be due, in part, to the fact that the various studies did not investigate all of the same consonant and vowel groups investigated by the present study. However, a major factor must be the nature of the subjects used in this study. Most of the studies reviewed used normal-hearing subjects. Other studies used subjects with hearing impairments of various types. In view of these differences in subjects, the similarities in the error patterns appear somewhat remarkable.

When the error patterns within the overall phoneme groupings are analyzed, substantial agreement between the present study and previous investigations is in evidence. As might be expected, there is a tendency for more error-pattern agreements between those few studies which used hearing-impaired subjects and the present study than between the larger number of studies which used only normal-hearing subjects and the present study.

The error-pattern agreement between previous studies which used normal-hearing subjects and the present study, nevertheless, is very good. There are only three outstanding error-pattern disagreements (out of 30 comparisons) between those previous studies that used normal-hearing subjects and the present study for both consonant and vowel error patterns. These three disagrements occurred in the Pederson (51) study. Pederson reported that the following substitutions occurred frequently in his study of normal-hearing subjects: /g/ for /b/, /u/ for /i/, and /i/ for /u/. These substitutions occurred only rarely in the present study. There is only one outstanding error disagreement (out of 15 comparisons) between the present study and previous studies in which only hearing-impaired subjects were used. Oyer and Doudna (50) noted that the /I/ for /ɛ/ substitution occurred frequently in their study of hearing-impaired subjects. The present investigator found that this substitution occurred only rarely.

The good agreement for error-pattern analyses between previous studies which used normal-hearing subjects and the present study suggests that severely hearing-impaired subjects and normal-hearing subjects listening in noise produce error patterns that are largely, though not perfectly, identical. Commonalities of the input signal and/or of the receptor mechanism are apparently sufficient to override, for the most part, the differences produced by a severe hearing impairment; or perhaps, to a large extent, a severe hearing impairment has an effect on the input signal (consonant and vowel phonemes) that is similar, at least in this regard, to that produced by a thermal noise.

The excellent error-pattern agreement between previous studies which used only hearing-impaired subjects and the present study suggests that the error patterns produced by hearing-impaired subjects do have some few commonalities which differentiate them from those produced by normal-hearing subjects. At the present time, however, these are little more than suggestions for further research into the nature of the effect of a severe hearing impairment on the perception of speech signals.

Previous attempts to evaluate auditory-speech-discrimination ability analytically utilizing open-response tests (PB-word lists) have been unsuccessful (8, 50, 57). Error-pattern matrices obtained in the present study demonstrate that the closed-response-set-test paradigm (UOST #6) can be used successfully to evaluate auditory-speech-discrimination ability analytically, at least on a group basis. These results also demonstrate that the types of phoneme errors produced by a group of hearing-impaired subjects occur in reasonably consistent patterns which are largely consistent with the patterns produced by normal-hearing subjects listening in noise. There are, however, some features unique to those patterns produced by severely hearing-impaired subjects.

CHAPTER V

SUMMARY

Many of the auditory-speech-discrimination tests used today in clinical-hearing evaluations are based on test paradigms that were originally designed for testing communication systems. These clinical tests have generally taken the form of monosyllabic, "phonetically balanced" (PB), meaningful word test item tests. The phonetic content of the lists is ostensibly representative of the statistical occurrence of the phonetic or phonemic elements within the parent language. The PB-word lists provide a numerical measurement of the subject's overall speech-discrimination ability. Attempts to evaluate speech discrimination analytically with the PB-word lists have also been made but have not proven successful.

Researchers concerned about the need for better tests to measure auditory-speech-discrimination ability of both normal and hearing-impaired listeners have developed some new test paradigms. The most recent of these efforts has been directed toward the development of a format which limits the response alternatives available to a subject (a closed-response set). These response alternatives are selected without regard to phonetic or phonemic balance. Rather, they are selected on the basis of such criteria as varying the speech sound (the test phoneme) in only one phoneme position across a set of CVC words. The number of articulatory-classification parameters varied within a test set may be controlled in this paradigm; for example, only the place of articulatory production of the test phonemes may be varied within the set.

There is an unfulfilled need for an auditory-speech-discrimination test specifically designed for use with the severely hearing-impaired population. When commonly used auditory-speech-discrimination tests are administered to members of this hearing-impaired population, the results are usually so poor as to indicate only that the person has a severe hearing loss, a finding which may be an oversimplification of his status.

The purpose of the present investigation was to contribute needed information about the performance of severely hearing-impaired children on a closed-response-set test called the University of Oklahoma Speech Test #6 (UOST #6). UOST #6 is a closed-response set, monosyllabic word, auditory-speech-discrimination test consisting of three independent subtests. These subtests include an 80-item initial-consonant subtest, an 80-item final-consonant subtest, and a 64-item medial-vowel subtest.

This closed-response set auditory-speech-discrimination test was presented to 23 severely hearing-impaired subjects (12 males and 11 females) between the ages of 9 years 3 months and 16 years 6 months. Their pure-tone averages for the experimental ear ranged from 60 to 110+ dB with a mean of about 88.5 dB. The data obtained from the performance of these subjects on UOST #6 were compared to other tests of hearing function performed in this study (CID W-22 test, pure-tone average, and speech-reception threshold). The data were also compared with such characteristics of the children as chronological age, I.Q. (performance section of the Wechsler Intelligence Scale for Children, W.I.S.C.), I.Q. (Leiter International Performance Scale), existing audiometric data, achievement test data (California Achievement Test), and teacher evaluations of each child's "speech production," "auditory speech understanding," "lipreading," "overall understanding," and "receptive problems other than hearing." These results are reported under Quantitative Analyses. Data obtained from subjects' performance on UOST #6 are also presented in error-matrix form and are reported under Qualitative Analyses. The following relationships are indicated.

Quantitative Analyses

None of the 23 subjects used in this study had been successfully tested prior to this investigation using the W-22 test, presumably because the clinician thought the test was too difficult or that it was inappropriate. Only nine of the 23 subjects produced supra-0 percent W-22 test scores in the present study. The W-22 scores for these nine subjects are very low, ranging from 12 to 36 percent. The 0 percent W-22 scores on 14 children and the very low W-22 test scores obtained from the others testify to the inappropriateness of the W-22 test for this hearing-impaired population.

Spearman rank order correlation coefficients indicated the following:

1. There is no relationship between the performance of these subjects on the two speech-discrimination tests (UOST #6 and W-22 test) and the following data: (a) subjects' chronological ages, (b) subjects' I.Q. scores (determined by the Leiter or the performance section of the W.I.S.C.), (c) subjects' lipreading ability (determined by teacher evaluation), and (d) subjects' receptive problems other than hearing (determined by teacher evaluation). These findings suggest an independence of auditory-speech-discrimination ability as measured by either test from factors such as I.Q., age, etc. This would appear to be a favorable characteristic in an auditory-speech-discrimination test.

2. The performance of these subjects on UOST #6 is strongly related to their performance on tests of

auditory sensitivity (pure-tone average and speech-reception threshold). The results of the W-22 test do not show such a relationship to the pure-tone-average or the speech-reception-threshold results.

3. The performance of these subjects on UOST #6 is related to teacher-evaluated "auditory speech understanding" and teacher-evaluated "overall understanding" results. This suggests that the closed-response-set test (UOST #6) is a fair predictor of auditory-speech-understanding ability and overall-understanding ability for a group of severely hearing-impaired subjects. The correlations between auditory speech discrimination as measured by the W-22 test and these two teacher-evaluated characteristics are approximately the same size but are not statistically significant. It does not appear unreasonable, however, to conclude that a larger N size would reveal a significant relationship here also.
4. The performance of these subjects on each of the three UOST #6 subtests is strongly related and suggests some degree of reliability for the test as a whole and interchangeability of its subtests.
5. Teachers of this group of hearing-impaired subjects appear able to rank them with great

consistency in the various categories (speech production, auditory speech understanding, lipreading, overall understanding, and receptive problems other than hearing).

Those subjects who chose a most-understandable-listening level (MULL) of approximately 90 dB HL (ANSI 1969 Norms) did not perform above 0 percent on the W-22 test. These same subjects also generally performed at or near chance performance level on each of the UOST #6 subtests.

In conclusion, the closed-response-set-test paradigm, as a test format for an auditory-speech-discrimination test, appears most appropriate for use with severely hearing-impaired subjects whose level of performance is very low (but not 0%) on the W-22 test. Performance scores for these subjects on the closed-response-set-test paradigm (UOST #6) occurred over a broader range of values than did their scores on the W-22 test. The UOST #6 scores are highly correlated to data dependent upon hearing function, whereas the W-22 test scores are not significantly correlated with other tests and evaluations of hearing function for this group of subjects.

Qualitative Analyses

The phoneme groups within the initial-consonant and final-consonant subtests are listed in order according to the highest response-error frequency. The predominant phoneme response errors (substitutions, additions, and omissions) are also listed in order under each phoneme group according to the highest frequency of occurrence.

Initial-Consonant Subtest

1. Voiceless-Fricative

   /θ/ for /f/ substitution

2. Voiced-Plosive

   /d/ for /g/ substitution

3. Voiced-Fricative

   /v/ for /-/ addition

   /ð/ for /v/ substitution

4. Voiceless-Plosive

   /k/ for /p/ substitution

5. Affricative-Blend

   /ʃ/ for /tʃ/ substitution

Final-Consonant Subtest

1. Voiced-Plosive

   /b/ for /g/ substitution

   /d/ for /g/ substitution

2. Voiceless-Fricative

   /θ/ for /f/ substitution

   (Response errors within the above two phoneme groups occurred with the same total frequency.)

3. Voiced-Fricative

   /-/ for /v/ omission

   /-/ for /ð/ omission

4. Affricative-Blend

/tʃ/ for /ʃ/ substitution

5. Voiceless-Plosive

/t/ for /k/ substitution

Medial-Vowel Subtest

The predominant phoneme-response errors (substitutions) within the medial-vowel subtest are listed in order according to frequency of occurrence as follows:

/ɛ/ for /I/	/ɛ/ for /i/
/ʌ/ for /ɛ/	/æ/ for /i/
/ɛ/ for /e/	/i/ for /I/
/ʌ/ for /æ/	/i/ for /e/
/o/ for /u/	/I/ for /o/

There is partial agreement between the present study and previous studies for the phoneme-error-pattern analyses for each of the UOST #6 subtests. That disagreements should occur does not seem surprising due to the fact that many of these previous studies used normal-hearing subjects or subjects with much less of a hearing impairment than those used in this study. The error-pattern agreement is, in fact, quite good between the present study and those studies which used normal-hearing subjects, and the agreement is very good between the present study and those that used only hearing-impaired subjects.

It appears that severely hearing-impaired subjects and normal-hearing subjects listening in noise produce error patterns that are largely, although not perfectly, identical.

Perhaps the characteristics of the input signal and/or of the receptor mechanism are sufficiently predominant to override, for the most part, the effect produced by a severe hearing impairment; or, perhaps, to a large extent, thermal noise has an effect on the input signal (consonant and vowel phonemes) that is similar to that produced by a hearing impairment.

The error-pattern matrices obtained in the present study indicate that a closed-response-set-test paradigm (UOST #6) can successfully demonstrate auditory-speech-discrimination-error patterns on a subject-group basis. The study also demonstrated that the types of phoneme errors produced by a group of hearing-impaired subjects are largely consistent with the errors produced by normal-hearing subjects listening in noise but have some few features unique to those patterns produced only by the severely hearing-impaired.

BIBLIOGRAPHY

1. Abbs, M. S., and Minifie, F. D., Effect of acoustic cues in fricatives on perceptual confusions in preschool children. J. Acoust. Soc. Amer., 46, 1535-1542 (1969).

2. Berger, K. W., A speech discrimination task using multiple-choice key words in sentences. J. Aud. Res., 9, 247-262 (1969).

3. Black, J. W., Multiple-choice intelligibility tests. J. Speech Hearing Dis., 22, 213-235 (1957).

4. Boothroyd, A., Statistical theory of the speech discrimination score. J. Acoust. Soc. Amer., 43, 362-367 (1967).

5. Bricker, W. A., Errors in the echoic behavior of children. J. Speech Hearing Res., 10, 67-76 (1967).

6. Carhart, R., and Jerger, J. F., The Hughson-Westlake method of pure-tone determination. J. Speech Hearing Dis., 24, 330-354 (1959).

7. Carrell, J., and Tiffany, W. R., Phonetics: Theory and Application to Speech Improvement. New York: McGraw-Hill (1960).

8. Castle, W. E., A conference on hearing-aid evaluation procedures, tests and measurements. ASHA, 2, 5-13 (1967).

9. Dale, D. M. C., Applied Audiology for Children. Springfield, Illinois: Charles C. Thomas (1962).

10. Davis, H., The articulation area and social adequacy index of hearing. Laryngoscope, 63, 761-778 (1948).

11. Dewey, G., Relative Frequency of English Speech Sounds. Cambridge, Mass.: Harvard University Press (1923). Cited in House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D., Articulation testing methods: consonantal differentiation with a closed-response set. J. Acoust. Soc. Amer., 37, 158-166 (1965).

12. Egan, J. P., Articulation testing methods. <u>Laryngoscope</u>, 58, 955-991 (1948).

13. Elpern, B. S., Difference in difficulty among the CID W-22 auditory tests. <u>Laryngoscope</u>, 70, 1560-1565 (1960).

14. Fairbanks, G., Test of phonemic differentiation: The Rhyme Test. <u>J. Acoust. Soc. Amer.</u>, 30, 596-600 (1958).

15. French, N. R., Carter, C. W., and Koenig, W., The words and sounds of telephone conversations. <u>Bell Systems Technical Journal</u>, 9, 290-324 (1958). Cited in Fairbanks, G., Test of phonemic differentiation: The Rhyme Test. <u>J. Acoust. Soc. Amer.</u>, 30, 596-600 (1958).

16. Goetzinger, C. P., and Rousey, C. L., A study of the Wechsler performance scale and the Knox cube test with deaf adolescents. <u>Amer. Annals Deaf</u>, 102, 388-399 (1957).

17. Griffiths, J. D., Rhyming minimal contrasts: a simplified diagnostic articulation test. <u>J. Acoust. Soc. Amer.</u>, 42, 236-241 (1967).

18. Haagen, C. H., Intelligibility measurements: twenty-four word multiple-choice tests. Psychological Corporation, New York, OSRD Report #5567 (1945). Cited in Studebaker, G., The audiologic application of tests of phonemic differentiation. University of Oklahoma Speech and Hearing Center, Rehabilitation Service Administration Grant #RD-2401-S-67 Report (1968).

19. Hanson, G. H., Hancock, B. B., and Kopra, L. L., Relationships among audiological status, linguistic skills, visual-motor perception, and academic achievement of deaf children. Final Report Project #6-2353, Office of Education, U.S. Depart. HEW, Depart. Speech, University of Texas, Austin (1969).

20. Harold, B. B., <u>The Effect of Variations in Intensity on the Capacity of Deaf Children and Adults to Hear Speech with Hearing Aids</u>, Ph.D. Dissertation, Manchester University (1954). Cited in Dale, D. M. C., <u>Applied Audiology for Children</u>. Springfield, Illinois: Charles C. Thomas (1962).

21. Harris, J. D., Some suggestions for speech reception testing. Naval Medical Research Laboratory, New London, Connecticut, Report #2, Project NM-003-021 (1948). Cited in Hirsh, I. J., <u>The Measurement of Hearing</u>. New York: McGraw-Hill (1952).

22. Haskins, H., A Phonetically Balanced Test of Speech Discrimination for Children, M. A. Thesis, Northwestern University (1949). Cited in Newby, H. A., Audiology. Second Edition. New York: Appleton-Century-Crofts (1964).

23. Heinz, J. M., and Stevens, K. N., On the properties of voiceless fricative consonants. J. Acoust. Soc. Amer., 17, 321-337 (1952).

24. Hirsh, I. J., Clinical application of two Harvard auditory tests. J. Speech Hearing Dis., 12, 151-158 (1947).

25. Hirsh, I. J., The Measurement of Hearing. New York: McGraw-Hill (1952).

26. Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., and Benson, R. W., Development of materials for speech audiometry. J. Speech Hearing Dis., 17, 321-337 (1952).

27. House, A. S., Williams, E. C., Hecker, M. H. L., and Kryter, K. D., Articulation-testing methods: consonantal differentiation with a closed-response set. J. Acoust. Soc. Amer., 37, 158-166 (1965).

28. Hutton, C., Curray, E. T., and Armstrong, M. D., Semi-diagnostic test materials for aural rehabilitation. J. Speech Hearing Dis., 24, 319-329 (1959).

29. Jerger, J. F., Behavioral correlates of hearing-aid performance. Bulletin of Prosthetics Research, BPR 10-7, 62-75 (1967).

30. Jerger, J. F., Malmquist, C., and Speaks, C., Comparison of some speech intelligibility tests in the evaluation of hearing aid performance. J. Speech Hearing Res., 9, 253-258 (1966).

31. Kendall, D. C., The Mental Development of Young Deaf Children, Ph.D. Dissertation, Manchester University (1953). Cited in Dale, D. M. C., Applied Audiology for Children. Springfield, Illinois: Charles C. Thomas (1962).

32. Kopra, L. L., Blosser, D., and Waldron, L., Comparison of Fairbanks Rhyme Test and CID Auditory Test W-22 in normal and hearing-impaired listeners. J. Speech Hearing Res., 11, 735-739 (1968).

33. Kruel, J. E., Nixon, J. C., Kryter, K. D., Bell, D. W., Lang, J. S., and Schubert, E. D., A proposed clinical test of speech discrimination. J. Speech Hearing Res., 7, 536-552 (1968).

34. Larr, A., and Cain, E., Measurement of native learning abilities of deaf children. Volta Review, 61, 160-162 (1959).

35. Lawrence, D. B., and Byers, V. W., Identification of voiceless fricatives by high frequency hearing impaired listeners. J. Speech Hearing Res., 12, 426-434 (1969).

36. Lavos, G., W.I.S.C. psychometric patterns among deaf children. Volta Review, 64, 547-552 (1962).

37. Lehiste, I., Readings in Acoustic Phonetics. Cambridge, Mass.: M.I.T. Press (1967).

38. Lehiste, G., and Peterson, I., Linguistic considerations in the study of speech intelligibility. J. Acoust. Soc. Amer., 31, 280-286 (1959).

39. Leiter, R. G., Evidence of the reliability and validity of the Leiter tests. Booklet published by Western Psychological Services, Beverly Hills, California (1948).

40. Levine, E. S., Studies in psychological evaluation of the deaf. Volta Review, 65, 499-503 (1963).

41. Ling, D., Three experiments on frequency transposition. Amer. Annals Deaf, 133, 283-294 (1968).

42. Ling, D., and Druz, W. S., Transposition of high frequency speech sounds by partial vocoding of the speech spectrum. J. Aud. Res., 7, 133-144 (1967).

43. MacFarland, D., Speech-hearing tests. Laryngoscope, 55, 71-115 (1945).

44. Miller, G. A., and Nicely, P. E., An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Amer., 27, 338-352 (1955).

45. Murray, N. E., Speech tests of hearing. Report I.R.4, Commonwealth Acoustic Laboratories, Australia (1952). Cited in Dale, D. M. C., Applied Audiology for Children. Springfield, Illinois: Charles C. Thomas (1962).

46. Myatt, B. D., and Landes, B. A., Assessing discrimination loss in children. Arch. Otolaryng., 77, 359-362 (1963).

47. Newby, H. A., Audiology. Second Edition. New York: Appleton-Century-Crofts (1964).

48. Owens, E., and Schubert, E. D., The development of consonant items for speech discrimination testing. J. Speech Hearing Res., 11, 656-667 (1968).

49. Owens, E., Talbott, C. B., and Schubert, E. D., Vowel discrimination of hearing-impaired listeners. J. Speech Hearing Res., 11, 648-655 (1968).

50. Oyer, H. J., and Doudna, M., Structural analysis of word responses made by hard-of-hearing subjects on a discrimination test. Arch. Otolaryng., 70, 357-364 (1959).

51. Pederson, O. T., The University of Oklahoma Speech Test #6, Ph.D. Dissertation, The University of Oklahoma (1969).

52. Pollack, I., Decker, L. R., and Rubenstein, H., Intelligibility of selected message sets. J. Acoust. Soc. Amer., 30, 643-644 (1958).

53. Pollack, I., Rubenstein, H., and Decker, L., Intelligibility of known and unknown message sets. J. Acoust. Soc. Amer., 31, 273-279 (1959).

54. Quick, M. A., A test for measuring achievement in speech perception among young deaf children. Volta Review, 55, 28-31 (1953).

55. Resnick, D., and Becker, M., Hearing aid evaluation - a new approach. ASHA, 5, 695-699 (1963).

56. Savin, H. B., Word-frequency effect and errors in the perception of speech. J. Acoust. Soc. Amer., 35, 200-206 (1963).

57. Schultz, M. D., Suggested improvements in speech discrimination testing. J. Aud. Res., 4, 1-14 (1964).

58. Seigel, Sidney., Non-parametric Statistics for the Behavioral Sciences. New York: McGraw-Hill (1956).

59. Shore, I., Bilger, R. C., and Hirsh, I. J., Hearing aid evaluations: reliability of repeated measurements. J. Speech Hearing Dis., 25, 152-170 (1960).

60. Shore, I., and Kramer, J., A comparison of two procedures for hearing-aid evaluation. J. Speech Hearing Dis., 28, 159-170 (1963).

61. Silverman, S. R., and Hirsh, I. J., Problems related to the use of speech in clinical audiometry. Annals Otol., Rhin., and Laryng., 64, 1234-1244 (1955).

62. Smith, C. S., The assessment of mental ability in partially deaf children. Teacher of the Deaf, 60, 216-224 (1962).

63. Sommerville, S., An Investigation of the University of Oklahoma Recordings of the Modified Northwestern University Auditory Test #6, M. A. Thesis, University of Oklahoma (1967).

64. Studebaker, G., The audiologic application of tests of phonemic differentiation. University of Oklahoma Speech and Hearing Center, Rehabilitation Service Administration Grant #RD-2401-S-67 Report (1968).

65. Thorndike, E. L., and Lorge, I., Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College, Columbia University (1944).

66. Tiegs, E. W., and Clark, W. W., California Achievement Tests manual. Booklet published by California Test Bureau, Del Monte Research Park, Monterey, California (1957).

67. Tillman, T. W., and Carhart, R., An expanded test for speech discrimination utilizing CNC monosyllabic words (N.U. Auditory Test No. 6). Technical Documentary Report #SAM-TDR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas (1966). Cited in Sommerville, S., An Investigation of the University of Oklahoma Recordings of the Modified Northwestern University Auditory Test #6, M. A. Thesis, University of Oklahoma (1967).

68. Tillman, T. W., Carhart, R., and Wilber, L., A test for speech discrimination composed of CNC monosyllabic words (N.U. Auditory Test No. 4). Technical documentary report #SAM-TDR-62-135, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas (1962). Cited in Sommerville, S., An Investigation of the University of Oklahoma Recordings of the Modified Northwestern University Auditory Test #6, M. A. Thesis, University of Oklahoma (1967).

69. Vernon, M., and Brown, D. W., A guide to psychological tests and testing procedures in the evaluation of deaf and hard of hearing children. J. Speech Hearing Dis., 29, 414-423 (1964).

70. Watson, L. A., and Tolan, T., Hearing Tests and Hearing Instruments. Baltimore: The Williams Wilkins Co. (1949).

71. Watson, T. J., Speech audiometry for children. Chapter in Ewing, A. W. G., Educational Guidance and the Deaf Child. Washington, D. C.: The Volta Bureau (1957).

72. Wechsler, David, Wechsler Intelligence Scale for Children. Booklet published by the Psychological Corporation, New York (1949).

73. Zerlin, S. A., A new approach to hearing aid selection. J. Speech Hearing Res., 5, 370-376 (1962).

APPENDIXES

Appendix A

RESEARCH PROJECT CONSENT

Subject's Name:

Address:

Age:

School:

Parents' Name:

I consent to allow my child to participate in a research project being conducted at the University of Oklahoma Speech and Hearing Center. I understand that my child's services are being used entirely for research purposes and not for the immediate personal benefit of my child. The purposes of the experiment and the procedures that my child will take part in have been explained to me fully, and I voluntarily agree to allow my child to participate in them.

____
Date

____
Parent's Signature

Appendix B

NURSERY RHYME USED FOR FINDING MOST-UNDERSTANDABLE-LISTENING LEVEL

Ba-a, Ba-a, Black sheep, have you any wool?

Yes, sir, yes, sir, three bags full:

One for my master, one for my dame,

And one for the little boy that lives in our lane.

Appendix C

SPONDEE WORD LIST SCRAMBLING

1.	AIRPLANE	25.	SNOWMAN	49.	TOOTHBRUSH	73.	BATHTUB
2.	TOOTHBRUSH	26.	TOOTHBRUSH	50.	SANDBOX	74.	SAILBOAT
3.	SAILBOAT	27.	HIGHCHAIR	51.	SAILBOAT	75.	SAILBOAT
4.	TOOTHBRUSH	28.	WRISTWATCH	52.	ICECREAM	76.	ICECREAM
5.	CUPCAKE	29.	CUPCAKE	53.	ICECREAM	77.	HIGHCHAIR
6.	HIGHCHAIR	30.	SAILBOAT	54.	CUPCAKE	78.	SNOWMAN
7.	BATHTUB	31.	HIGHCHAIR	55.	BATHTUB	79.	SNOWMAN
8.	SNOWMAN	32.	BATHTUB	56.	WRISTWATCH	80.	HIGHCHAIR
9.	CUPCAKE	33.	TOOTHBRUSH	57.	BATHTUB	81.	ICECREAM
10.	HIGHCHAIR	34.	SNOWMAN	58.	AIRPLANE	82.	ICECREAM
11.	TOOTHBRUSH	35.	BATHTUB	59.	AIRPLANE	83.	TOOTHBRUSH
12.	ICECREAM	36.	SNOWMAN	60.	SNOWMAN	84.	ICECREAM
13.	COWBOY	37.	SNOWMAN	61.	SNOWMAN	85.	TOOTHBRUSH
14.	SNOWMAN	38.	SANDBOX	62.	SANDBOX	86.	SNOWMAN
15.	FIRETRUCK	39.	BATHTUB	63.	BATHTUB	87.	SAILBOAT
16.	SNOWMAN	40.	AIRPLANE	64.	TOOTHBRUSH	88.	CUPCAKE
17.	ICECREAM	41.	SANDBOX	65.	COWBOY	89.	CUPCAKE
18.	HIGHCHAIR	42.	CUPCAKE	66.	TOOTHBRUSH	90.	BATHTUB
19.	CUPCAKE	43.	HIGHCHAIR	67.	HIGHCHAIR	91.	BATHTUB
20.	TOOTHBRUSH	44.	AIRPLANE	68.	CUPCAKE	92.	SAILBOAT
21.	BATHTUB	45.	CUPCAKE	69.	HIGHCHAIR	93.	SAILBOAT
22.	AIRPLANE	46.	BATHTUB	70.	HIGHCHAIR	94.	BATHTUB
23.	AIRPLANE	47.	WRISTWATCH	71.	SANDBOX	95.	CUPCAKE
24.	AIRPLANE	48.	CUPCAKE	72.	SANDBOX	96.	CUPCAKE

Appendix D

INSTRUCTIONS TO SUBJECTS

Initial-Consonant and Final-Consonant Subtests

These are the instructions. Please read them carefully. In one ear you will hear, "The word is ____." For example, "The word is bat." Underline the word on the response sheet you think you heard.

pat   bat   mat   at

Do you have any questions?

Medial-Vowel Subtest

These are the instructions. Please read them carefully. In one ear you will hear, "The word is ____." For example, "The word is look." Put a check in the square under the word on the response sheet you think you heard.

Presentation #	lack	leak	look	like	lock	luck	lick	lake
1.			X					

Do you have any questions?

Appendix E

STUDENT EVALUATION FORM

Subject	Speech Production	Auditory Speech Understanding	Lipreading	Overall Understanding	Receptive Problems Other Than Hearing
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					